Cosparse regularization of physics-driven inverse problems

Srdan Kitic

ANNÉE 2015

THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Traitement du Signal et Télécommunications

Ecole doctorale MATISSE

présentée par

Srđan Kitić

préparée à l'unité de recherche 6074 IRISA
(Institut de Recherche en Informatique et Systèmes Aléatoires)

Cosparse regularization of physics-driven inverse problems

Thèse soutenue à l'IRISA
le 26 novembre 2015

devant le jury composé de :

Laurent DAUDET
Professeur, Université Paris 7, Laboratoire Ondes et Acoustiques / rapporteur

Patrick FLANDRIN
Directeur de recherche CNRS, ENS de Lyon, Membre de l'Académie des Sciences / rapporteur

Michael DAVIES
Professeur, University of Edinburgh / examinateur

Laurent DUVAL
Ingénieur de Recherche, chef de projet, IFP Energies Nouvelles / examinateur

Laurent ALBERA
Maître de conférences, Université de Rennes 1, LTSI / membre invité

Rémi GRIBONVAL
Directeur de recherche INRIA / directeur de thèse

Nancy BERTIN
Chargé de recherche CNRS / co-directrice de thèse

Dedicated to my family

Acknowledgements

I am deeply thankful to my supervisors for their guidance during the last three years. I feel flattered and very fortunate to become Rémi's student - the depth of his mathematical insight and vision in broad range of subjects continues to amaze me. Despite my numerous limitations, he was always very patient and supportive. On top of that, he was an extremely gentle and carrying advisor. A huge thank you also goes to Nancy, both for her invaluable advices, but also for her help in many other ways - enabling us to settle down was only one of them. Simply saying that she is a wonderful person is diminishing. Many thanks and credits also to Laurent Jacques and Christophe De Vleeschouwer of UCL, my first scientific guides, and the most influential for my decision to pursuit a scientific career.

Additionally, I would like to thank the jury: Laurent Daudet, Patrick Flandrin, Mike Davies and Laurent Duval for taking time off their duties to review the thesis and/or participate in examination. Their feedback and suggestions were a highly valuable asset. A special acknowledgment for Laurent Albera, who was not only my main collaborator, but also an indispensable source of knowledge, kindness and support. In the same line, I need to thank Çağdaş and Gilles for tutoring me on various aspects of optimization, and to Çağdaş and Jules for taking part in the infamous "string theory" experiment (keep the faith!).

It was very pleasant being a member of PANAMA (*i.e.* METISS) team, and the credit for that goes to all my colleagues and friends: Alexis, Anaik, Anthony, Çağdaş, Cedric, Clement, Corentin(s), Ewen, Frederic, Gabriel, Gilles, Jeremy, Joachim, Jules, Laurence, Laurent, Luc, Nicolas(es), Patrick, Pierre, Roman, Stanislaw, Stefan, Stephanie, Thomas and Yann (if I forgot to mention someone it was an honest mistake). Thanks a lot guys for enjoyable coffee-breaks and developing a caffeine addiction.

Likewise, I am very grateful to my parents and my sister, for their warming love, understanding and sacrifice during... well, entire life. Finally, the biggest thanks is for my little family, my two sweethearts Dragana and Nevena, for their endless love and patience, and because of whom all this actually makes sense.

Rennes, S. K.
November 26, 2015.

Abstract

Inverse problems related to physical processes are of great importance in practically every field related to signal processing, such as tomography, acoustics, wireless communications, medical and radar imaging, to name only a few. At the same time, many of these problems are quite challenging due to their ill-posed nature.

On the other hand, signals originating from physical phenomena are often governed by laws expressible through linear Partial Differential Equations (PDE), or equivalently, integral equations and the associated Green's functions. In addition, these phenomena are usually induced by sparse singularities, appearing as sources or sinks of a vector field. In this thesis we primarily investigate the coupling of such physical laws with a prior assumption on the sparse origin of a physical process. This gives rise to a "dual" regularization concept, formulated either as sparse analysis (*cosparse*), yielded by a PDE representation, or equivalent sparse synthesis regularization, if the Green's functions are used instead. We devote a significant part of the thesis to the comparison of these two approaches. We argue that, despite nominal equivalence, their computational properties are very different. Indeed, due to the inherited sparsity of the discretized PDE (embodied in the analysis operator), the analysis approach scales much more favorably than the equivalent problem regularized by the synthesis approach.

Our findings are demonstrated on two applications: acoustic source localization and epileptic source localization in electroencephalography. In both cases, we verify that cosparse approach exhibits superior scalability, even allowing for full (time domain) wavefield extrapolation in three spatial dimensions. Moreover, in the acoustic setting, the analysis-based optimization benefits from the increased amount of observation data, resulting in a speedup in processing time that is orders of magnitude faster than the synthesis approach. Numerical simulations show that the developed methods in both applications are competitive to state-of-the-art localization algorithms in their corresponding areas. Finally, we present two sparse analysis methods for blind estimation of the speed of sound and acoustic impedance, simultaneously with wavefield extrapolation. This is an important step toward practical implementation, where most physical parameters are unknown beforehand. The versatility of the approach is demonstrated on the "hearing behind walls" scenario, in which the traditional localization methods necessarily fail.

Additionally, by means of a novel algorithmic framework, we challenge the audio declipping problem regularized by sparsity or cosparsity. Our method is highly competitive against state-of-the-art, and, in the cosparse setting, allows for an efficient (even real-time) implementation.

Résumé en français

Le résumé suivant propose un survol intuitif du contenu de cette thèse, en langue française. Un panorama de l'état de l'art, le détail des méthodes proposées et les perspectives futures ouvertes par notre travail sont disponibles (en anglais) dans le reste du manuscrit.

Introduction Si l'on devait décrire de la manière la plus concise possible le traitement du signal en tant que discipline, on pourrait probablement dire qu'il s'agit de la discipline s'attachant à *résoudre des problèmes inverses*. En effet, pratiquement toutes les tâches de traitement de signal, aussi naïves fussent-elles, peuvent être formulées comme des problèmes inverses. Malheureusement, beaucoup de problèmes inverses sont *mal posés* ; ils sont généralement abordés par le biais de *techniques de régularisation* appropriées.
La régularisation au moyen d'un modèle parcimonieux des données (également appelé modèle de parcimonie à la synthèse, ou tout simplement *parcimonie*) est une tendance désormais bien installée (elle dure depuis plus de vingt ans !) et qui a été appliquée avec succès à de nombreux cas. Son succès est attribué à une explication intuitive, selon laquelle les signaux de la Nature admettent des descriptions « simples » – dans le cas de la parcimonie à la synthèse, une combinaison linéaire de quelques atomes choisis dans un dictionnaire. Plus récemment, une régularisation alternative (ou complémentaire) a émergé : le modèle de parcimonie à l'analyse (ou *coparcimonie*), dans lequel on suppose que le signal peut être rendu parcimonieux par l'application d'une transformation linéaire bien choisie, désignée sous le nom d'*opérateur d'analyse*. Ces deux modèles sont fondamentalement différents, en dehors du cas particulier où le dictionnaire et l'opérateur sont inverses l'un de l'autre. En règle générale, on ne peut répondre catégoriquement à la question : « quel est le meilleur modèle ? ». Il est plutôt supposé que leur utilité dépend principalement du problème particulier que l'on est en train de considérer. Cependant, les études qui comparent vraiment ces deux modèles, en dehors du contexte purement théorique, sont extrêmement rares. Dans les travaux que nous présentons, nous visons à faire la lumière sur cette question, en nous concentrant sur une classe de problèmes inverses liés aux processus physiques, que nous baptisons *problèmes inverses gouvernés par la Physique*.

Prologue : la désaturation audio Avant de plonger dans nos contributions principales, nous prendrons un détour. Nous explorons le problème inverse de la désaturation des signaux audibles, régularisé par un modèle parcimonieux ou coparcimonieux. La saturation d'amplitude,

en anglais *clipping*, se produit souvent lors d'un enregistrement audio, de sa restitution ou lors des conversions analogique-numérique. Ce problème commun en traitement de signal audio existe également dans les domaines du traitement de l'image ou des communications numériques (par exemple, en OFDM).
L'observation-clef est que la saturation produit des discontinuités, qui se traduisent en une dispersion de l'énergie dans le plan temps-fréquence. Cette constatation peut être exploitée pour inverser le processus : construire un estimateur du signal d'origine qui soit cohérent avec les contraintes liées à la saturation et dont l'énergie soit concentrée en temps-fréquence. Notre but est de développer un algorithme de désaturation audio, compétitif face à l'état de l'art, qui puisse intégrer de la même manière une hypothèse de parcimonie à la synthèse ou à l'analyse, de manière à former un bon indicateur de comparaison des deux modèles. Ce but est atteint dans le cadre algorithmique que nous avons baptisé *SParse Audio DEclipper (SPADE)*. Il déploie la régularisation parcimonieuse ou coparcimonieuse par une approche gloutonne non-convexe, fondée sur les algorithmes de type *Alternating Direction Method of Multipliers (ADMM)*.
Les résultats sont présentés en termes de performances numériques et d'évaluation perceptive, et incluent une comparaison avec l'état de l'art. Ils nous ont amenés à la conclusion que la méthode fondée sur la parcimonie à la synthèse est légèrement plus performante en termes de reconstruction du signal, mais au prix d'un coût computationnel énorme. D'autre part, la version fondée sur la parcimonie à l'analyse se situe à peine en-dessous en termes de performance, mais permet une mise en œuvre extrêmement efficace, permettant même un traitement en temps-réel. De surcroît, les deux versions de SPADE sont tout-à-fait compétitives face aux approches de l'état de l'art.

Problèmes inverses gouvernés par la Physique Nous poursuivons nos investigations avec des problèmes inverses soulevés dans un contexte physique (que nous appelons « gouvernés par la Physique »), qui sont des problèmes d'une grande importance pratique dans bien des domaines reliés au traitement du signal. Ils sont fondamentaux dans des applications telles que la tomographie, l'acoustique, les communications sans fil, le radar, l'imagerie médicale, pour n'en nommer que quelques unes. Dans le même temps, beaucoup de ces problèmes posent de grands défis, en raison de leur nature mal-posée. Cependant, les signaux qui émanent de phénomènes physiques sont souvent gouvernés par des lois connues, qui s'expriment sous la forme d'équations aux dérivées partielles (EDP). Pourvu que certaines hypothèses sur l'homogénéité des conditions initiales et aux limites soient vérifiées, ces lois possèdent une représentation équivalente sous la forme d'équations intégrales, et des fonctions de Green associées.
De plus, les phénomènes physiques considérés sont souvent induit par des singularités que l'on pourrait qualifier de parcimonieuses, décrites comme des *sources* ou des *puits* dans un champ vectoriel. Dans cette thèse, nous étudions en premier lieu le couplage entre de telles lois physiques et une hypothèse initiale de parcimonie des origines du phénomène physique. Ceci donne naissance à un concept de dualité des régularisations, formulées soit comme un problème d'analyse coparcimonieuse (menant à la représentation en EDP), soit comme une

parcimonie à la synthèse équivalente à la précédente (lorsqu'on fait plutôt usage des fonctions de Green). Nous nommons ce concept *cadre (co)parcimonieux gouverné par la Physique (physics-driven (co)sparse framework)* et dédions une part significative de notre travail à la comparaison entre les approches de synthèse et d'analyse. Nous défendons l'idée qu'en dépit de leur équivalence formelle, leurs propriétés computationnelles sont très différentes. En effet, en raison de la parcimonie héritée par la version discrétisée de l'EDP[1] (incarnée par l'opérateur d'analyse), l'approche coparcimonieuse passe bien plus favorablement à l'échelle que le problème équivalent régularisé par parcimonie à la synthèse. Afin de résoudre les problèmes d'optimisation convexe découlant de l'une et l'autre des approches de régularisation, nous développons une version générique et sur-mesure de l'algorithme *Simultaneous Direction Method of Multipliers (SDMM)*, baptisé *Weighted SDMM*. Nos constatations sont illustrées dans le cadre de deux applications : la localisation de sources acoustiques, et la localisation de sources de crises épileptiques à partir de signaux électro-encéphalographiques.

Application 1 : localisation de sources acoustiques Parmi bien d'autres applications, la localisation de sources acoustiques (ou sonores) est notamment utilisée pour le débruitage, la déréverbération, le suivi de sources, le positionnement de robots, ou l'imagerie sismique ou médicale. Les méthodes traditionnelles de localisation de source sont fondées sur l'estimation de la différence de temps d'arrivée (en anglais TDOA pour *Time Difference Of Arrival*) ou sur des techniques de formation de voies *(beamforming)*. Toutes ces approches, qu'elles soient plus ou moins performantes en terme de robustesse ou de précision, souffrent invariablement de la réverbération (c'est-à-dire l'existence de trajets acoustiques multiples) et visent à en supprimer les effets. Pourtant, dans d'autres domaines tels que les communications sans fil, les chemins multiples sont régulièrement et efficacement exploités en tant que source supplémentaire d'information, souvent dans le but d'améliorer le Rapport Signal-sur-Bruit (en anglais SNR pour *Signal-to-Noise Ratio*). Inspirés par ces techniques et motivés par le succès de plusieurs travaux récents dans cette direction, nous proposons une méthode générique de localisation de sources sonores qui s'appuie sur l'interpolation du champ sonore.
Après avoir rappelé que la propagation du son dans l'air est modélisée par une EDP linéaire dépendant du temps appelée *équation des ondes*, nous la discrétisons et l'embarquons dans un opérateur d'analyse qui s'exprime sous forme matricielle (et qui incorpore également les conditions initiales et aux bords). Dans ce cadre, la représentation équivalente par les fonctions de Green est obtenue en formant un dictionnaire de synthèse qui n'est autre que l'inverse matriciel de l'opérateur d'analyse. En supposant que le nombre de sources est petit par rapport à l'ensemble de l'espace discrétisé, nous pouvons alors formuler un problème inverse régularisé d'interpolation du champ sonore, c'est-à-dire un problème d'estimation du champ de pression acoustique à *toutes* les coordonnées de l'espace-temps discrétisé. L'estimateur obtenu est alors seuillé afin de déterminer les positions potentielles des sources sonores.
Nos simulations indiquent que les deux approches, parcimonieuse comme coparcimonieuse, atteignent de hautes performances de localisation, et, comme prévu, qu'elles produisent des

[1] Pourvu que la méthode de discrétisation choisie soit à support local.

estimées identiques (à la précision numérique près). Cependant, la seconde démontre une meilleure capacité de passage à l'échelle ($O(st)$ *vs* $O(mst^2)$, où m, s et t désignent respectivement le nombre de microphones, de points dans l'espace et d'échantillons temporels), au point qu'elle permet même une interpolation complète du champ de pression dans le temps et en trois dimensions. De plus, l'optimisation fondée sur le modèle d'analyse *bénéficie* d'une augmentation du nombre de données observées, ce qui débouche sur une accélération du temps de traitement, qui devient plus rapide que l'approche de synthèse dans des proportions atteignant plusieurs ordres de grandeur. Enfin, l'approche proposée est compétitive face à la version stochastique de l'algorithme SRP-PHAT, qui constitue actuellement l'état de l'art dans la tâche de localisation de source.

Scénarios avancés de localisation coparcimonieuse de sources sonores L'approche précédemment introduite repose lourdement sur la connaissance explicite de la géométrie spatiale de la pièce, de la paramétrisation des conditions aux limites, et du milieu de propagation. Afin de relâcher ces hypothèses incommodes, nous proposons deux algorithmes réalisant l'estimation simultanée du champ de pression acoustique et de certains de ces paramètres physiques.

En premier lieu, nous considérons le cas d'une vitesse du son inconnue, ce qui est pertinent d'un point de vue pratique, en raison par exemple de l'existence d'un gradient de température dans la pièce. Nous introduisons l'hypothèse raisonnable que la vitesse du son est constante dans le temps et fonction assez régulière de l'espace. Concrètement, nous considérons qu'elle peut être approchée par un polynôme discrétisé d'ordre r, ce qui réduit drastiquement le nombre de degrés de liberté pour ce paramètre ($O(r^d)$ *vs* $O(st)$, où d est le nombre de dimensions). Le problème de l'estimation simultanée de la vitesse du son et du champ de pression sonore est biconvexe, et nous lui appliquons une heuristique de type ADMM non-convexe pour en approcher la solution. Cette méthode est baptisée *Blind Localization and Sound Speed estimation (BLESS)*. Les résultats préliminaires indiquent qu'une estimation presque parfaite est possible lorsque $r = 1$ ou $r = 2$, au prix d'une augmentation modérée du nombre de microphones (par rapport aux cas où la vitesse de propagation du son est parfaitement connue au préalable).

Dans un second scénario, nous étudions la possibilité d'estimer simultanément le champ de pression acoustique et le coefficient d'impédance acoustique spécifique, qui paramétrise les bords du domaine. C'est également un problème qui a des implications pratiques importantes, car il est généralement très difficile de deviner précisément et à l'avance la valeur de ce paramètre physique. Pour contourner le caractère mal-posé de ce problème, nous supposons que l'impédance est constante par morceaux, ce qui est justifié physiquement par le fait que les bords sont habituellement constitués de structures macroscopiquement homogènes, telles que des murs, des portes ou des fenêtres. L'hypothèse nous suggère qu'une régularisation de type *variation totale* peut être utilisée pour promouvoir des solutions de cette nature. À nouveau, l'estimation simultanée est formulée comme un problème biconvexe et résolue par une forme non-convexe d'ADMM. Les résultats de simulation sont étonnamment optimistes, puisque notre méthode, baptisée *Cosparse Acoustic Localization, Acoustic Impedance estima-*

tion and Signal recovery (CALAIS), atteint des résultats presque identiques aux résultats d'une localisation coparcimonieuse standard en présence de conditions aux bords parfaitement connues.
Pour finir, nous démontrons la capacité de la localisation coparcimonieuse à aborder un problème où les méthodes traditionnelles échoueraient nécessairement. Dans ce scénario, que nous appelons « entendre derrière les murs », les sources sonores et les microphones sont séparés par un obstacle acoustiquement opaque qui empêche toute observation du chemin direct de propagation (mais permet à des réflexions d'atteindre les microphones). Tandis que l'absence de la contribution du chemin direct à l'observation interdit toute application des méthodes classiques fondées sur le TDOA, la localisation coparcimonieuse exploite l'information contenue dans les échos pour réaliser une localisation précise des sources, même lorsque la « porte » qui permet le passage de ces chemins multiples est relativement petite.

Application 2 : localisation de sources dans le cerveau Notre dernière application cible est l'électro-encéphalographie (EEG), ou plus précisément, la localisation de sources de crises épileptiques à partir des mesures du potentiel électrique sur le scalp. Le modèle physique sous-jacent, qui lie les potentiels en surface et les sources épileptiques, c'est-à-dire les courants électriques distincts dans le cerveau, est gouverné par l'équation de Poisson. En sus, les sources sont modélisées comme des dipôles électriques, ce qui mime l'activité corrélée de groupes de neurones parallèles. Enfin, il est physiologiquement admis que les sources pertinentes se situent exclusivement dans la région du cortex, et sont orientées perpendiculairement à la matière grise. Ces hypothèses facilitent la résolution du problème inverse émergeant du système de mesures (limité à des électrodes sur la surface de la tête), qui serait autrement très mal posé.
Malheureusement, ces connaissances et hypothèses préalables restent insuffisantes pour assurer que le problème inverse de localisation de sources en EEG soit bien posé. Par conséquent, le problème est généralement abordé par des techniques variées, par exemple statistiques (fondées sur les moments ou les cumulants d'ordre supérieur), ou variationnelles (par exemple la régularisation de Tikhonov). Plusieurs méthodes récentes supposent que les sources sont spatialement parcimonieuses, ce qui est également l'approche que nous avons choisie.
La méthode que nous proposons découle tout naturellement de notre cadre de régularisation (co)parcimonieuse gouvernée par la physique. La discrétisation de l'équation de Poisson et l'ajout du modèle de sources dipolaires conduit à l'expression de l'opérateur d'analyse. Le dictionnaire de synthèse correspondant se réduit, à nouveau, à l'inverse matriciel de l'opérateur d'analyse. Comme dans le cas de l'acoustique, la version c« analyse » passe bien mieux à l'échelle que la version « synthèse », qui toutes les deux fournissent des performances compétitives devant l'état de l'art. La méthode proposée se révèle particulièrement robuste au cas où les sources épileptiques sont mutuellement dépendantes. Dans ce cas, les performances des méthodes d'inversion statistiques (par exemple, la bien connue méthode *MUltiple SIgnal Classification – MUSIC*) décroissent très significativement.

Publications

Journal:

S. Kitić, L. Albera, N. Bertin, and R. Gribonval. Physics-driven inverse problems made tractable with cosparse regularization. To appear in *IEEE Transactions on Signal Processing*, 2015.

Conference:

N. Bertin, S. Kitić, and R. Gribonval. Joint estimation of sound source location and boundary impedance with physics-driven cosparse regularization. To appear in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2016.

S. Kitić, N. Bertin and R.Gribonval. Sparsity and cosparsity for audio declipping: a flexible non-convex approach. In *Latent Variable Analysis and Signal Separation*, pages 243-250. Springer, 2015.

L. Albera, S. Kitić, N. Bertin, G. Puy and R.Gribonval. Brain source localization using a physics-driven structured cosparse representation of EEG signals. In *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1-6. IEEE, 2014.

S. Kitić, N. Bertin and R.Gribonval. Hearing behind walls: localizing sources in the room next door with cosparsity. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3087-3091. IEEE, 2014.

S. Kitić, N. Bertin, and R.Gribonval. A review of cosparse signal recovery methods applied to sound source localization. In *Le XXIVe colloque Gretsi*, 2013.

Workshop:

S. Kitić, N. Bertin and R.Gribonval. Time-data tradeoff for the sparse and cosparse regularizations of physics-driven inverse problems, *SPARS15 - Signal Processing with Adaptive Sparse Structured Representations*, Cambridge, United Kingdom, 2015.

Ç. Bilen, S. Kitić, N. Bertin and R.Gribonval. Sparse acoustic source localization with blind calibration for unknown medium characteristics. In *iTwist-2nd international-Traveling Workshop on Interactions between Sparse models and Technology*, 2014.

S. Kitić, N. Bertin and R.Gribonval. Audio declipping by cosparse hard thresholding. In *iTwist-2nd international-Traveling Workshop on Interactions between Sparse models and Technology*, 2014.

Contents

List of symbols

x	Real or complex scalar
x	Scalar integer
$\mathbf{x}$	Vector
$\mathbf{X}$	Matrix
ξ	Scalar constant
x	Functional
X	Linear operator
X	Random variable
$\mathfrak{X}$	Nonlinear operator
$\Re$	Real part of the complex number (linear operator)
$\Im$	Imaginary part of the complex number (linear operator)
Ξ	Set
χ_Ξ	Characteristic function of a set Ξ
$\boldsymbol{\xi}$	Tuple
$\mathscr{X}$	Vector space
$\mathbb{R}^\mathrm{n}$	n-dimensional real vector space ($\mathbb{R}^1 = \mathbb{R}$)
$\mathbb{C}^\mathrm{n}$	n-dimensional complex vector space ($\mathbb{C}^1 = \mathbb{C}$)
$\mathbb{N}_0$	Set of nonnegative integers
$L^\mathrm{p}(\Xi)$	Space of Lebesgue p-integrable functions on Ξ
$\mathbb{E}$	Expected value
∇	Gradient
$\mathtt{div}, \nabla\cdot$	Divergence
$\triangle$	Laplacian
$\Box$	D'Alembertian (wave operator)
O	Asymptotic behavior of a function ("big O" notation)
$\mathbf{X}_\Xi$	The restriction of a matrix $\mathbf{X}$ to rows or columns indexed by Ξ

Manuscript conventions

$\mathtt{n}$	Signal dimension
$\mathtt{k}$	Sparsity level or the number of sources
$\mathtt{m}$	Number of measurements or the number of sensors
$\mathtt{s}$	Dimension of discretized space
$\mathtt{t}$	Dimension of discretized time
Γ	Continuous or discrete spatial domain
Ω	Continuous or discrete spatiotemporal domain ($\Gamma \times [\mathrm{t}_1, \mathrm{t}_2]$)
$\mathfrak{M}, M, \mathbf{M}$	Measurement (observation) operators
$y, \mathbf{y}$	Measurements (observations)
$\mathbf{I}$	Identity matrix
$\mathbf{A}$	The analysis operator
$\mathbf{D}$	The synthesis dictionary
P_Ξ	Orthogonal projection to a set Ξ
$\mathbf{r}$	Vector of spatial coordinates
t	Temporal coordinate
$\mathbf{n}$	Outward normal vector
$\mathbf{v}$	Usually a prototype vector
c	Wave propagation speed
ω	Angular frequency
$\boldsymbol{\omega}$	An element of a set Ω
`nnz`	Number of non-zero elements
`null`($\mathbf{V}$)	The null space of the matrix $\mathbf{V}$
`diag`($\mathbf{v}$)	Diagonal matrix with the elements of $\mathbf{v}$ on the main diagonal

All symbols inconsistent with the notation are disambiguated in the text.

List of Figures

1 Introduction

The history of the human race is a tale riddled with examples of how our inherent curiosity and the need to discover and understand can be hindered by the limitations of our perception and technology. "*What is the shape of the Earth?*", for instance, is one fundamental question to which various cultures provided different answers that were often quite creative yet unequivocally wrong, with some persisting until as late as the 17th century. The most common misconception was "the flat Earth" model, where Earth was thought of as a disk or a square, for some even magically supported on the back of a giant turtle [176]. The first ones to come close to reality were the Pythagoreans, who proposed a spherical model of the Earth around 6th century BC. This model was later elaborated on by Aristotle, who offered in its favor the argument that the shadow that the Earth casts onto the Moon during lunar eclipses is round [116]. Today, the generally accepted model is a refinement of the Pythagorean one, and Earth is modeled as an oblate ellipsoid.

Why have we taken this brief stroll through history? Our aim was to illustrate how, in the absence of direct observation, philosophers and scientists are forced to rely on indirect observation to develop models that fit reality. Such problems are referred to as *inverse problems* in the world of science, and by offering a solution for any of them, we try to shed new light on the inner workings of Nature.

In an effort to achieve this goal, numerous scientists have provided their contributions throughout history. The result of this joint endeavor is a collection of mathematical models through which the observed physical manifestations are explained. Needless to say, all of these models are wrong [38] (or, to put it another way, not perfectly accurate), but some of them can be highly useful. Often, we are incapable of directly observing certain phenomena, and the model that we have remains our best *guess*. One illustrative example comes from astronomy, where we determine the chemical structure of a star based on the observations of its light emissions. There, the objects of interest are usually (extremely) far and out of our reach, and the bigger part of our knowledge base comes from models based on scarce observations.

The quality of such models naturally varies with the amount and quality of the information

that is available — our use of the word "guess" in the previous paragraph was hardly accidental. In many cases, an inverse problem, if tackled from a single standpoint, is ambiguous and ill-posed. Returning to the shape of the Earth — If we only take the point of view of a single individual standing on the surface, one understands how Earth can appear to be flat. Also, the "lunar shadow argument" by Aristotle is insufficient on its own as proof that Earth is not flat or is round. Indeed, if the Earth and the Sun were static relative to each other, a flat disc could project a shadow of the same shape. Therefore, the problem of describing the shape of the Earth based solely on the shape of its shadow on the Moon is ill-posed. Whenever we have an ill-posed inverse problem, our goal is to make it well-posed or regularize it, by adding extra information until we have unambiguous confirmation of the model.

Before going further with more formal discussion on inverse problems in signal processing, we temporary narrow the scope and give a flavor of what is the subject of this thesis. We are interested in so-called "*inverse source problems*", which lay on the blurry boundary between physics and signal processing. These problems are usually severely ill-posed and to address them we exploit particular regularization methods discussed later in this Chapter. On the practical side, we are primarily interested in the inverse source problems in acoustics and electroencephalography, which we informally describe in the following text.

Acoustics is the scientific discipline that investigates propagation of sound. Sound can be interpreted as the manifestation of mechanical vibration of fluid, such as air or water, which is called a propagating medium. Sound is not necessarily audible, *i.e.* it is not always possible to register it by the human auditory system. For instance, whales are using infrasound to communicate, and the ultrasound is used in sonography to visualize internal structure of human body. Both of these are usually not perceivable by humans due to physical limitations of our auditory system in terms of *frequency*. Loosely speaking, sound frequency is the rate of oscillations of particles in the propagating fluid. What distinguishes sound from other types of propagations, such as electromagnetic waves, is its mechanical nature (thus it cannot exist in vacuum) and the frequency range. A *sound source* is a vibrating region of space that produces sound, such as a guitar string. Imagine two guitarists playing in a small room. After a guitar string has been hit, depending on the *speed of sound*, the vibrations amplified by the soundboard will eventually propagate everywhere in the room (which can be justified by the fact that we would be able to hear it wherever we stand in the room). If we had microphones, placed in different regions of the room, they would all record sound which is a "mixture" of music, played by both guitarists. In the acoustic inverse source problem we are interested in characterizing "guitarists" on the basis of the microphone recording. By characterization, we mean inferring their locations within the room, the content of their plays, and the time instants when they start and stop playing. Replace "guitarists" by any sound sources, and this is an inverse acoustic source problem in an enclosed environment. In this thesis, we are primarily interested in the first property, namely, the locations of the sound sources. Sound (acoustic) source localization has many uses, starting from robot navigation, audio and speech enhancement, aeroacoustic noise suppression, ocean acoustics etc. More details will be given in Chapter 4.

Electroencephalography (EEG) is concerned with measuring electrical activity of brain sources, usually in a non-invasive way. It is by now widely known that neurons in the brain produce some amount of electrical activity, which can be passively measured by placing electrodes along human scalp. By investigating electrode recordings, one can infer information useful for diagnosing epileptic seizures, tumors, stroke or brain death in comatose patients. Nowadays, it is mostly used for diagnosing epileptic sources in the brain, which is an inverse source problem. This is an important information for the physician, as he may decide to surgically remove the part of the brain responsible for the seizures. The issues related to solving this inverse problem are presented in Chapter 6.

1.1 Inverse problems

In this section we informally and formally define inverse problems, with an emphasis on inverse problems in signal processing. However, the demarcation line between scientific disciplines is not clear, as we will see in what follows.

The significance of inverse problems in science and engineering cannot be overstressed. In signal processing, they are omnipresent, and examples are numerous: source localization [62, 243, 45], radar imaging [210, 18, 9], image processing [112, 24, 95], acoustic imaging and tomography [182, 79], medical imaging [262, 233, 203], compressed sensing [50, 96, 106], tracking [174, 252, 3], denoising [93, 19], declipping [227, 75, 144] etc, to cite only a few. Their generality is of such a wide scope that one may even argue that solving inverse problems is what signal processing is all about. Albeit being true, this is not a particularly useful generalization, quite similar to "everything is an optimization problem" [254] paradigm. What matters is identifying which inverse problems are solvable and finding a means to solve them.

Generally, two problems are considered being inverse to each other if the formulation of one problem requires the solution of the other [97]. A traditional way to define an inverse problem is to see it as the inverse of a *forward (direct)* problem. Despite the lack of scientific consensus on this type of categorization [138], in this thesis we will maintain the conventional wisdom and consider (informally) that forward problems start from the known input while inverse problems start from the known output [220]. Thus, the forward problem is usually the "easier" one, where certain conditions ("parameters") generate observable effects. In time-dependent systems, the forward problem usually respects the causality principle: the solution of the forward problem at time $t = t_1$ does not depend on the input at time $t = t_2$, when $t_1 < t_2$. Conversely, inverse problems are often non-causal and non-local, which makes them more challenging to solve.

More formally, the forward problem is often defined through a mathematical model sublimed in *the measurement operator* $\mathfrak{M}$ [16]. This operator maps the objects of interest x from the *parameter space* $\mathscr{X}$, to the space of observations or measurements $y \in \mathscr{Y}$:

$$y = \mathfrak{M}(x). \tag{1.1}$$

Then, the inverse problem is to recover (or estimate) the parameters x from the observed data y by means of a method that "reverts" $\mathfrak{M}$. We will assume that $\mathscr{X}$ and $\mathscr{Y}$ are vector spaces, sometimes equipped with an additional structure (*e.g.* norm and/or inner product). In the context of signal processing, we often call these quantities *signals*.

In practice, one often encounters *linear* inverse problems, characterized by the fact that the measurement operator $\mathfrak{M} := M$ is linear:

$$y = \mathfrak{M}(x) = Mx. \tag{1.2}$$

In the spirit of "all models are wrong" principle, one should replace the equality sign in (1.1) and (1.2) by the approximation ("$\approx$"). However, the meaning of "$\approx$" is more subtle, and needs to be specifically defined for a particular problem. Even if the observation model $\mathfrak{M}$ is perfect, in practice, the observation data y is inaccurate, due to imperfections in physical measurement systems. Often, this collection of model and measurement inaccuracies is labeled as "*noise*". The most common *model* for noisy observations is to consider an additive noise:

$$y = \mathfrak{M}(x) + e, \tag{1.3}$$

where e denotes the noise vector. This model implicitly assumes that x and e are statistically independent, which is not always true (*e.g.* when the noise is caused by quantization [117]). However, its convenient form makes it widely used in signal processing.

Concerning "solvability" of a particular problem, Jacques Hadamard introduced [121] (by now widely-accepted) conditions to determine if a problem is well-posed:

Existence: There exist a solution for the problem.

Uniqueness: The solution is unique.

Stability: The solution depends continuously on the data.

For the inverse problems defined previously, these conditions translate into properties of the measurement operator $\mathfrak{M}$. Assuming that the forward problem is well-posed and that $\{\forall\, y \in \mathscr{Y}, \exists x \in \mathscr{X} \mid y = \mathfrak{M}(x)\}$, to have a well-posed inverse problem the operator $\mathfrak{M}$ needs to be:

1. *Injective:* $\mathfrak{M}(x_1) = \mathfrak{M}(x_2) \Rightarrow x_1 = x_2$ and
2. *Stable:*[1] $x \to x^*$ is equivalent to $y \to y^*$, where $y = \mathfrak{M}(x)$ and $y^* = \mathfrak{M}(x^*)$.

Unfortunately, many interesting inverse problems are ill-posed in the sense of Hadamard. A rather famous example is the deconvolution problem, where direct inversion of the transfer

[1] By $u \to v$ we mean $d(u - v) \to 0$, where $d(\cdot)$ is a distance functional.

function results in instabilities at high frequencies [135]. Moreover, even if a problem is well-posed, it may be badly conditioned, which is another practical restriction. A simple example is the finite-dimensional linear system $\mathbf{y} = \mathbf{M}\mathbf{x}$ where $\mathbf{M}$ is invertible, but the *condition number* $\kappa(\mathbf{M}) = \|\mathbf{M}\|_2\|\mathbf{M}^{-1}\|_2$ is large, leading to numerical instabilities and erroneous results [111].

Tikhonov argued [236] that some of these problems may be solved by restricting the set of solutions to $\mathscr{C} \subset \mathscr{X}$, where $\mathscr{C}$ is termed *the correctness class*. For the solutions $x \in \mathscr{C}$, the measurement operator $\mathfrak{M}$ is indeed injective and stable, and the inverse problem is called *conditionally correct* according to Tikhonov [129]. Promoting solutions from the correctness class $\mathscr{C}$ is called regularization of the inverse problem and its success depends on the nature of $\mathscr{C}$ and the applied numerical method. In signal processing community, a correctness class is often called a *data model*.

1.2 Regularization

Regularization can be considered from different points of view. Roughly speaking, regularization approaches can be seen as *deterministic* (which are exploited in this work) or *stochastic*.

The most well-known deterministic approach is the *variational method*, to which subection 1.2.1 is devoted. Another, by now well-established approach, is the so-called *low-complexity regularization*, which is at the core of this thesis. We discuss the ideas behind this approach in subsection 1.2.2. Beside the variational method and low-complexity regularization, there are other deterministic approaches widely used in practice, such as the *truncated singular value decomposition* and *truncated iterative linear solvers*. These are usually applied to well-posed, but badly conditioned linear systems [135].

Overall, this section serves as an introductory review of deterministic regularization methods. However, stochastic regularization, or *statistical inversion methods*, are equally important. In a stochastic framework, all involved quantities are treated as random variables with an associated probability distribution. To apply regularization through the *Bayesian framework*, one builds a posterior probability distribution given the observed data used for computing a point estimate, such as the conditional mean or the posterior mode [135]. A potential difficulty with the former point estimate is the necessity to numerically integrate the posterior distribution. This is usually infeasible, and instead approximated by Markov Chain Monte Carlo methods [108] such as the *Gibbs sampler* [52]. Computing the posterior mode (*i.e. Maximum A Posteriori (MAP)* estimation) is sometimes more practical, and leads to an optimization problem that may be easier to solve. In some cases, the generated optimization problem coincides with a deterministic regularization problem, but one should be cautious when drawing conclusions concerning connections between the two [119].

1.2.1 Variational regularization

In many cases the term "regularization" is used interchangeably with variational regularization, which indicates the longevity of the variational method. The theory of variational regularization is rich and vast, based on variational calculus and convex analysis in Banach and Hilbert spaces. Therefore, in this introduction we do not limit the concept to finite dimensional spaces. Moreover, the problems we are going to tackle are genuinely continuous, although, in practice, we will handle them in a finite-dimensional manner. Hence, in the context of variational regularization, we consider the signal of interest to be a continuous *functional* $x = x(\boldsymbol{\omega})\colon \Omega \mapsto \mathbb{C}$ defined over some domain Ω (*e.g.* $\mathbb{R}^{\mathsf{n}}$). Furthermore, we assume that x belongs to $H^{\mathsf{d}}(\Omega)$ Sobolev space, *i.e.* the space of square-integrable functions whose first d weak derivatives[2] are also in $L^2(\Omega)$.

Intuitively, the idea behind the variational approach is to yield an estimate by minimizing an "energy" represented by a sum of functionals. These are chosen such that they promote solutions from the correctness class $\mathscr{C}$. Particularly, an inverse problem consisting in estimation of x from the observed data y is formulated as an optimization problem of the following form:

$$\hat{x} = \arg\inf_{x} f_d\left(y, \mathfrak{M}(x)\right) + f_r\left(x\right), \tag{1.4}$$

where f_d is known as a *data fidelity* or *discrepancy* term and f_r is a *regularizer*. The role of the data fidelity functional is to ensure that the estimate $\hat{x}$ in some sense complies with the observed data y, while the f_r penalty actually embodies regularization. Informally, f_r is also called a "*prior*", indicating that it arose from a prior knowledge we have about the estimate. The choice of penalties f_r and f_d is dictated by the particular problem at hand.

Certainly, the most common choice of the data fidelity is the quadratic penalty $f_d = f_q^{\cdot}$, such as the well-known squared difference[3]:

$$f_q^{\mathscr{Y}}\left(y - \mathfrak{M}(x)\right) = \int_{\Omega} \left(y - \mathfrak{M}(x)\right)^2 d\omega = \|y - \mathfrak{M}(x)\|_{\mathscr{Y}}^2, \tag{1.5}$$

where the norm $\|\cdot\|_{\mathscr{Y}}$ is the usual norm associated with an inner product $\langle\cdot,\cdot\rangle_{\mathscr{Y}}$. The penalty $f_q^{\mathscr{Y}}$ puts large weight on the large components of the residual $r = y - \mathfrak{M}(x)$ and vice-versa, a small weight on the small components. Effectively, it promotes solutions $\hat{x}$ such that the residual r contains a large amount of low-magnitude components.

If both f_d and f_r are given in the form of $f_q^{\cdot}$, with $f_d = f_q^{\mathscr{Y}}\left(y - \mathfrak{M}(x)\right)$ and $f_r = \lambda f_q^{\mathscr{V}}(Lx)$, the regularization approach is known as *(generalized) Tikhonov regularization* [222]:

$$\inf_{x} f_d\left(y, \mathfrak{M}(x)\right) + f_r(x) = \inf_{x} \|y - \mathfrak{M}(x)\|_{\mathscr{Y}}^2 + \lambda \|Lx\|_{\mathscr{V}}^2, \tag{1.6}$$

[2] The generalization of the notion of derivative: $\mathbf{d}(\boldsymbol{\omega})$ is the d^{th} weak derivative of $\mathbf{x}(\boldsymbol{\omega})$ if and only if $\forall \mathbf{u}(\boldsymbol{\omega}) \in C^{\infty}(\Omega)\colon \int_{\Omega} \mathbf{x}(\boldsymbol{\omega}) \partial^{\mathsf{d}} \mathbf{u}(\boldsymbol{\omega}) = (-1)^{|\mathsf{d}|} \int_{\Omega} \mathbf{d}(\boldsymbol{\omega}) \mathbf{u}(\boldsymbol{\omega})$.

[3] Assuming the integral exist and is finite.

with $L\colon \mathscr{X} \mapsto \mathscr{V}$ a bounded linear mapping[4]. In practice, L is often a differential operator which promotes some level of smoothness on the estimated solution $\hat{x}$ (*e.g.* L is the gradient operator ∇). In this case, $\mathscr{C}$ is the subset of solutions $\{x \mid y \approx \mathfrak{M}(x)\}$ such that $\hat{x} \in \mathscr{C}$ is smooth in the corresponding H^s sense. If, for instance, $L = I$ (where I is the identity operator), Tikhonov regularization encourages estimates from the correctness class of "minimal L^2-norm" [14].

Tikhonov regularization is widely used nowadays, since smoothness of the objective is attractive from a numerical optimization standpoint. Namely, if $\mathfrak{M} = \mathbf{M}$ and $L = \mathbf{L}$ are finite-dimensional linear mappings, (1.6) is given as:

$$\inf_x f_d(y, \mathfrak{M}(x)) + f_r(x) \Leftrightarrow \underset{\mathbf{x}}{\text{minimize}} \ \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + \lambda \|\mathbf{L}\mathbf{x}\|_2^2, \tag{1.7}$$

which admits a closed-form solution

$$\hat{\mathbf{x}} = (\mathbf{M}^{\mathsf{H}}\mathbf{M} + \lambda \mathbf{L}^{\mathsf{H}}\mathbf{L})^{-1}\mathbf{M}^{\mathsf{H}}\mathbf{y}.$$

Note that, for general non-linear $\mathfrak{M}$, the optimization problem may be non-convex and thus, difficult to solve exactly.

While Tikhonov regularization leads to an unconstrained optimization problem, the two related regularizations are based on constraints embedded in either f_d or f_r. *Morozov regularization* [135] is defined as follows:

$$\inf_x \|x\|_{\mathscr{V}}^2 \text{ subject to } \|y - \mathfrak{M}(x)\|_{\mathscr{Y}} \leq \varepsilon, \tag{1.8}$$

where ε can be interpreted as the standard deviation of the noise in the measurement data. *Ivanov regularization* [180], on the other hand, bounds the regularizing norm:

$$\inf_x \|y - \mathfrak{M}(x)\|_{\mathscr{Y}}^2 \text{ subject to } \|x\|_{\mathscr{V}} \leq \tau. \tag{1.9}$$

The two latter types of regularization do not have a closed-form solution even in the linear case. However, they can be expressed as an unconstrained Tikhonov regularization (1.6) by an appropriate choice of the parameter λ [35].

Although very useful, minimization of the squared inner product norm is, by no means, the only available regularization method. Another useful regularizer is the L^1 norm[5]. In the linear setting, with a usual quadratic discrepancy term, it corresponds to the following optimization problem:

$$\inf_x f_d(y, \mathfrak{M}(x)) + f_r(x) = \inf_x \|y - Mx\|_{\mathscr{Y}}^2 + \lambda \int_\Omega |Lx| d\omega. \tag{1.10}$$

A very common choice for the operator L is, again, $L = \nabla$. This type of regularization is known as *total variation* minimization [222], and it, intuitively, favors piecewise constant estimates $\hat{x}$.

[4] Same as before, $\|\cdot\|_{\mathscr{V}}$ is the usual norm associated with the inner product $\langle \cdot, \cdot \rangle_{\mathscr{V}}$.

[5] We implicitly assumed that x and the associated mappings now also belong to L^1 space.

There are many other variational priors beyond those mentioned, but these are out of the scope of this thesis. Lastly, we emphasize that all optimizations are necessarily performed in finite dimensions. Therefore, *discretization* of continuous problems plays an important role.

1.2.2 Low complexity regularization

The idea of "simple" models is not new, but after being rediscovered in the beginning of 1990s, it somewhat revolutionized modern signal processing. The idea behind this type of regularization is inspired by the *Occam's razor* principle: "*Numquam ponenda est pluralitas sine necessitate*" (*Plurality must never be posited without necessity*). This can be interpreted as "the underlying model should not be more complicated than necessary to accurately represent the observations". In the context of prediction theory, Solomonoff [228] mathematically formalized this principle and shown that shorter theories do have larger weight in computing the probability of next observation.

Nevertheless, in regularization, Occam's razor emerges from the empirical observation that many natural signals can be approximated by a "simple" *representation*, although their "standard" or "altered" (*e.g.* noisy) representation may be complex. The notion of what do we mean by "simple" is important. Here, we identify with "simple" a signal whose *intrinsic* dimension is much smaller that *the ambient space*. The ambient space is the space "surrounding" a mathematical object, while the intrinsic dimension is the "true" number of degrees of freedom of an object (for instance, a plane in 3D ambient space has an intrinsic dimension two).

Implicitly, this fact is the foundation of compression, which exploits redundancy of "standard" representations to reduce the amount of data necessary to archive or transmit a signal. Some examples are transform coding schemes embedded in MP3, MPEG, JPEG and JPEG2000 standards. Beyond compression, applications such as in radar [125, 18], seismic [77] and medical imaging [241, 164], telecommunications [198, 207], computer vision [259, 53] and genomics [201], confirm the "simple" or "low complexity" intuition in practice.

In this subsection we restrict the discussion to finite dimensional spaces[6], as the theory of infinite dimensional "low complexity" regularization is not unified and, in many segments, not fully mature yet (although there are some notable works [123, 208]). Still, even with this constraint, one can imagine different types of simplicity in signals, such as (among many):

k-sparse signals For a signal $\mathbf{x} \in \mathbb{R}^n$ with $k < n$ non-zero elements, the ambient dimension is n, and the intrinsic dimension is k [33].

l-cosparse signals Given a matrix $\mathbf{A} \in \mathbb{R}^{p \times n}$, a signal $\mathbf{x} \in \mathbb{R}^n$ is l-cosparse if the product $\mathbf{Ax}$ contains only $p - l$ non-zero components; the ambient dimension is again n, but the intrinsic dimension is typically $n - l$ [187].

[6] $\mathbb{R}^n$, $\mathbb{C}^n$ and, technically, any finite dimensional Hilbert space, since these are isomorphic to $\mathbb{R}^n$, $\mathbb{C}^n$.

Rank-r matrices A matrix $\mathbf{R} \in \mathbb{R}^{n_1 \times n_2}$ of rank r has the ambient dimension $n_1 n_2$ and the intrinsic dimension $r(n_1 + n_2) - r^2$ [49].

Hermitian Toeplitz matrices For a Hermetian Toeplitz matrix of size $n \times n$, the ambient dimension is n^2, but the intrinsic dimension is $2n - 1$ [212].

The first two signal models are widely applicable concepts in many settings: for instance, previously mentioned compression coding schemes (implicitly) rely on the sparse / cosparse prior. Low rank matrix models are used, *e.g.* in recommender systems [146], while the covariance matrices of wide-sense stationary signals have Hermitian Toeplitz structure [212]. Furthermore, signals may have a structure expressed by an "interSection of models", *i.e.* they exhibit properties that can be characterized by several models ([162, 12, 246], for example).

The sparse synthesis data model

The sparse synthesis data model, or simply ***sparsity***, has been an active area of research for more than two decades. Impressive results were obtained in many applications, with compressive sensing [51, 96, 106] being the flagship of the field.

It is illustrative to introduce the model using the variational form given in (1.4) (bearing in mind that it should not be confused with variational regularization, which is elaborated later in the text). Hence, sparse synthesis regularization may be seen as an optimization problem involving the ℓ_0 "norm" as a regularizer:

$$\underset{\mathbf{x}}{\text{minimize}}\ f_d\left(\mathbf{y}, \mathfrak{M}(\mathbf{x})\right) + f_r(\mathbf{x}) = \underset{\mathbf{x}}{\text{minimize}}\ f_d\left(\mathbf{y}, \mathfrak{M}(\mathbf{x})\right) + \|\mathbf{x}\|_0. \tag{1.11}$$

Here $f_r(\mathbf{x}) = \|\mathbf{x}\|_0$ is the count of non-zero elements in $\mathbf{x} \in \mathbb{R}^d$, which is not absolutely homogeneous ($\|\lambda\mathbf{x}\|_0 \neq |\lambda|\|\mathbf{x}\|_0$), and therefore, not a true norm. An alternative is *the characteristic function*[7] $\chi_{\ell_0 \leq k}(\mathbf{x})$, which imposes the constraint that the estimated vector $\mathbf{x}$ cannot have more than k non-zero components. In any case, the assumption is that $\|\mathbf{x}\|_0 \ll d$, where d is the ambient dimension.

A common extension is the model in which the signal is sparse in a *dictionary*, expressed as follows:.

$$\underset{\mathbf{x}}{\text{minimize}}\ f_d\left(\mathbf{y}, \mathfrak{M}(\mathbf{D}\mathbf{x})\right) + \|\mathbf{x}\|_0, \tag{1.12}$$

where the dictionary matrix $\mathbf{D} \in \mathbb{C}^{n \times d}$ can be *overcomplete* ($n < d$), making it more general than basis. Its columns are often called *atoms*. The set Φ of indices of non-zero components in a vector $\mathbf{x}$ is termed *support* of $\mathbf{x}$. We will denote by $\mathbf{D}_\Phi$ the restriction of the matrix $\mathbf{D}$ to the *columns* of the dictionary corresponding to support.

Unfortunately, sparse synthesis regularization leads to NP-hard (combinatorial-type) prob-

[7] Formally defined in appendix A.

lems cf. [105]. Moreover, the ℓ_0 "norm" is discontinuous, and, from a numerical point of view, very unstable: small perturbations of the signal may have a large impact on the results. This is also a reason why one should not confuse (conceptual) sparse and cosparse regularization to a variational approach, which requires a well-posed minimization problem [165]. On the other hand, in many cases *convex relaxations* and *greedy algorithms* are very effective in recovering sparse signals.

Convex relaxations substitute ℓ_0 by a convex penalty, in which case (co)sparse regularization coincides with variational methods, discussed in the previous subsection. Their principal advantage is that convexity ensures that the attained minimum is indeed global (which does not necessarily mean that the estimate is unique - this is reserved for *strictly convex* functions [40]). Relating convex penalties to sparse signal recovery is somewhat technical [54], but the common intuition is that the ℓ_1 norm is known to promote generally sparse solutions. If a signal has additional structure, such as group sparsity[8], this can be also promoted by appropriate choice of a *group norm* [14, 132]. The goal of greedy algorithms is to recover the support of a signal, and therefore the signal itself, by an iterative estimation procedure. There are many variants, such as *Matching Pursuit* [173], *Orthogonal Matching Pursuit (OMP)* [204, 189], *Iterative Hard Thresholding (IHT)* [32] or *Compressive Sampling Matching Pursuit (CoSaMP)* [191]. These algorithms can be seen as non-convex heuristics, but they come with certain performance guarantees.

Theoretical aspects of sparse signal recovery has been exhaustively researched. The most referenced concept is the *Restricted Isometry Property (RIP)* or *uniform uncertainty principle* [51, 106]. In linear inverse problems ($\mathfrak{M} = \mathbf{M}$), for the class of signals sparse in a dictionary $\mathbf{D}$, it characterizes the *sensing matrix* $\mathbf{MD}$ with the so-called *restricted isometry constant*. It is defined as the smallest constant $\delta_k > 0$ such that the following holds:

$$(1-\delta_k)\|\mathbf{x}\|_2^2 \leq \|\mathbf{MDx}\|_2^2 \leq (1+\delta_k)\|\mathbf{x}\|_2^2, \quad \forall \{\mathbf{x} \mid \|\mathbf{x}\|_0 \leq k\}. \tag{1.13}$$

For some applications and classes of sensing matrices, recovery conditions based on the restricted isometry constants have been established. For instance, a famous result in compressed sensing states that, with *e.g.* Gaussian or Bernoulli sensing matrices [106], one needs only $m \propto k \ln \frac{n}{k}$ measurements to obtain robust signal recovery. Moreover, this has been accomplished not only for convex decoders[9], but also for several greedy algorithms (including the ones mentioned above). However, evaluating δ_k in (1.13), for general matrices, is also NP-hard [238]. In cases where the restricted isometry constant cannot be calculated, (suboptimal) recovery results based on *coherence* [106] may be attainable.

A recent research direction is the generalization of the RIP concept to more general decoders, or even to different signal models (*e.g.* [37]).

[8] We will evoke some group sparse structures in subSection 4.3.4 of Chapter 4.

[9] *Decoder* is a means of estimating $\mathbf{x}$ from the measurements $\mathbf{y}$.

The sparse analysis data model

The sparse analysis data model or ***cosparsity*** [187], has only recently attracted as much attention of the scientific community as the synthesis model. It differs from sparsity due to the presence of a matrix $\mathbf{A}$, termed *the analysis operator* :

$$\underset{\mathbf{x}}{\text{minimize}}\ f_d\left(\mathbf{y}, \mathfrak{M}(\mathbf{x})\right) + \|\mathbf{A}\mathbf{x}\|_0. \tag{1.14}$$

The analysis operator $\mathbf{A} \in \mathbb{R}^{p \times n}$ can be overcomplete, in the sense that it contains more rows than columns ($p > n$). The index set Λ of rows of $\mathbf{A}$ orthogonal to a vector $\mathbf{x}$ is termed *cosupport* of $\mathbf{x}$. The restriction of the matrix $\mathbf{A}$ to the set of *rows* referenced by cosupport is denoted $\mathbf{A}_\Lambda$. One of the most common use cases of cosparsity is the well-known total variation regularization (1.10), for which the matrix $\mathbf{A}$ is an approximation of the gradient operator ∇.

One could see the cosparse model as a generalization of the synthesis model (which is recovered by $\mathbf{A} = \mathbf{I}$), but a more appropriate view is that sparse and cosparse signals belong to different *unions of subspaces* [166, 33]. While k-sparse signals are members of the union of all k-dimensional subspaces spanned by columns of $\mathbf{D}_\Phi$ ($|\Phi| = k$), l-cosparse signals belong to the union of all $(n-l)$-dimensional subspaces spanned by columns of $\texttt{null}(\mathbf{A}_\Lambda)$ ($|\Lambda| = l$). In fact, there is only one particular case where the two models are equivalent, and that is when $\mathbf{D} = \mathbf{A}^{-1}$, which requires the two matrices to be square-invertible [94]. In many practical scenarios, however, the dictionary and the analysis operator are overcomplete, rendering the two models different. For the cosparse model, we assume that $\|\mathbf{A}\mathbf{x}\|_0 \ll p$. If the matrix $\mathbf{A}$ is of full column rank, there is an obvious limit to the number of possible rows orthogonal to $\mathbf{x}$ (the product $\mathbf{A}\mathbf{x}$ is usually "less sparse" compared to the synthesis case). However, it should not be interpreted as a weakness of this data model.

We know that ℓ_0 minimization is computationally intractable. As in the synthesis setting, convex relaxations and greedy methods have been proposed to approximate the solutions of (1.14). Assuming no additional structure in $\mathbf{A}\mathbf{x}$, one would again use the ℓ_1 norm $\|\mathbf{A}\mathbf{x}\|_1$ as a convex relaxation [187, 242]. Analogously, if the structure is available, more appropriate objectives can be used. Concerning greedy methods, counterparts of the most common synthesis-based algorithms have been proposed in the analysis context. Some of these are *Greedy Analysis Pursuit (GAP)* [186, 187], *Analysis Iterative Hard Thresholding (Analysis IHT)* and *Analysis Compressive Matching Pursuit (Analysis CoSaMP)* [109].

The necessary condition to have any hope of cosparse signal recovery is that the null spaces of $\mathbf{A}$ and $\mathbf{M}$ intersect only trivially:

$$\texttt{null}\left(\begin{bmatrix}\mathbf{A}\\ \mathbf{M}\end{bmatrix}\right) = \{\mathbf{0}\}. \tag{1.15}$$

Otherwise, obviously, uniqueness cannot be assured.

An adaptation of the RIP to the sparse analysis data model, known as D-RIP [46, 133], is the

following:

$$(1-\delta_l)\|\mathbf{x}\|_2^2 \leq \|\mathbf{M}\mathbf{x}\|_2^2 \leq (1+\delta_l)\|\mathbf{x}\|_2^2, \quad \forall \left\{\mathbf{x}=\mathbf{A}^\mathsf{T}\mathbf{z} \mid \|\mathbf{z}\|_0 \leq l\right\}, \tag{1.16}$$

i.e., it should hold for all l-cosparse vectors $\mathbf{x}$. For $\mathbf{A}=\mathbf{I}$ and $\mathbf{M}:=\mathbf{MD}$, the RIP condition for sparse-in-a-dictionary signals (1.13) is recovered. However, as argued in [187], the D-RIP tends to keep the sparse synthesis perspective, and the results derived in this spirit do not reflect the true ability of sparse analysis regularization to recover cosparse signals. Different conditions, such as the ones derived using the so-called *Null Space Property* [242, 133, 187] are more applicable.

1.3 Thesis guideline

The main objective of this thesis is to gain a profound understanding of potentials and drawbacks of the sparse synthesis and sparse analysis regularizations in the context of so-called *physics-driven* signal processing inverse problems. This class of problems is inspired by certain physical laws and characterized by an a priori knowledge that the involved signals can be modeled as sparse or cosparse. The "dual" nature of the problems, poses the following questions:

1. Which of the two data models is more appropriate?
2. How to choose a synthesis dictionary / an analysis operator for the given problem?
3. How to efficiently solve the regularized inverse problem?

Our goal in this work is to shed some light on these fundamental issues.

The thesis is organized as follows:

- Chapter 1 was the thesis introduction.
- Chapter 2 is a prelude to physics-driven (co)sparse regularization. It is concerned with the audio declipping inverse problem regularized by sparsity and cosparsity, highlighting the qualitative and numerical differences of the two priors in this context.
- In Chapter 3 the physics-driven (co)sparse framework is introduced. Computational complexity of using either of the two approaches is discussed.
- Acoustic source localization addressed by physics-driven (co)sparse regularization is discussed in Chapter 4. We review state-of-the-art and provide comprehensive simulation results with both synthesis and analysis approaches.
- The ability of the cosparse localization approach to handle difficult scenarios –where physical parameters are unknown or the environment is adverse–, is presented in Chapter 5.

- The brain source localization problem in EEG is challenged by means of the physics-driven (co)sparse framework in Chapter 6. The approach is confronted to several state-of-the-art methods.
- Conclusions, final remarks and perspectives are discussed in Chapter 7.
- Appendices A, B, C and D provide background material occasionally recalled in the thesis.

Except for the Introduction, Conclusions and Appendices, every chapter begins with a statement of the main topic, a note on the related publications resulted from this work, and an introduction of the chapter structure. Chapters end with a summary of the presented material and a short remark concerning our contributions.

2 Audio inpainting

In the first chapter we argued that, although perhaps deceivingly similar, the sparse synthesis and sparse analysis data models are fundamentally different. There is only one *modus operandi* where the two models become nominally equivalent, and that is when the dictionary and the analysis operator are inverse of each other. The reminder of the thesis will revolve around this special case, due to the particular class of problems we are interested in (these are elaborated in chapter 3). The goal of this chapter is to give a taste of practical implications of using the two models in non-equivalent settings.

We discuss the so-called inverse problem of *audio inpainting*, which is a term that first appeared in [157, 2]. The purpose was to highlight similarities with well-known *image inpainting* problem [24, 68, 95], concerned with partial recovery or object removal in images. Analogously, in audio inpainting one is interested in recovering a part of audio signal, which is either missing or degraded, using the surrounding "*reliable*" (undistorted) audio data. The problem is very general, and includes cases such as packet loss in Voice-over-IP networks [205], impulsive noise ("*clicks*"), scratches and breakages in the recording medium [110], *clipping* [134], audio bandwidth extension [155] etc (an exhaustive list of references is available in [2]). The scope of this work is inpainting the audio signals degraded by clipping, or *magnitude saturation*. We first formally define the goal:

Given the saturated single channel recording, estimate the original audio signal.

This is a well-known problem in signal processing, arising not only in audio, but also in image processing [13, 183] and digital communications [161]. Still, the focus is on single channel audio, although the main principles and algorithms may be generalized.

The chapter proceeds as follows: in the first section, we introduce the inverse problem of audio declipping, which is followed, in the second section, by the review of state-of-the-art approaches in the field. The third section proposes new sparse-cosparse algorithmic framework for audio declipping. In the fourth section we provide numerical and perceptual performance evaluation. The material in this chapter is based on publications [143, 141].

2.1 The inverse problem of audio declipping

Audio signals become saturated usually during acquisition, reproduction or A/D (Analogue-to-Digital) conversion. The perceptual manifestation of clipped audio depends on the level of clipping degradation and the audio content. In case of mild to moderate clipping, the listener may notice occasional "clicks and pops" during playback. When clipping becomes severe, the audio content is usually perceived as if it was contaminated with a high level of additive noise, which may be explained by the introduction of a large number of harmonics caused by the discontinuities in the degraded signal. In addition to audible artifacts, some recent studies have shown that clipping has a negative impact on Automatic Speech Recognition [234, 124] and source separation [27] performance.

In the following text, a sampled audio signal is represented by the vector $\mathbf{x} \in \mathbb{R}^n$ and its clipped version is denoted by $\mathbf{y} \in \mathbb{R}^n$. The latter can be easily deduced from $\mathbf{x}$ through the following nonlinear measurement model, called *hard clipping*:

$$\mathbf{y}_i = \mathfrak{M}(\mathbf{x})_i = \begin{cases} \mathbf{x}_i & \text{for } |\mathbf{x}_i| \leq \tau, \\ \operatorname{sgn}(\mathbf{x}_i)\tau & \text{otherwise}^1. \end{cases} \tag{2.1}$$

While idealized, this clipping model is a convenient approximation allowing to clearly distinguish the clipped parts of a signal by identifying the samples having the highest absolute magnitude. Indices corresponding to "reliable" samples of $\mathbf{y}$ (not affected by clipping) are indexed by Ω_r, while Ω_c^+ and Ω_c^- index the clipped samples with positive and negative magnitude, respectively. An illustrative example of a clipped sinusoidal signal is given in figure 2.1a.

Our goal is to estimate the original signal $\mathbf{x}$ from its clipped version $\mathbf{y}$, *i.e.* to "declip" the signal $\mathbf{y}$. Ideally, the estimated signal $\hat{\mathbf{x}}$ should satisfy natural magnitude constraints in order to be consistent with the clipped measurements. Thus, we seek an estimate $\hat{\mathbf{x}}$ which meets the following criteria:

$$\hat{\mathbf{x}} \in \Xi = \left\{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{M}_r \mathbf{x} = \mathbf{M}_r \mathbf{y}, \quad \mathbf{M}_c^+ \mathbf{x} \geq \mathbf{M}_c^+ \mathbf{y}, \quad \mathbf{M}_c^- \mathbf{x} \leq \mathbf{M}_c^- \mathbf{y} \right\}, \tag{2.2}$$

where the matrices $\mathbf{M}_r$, $\mathbf{M}_c^-$ and $\mathbf{M}_c^+$ are *restriction operators*. These are simply row-reduced identity matrices used to extract the vector elements indexed by the sets Ω_r, Ω_c^+ and Ω_c^-, respectively. We write the constraints (2.2) as $\hat{\mathbf{x}} \in \Xi$.

Obviously, consistency alone is not sufficient to ensure uniqueness of $\hat{\mathbf{x}}$, thus one needs to further regularize the inverse problem. The declipping inverse problem is amenable to several regularization approaches proposed in the literature, such as based on linear prediction [131], minimization of the energy of high order derivatives [124], psychoacoustics [75], sparsity [2, 144, 227, 75, 256] and cosparsity [143, 141]. The latter two priors are based on the fact that the energy of audio signals is often concentrated either in a small number of frequency

[1] $\operatorname{sgn}(\cdot)$ is component-wise sign operator.

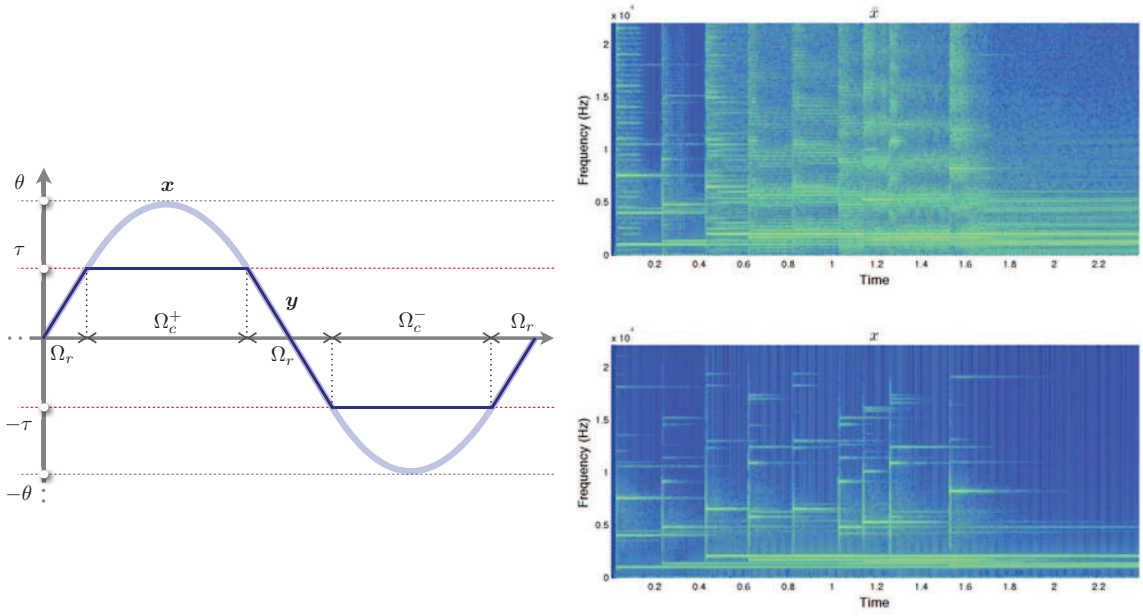


Figure 2.1 – Hard clipping example (left) and spectrograms of the clipped (top right) and the original (bottom right) audio signals.

components, or in short temporal bursts [206], *i.e.* they are (approximately) time-frequency sparse. This observation enables some state-of-the-art methods in clipping restoration (for illustration, figure 2.1b shows the spectrogram effects of audio clipping).

2.1.1 Theoretical considerations

From the theoretical perspective, the declipping inverse problem regularized by sparse representations has not been investigated. One of the reasons is that the measurement system violates principal assumptions of the compressed sensing theory. Namely, the sampling process is signal-dependent, and, thus, cannot be modeled in a standard way (*e.g.* by uniform distribution). Moreover, the class of signals Ξ imposes rather complex structure, beyond the ones commonly considered in compressive sensing (*i.e.* k-sparse or group sparse signals). In addition, there is some practical evidence [256, 75, 2] that standard convex relaxation methods underperform when compared to greedy declipping heuristics. This is another indicator that the declipping inverse problem is much different than compressed sensing, despite the fact that the underlying regularizers are the same.

For these reasons, all research results thus far (including ours) are of empirical nature. In practice, particularly in audio, these can actually be more relevant. The end users are interested in audible, rather than numerical improvements. On the other hand, when a declipping algorithm serves as a pre-processing block for some other application (*e.g.* a speech recognizer), the estimation accuracy may be more important.

2.2 Prior art

Since declipping is a special case of audio inpainting, any algorithm that addresses the latter, can be used for solving the former inverse problem. Historically, one of the first such "audio interpolation" approaches, due to Janssen et al. [131], is based on autoregressive (AR) modeling of audio signals. Autoregression is known to be a good model of human glottal tract [245], and therefore, good model for speech signals. However, the approach does not generalize well to other audio signals (*e.g.* music). Moreover, it is successful only for relatively mild clipping (usually corresponding to small "gaps" in audio), since it does not take into account the particularities of clipping.

Methods that specifically target the declipping inverse problem are relatively new. Most of them are based on some form of the sparse synthesis prior and time-frequency dictionaries. Adler et al. [2] proposed a two-stage declipping method based on the OMP algorihtm, termed *constrained Orthogonal Matching Pursuit*. In the first stage, the algorithm uses reliable part of the signal to estimate the support in transform domain. Then, in the second, refinement stage, the estimate is projected to the constraint set Ξ, while preserving only those atoms of the dictionary agreeing with the support set. Since the support estimation is performed without exploiting clipping constraints, the first stage of the algorithm is highly susceptible to errors. This has been demonstrated in [144], on both simulated and real audio data.

In the same paper, a new declipping algorithm based on *Iterative Hard Thresholding for Compressed Sensing* [32] was proposed. By introducing additional penalties in the objective, this algorithm, termed *Consistent IHT*, simultaneously enforces sparsity and clipping consistency of the estimate. At the same time, the low complexity feature of the original IHT algorithm is preserved. The algorithm iterates the following expression:

$$\mathbf{z}^{(i+1)} = \mathfrak{H}_{i+1}[\mathbf{z}^{(i)} + m^{(i)}\,\mathbf{D}^{\mathsf{T}}\mathfrak{B}(\mathbf{y} - \mathbf{D}\mathbf{z}^{(i)})], \tag{2.3}$$

where the operator $\mathfrak{H}_k(\mathbf{v})$ performs hard thresholding, *i.e.* sets all but k highest in magnitude components of $\mathbf{v}$ to zero, thereby encouraging sparsity. Since $k := i + 1$, the algorithm incrementally allows more non-zero components to be transferred between iterations, acting as a heuristic sparsity-learning strategy. The operator $\mathfrak{B}(\mathbf{v})$ is part of the negative gradient term, which incorporates the data fidelity and clipping consistency penalization:

$$\mathfrak{B}(\mathbf{v})_j = \begin{cases} v_j & j \in \Omega_r \\ (v_j)_+ & j \in \Omega_c^+ \\ (v_j)_- & j \in \Omega_c^-, \end{cases} \tag{2.4}$$

where $(\cdot)_+$ and $(\cdot)_-$ are positive- and negative- thresholding operators, respectively. The multiplier $m^{(i)}$ is computed by line-search. Both constrained OMP and Consistent IHT use block-based processing: the algorithms operate on individual blocks of audio data, which is subsequently resynthesized by means of the overlap-add scheme.

Recently, a new sparsity-based declipping algorithm was introduced by Siedenburg et al. [227]. The algorithm exploits a structured sparsity prior known as *social sparsity* [148], which encourages estimates whose time-frequency support is clustered ("neighborhood" sparsity). This improves estimation and declipping performance, but at the expense of higher computational cost. Namely, the algorithm requires batch processing of the entire audio signal, in order to be able to perform accurate time-frequency clustering. This limitation can become cumbersome in applications where real-time processing is required. More details and the pseudocode are provided in appendix C.

Another approach from the sparse synthesis family was proposed by Defraene et al. [75]. It is the only considered algorithm that uses convex optimization, although it was previously argued that, for the declipping problem, greedy heuristics seem to outperform convex relaxations. The distinct feature of this approach is that it uses perceptually weighted dictionary, and, as such, is better adapted to human auditory system. Unfortunately, due to the applied convex relaxation, it performs worse than plain (clipping-unaware) OMP algorithm in terms of the signal recovery. It is possible, however, that the performance can be improved by applying non-convex heuristics, such as reweighted ℓ_1 minimization proposed in [256].

Finally, one method that does not rely on the sparse synthesis model was proposed by Harvilla et al. in [124]. The approach is based on the assumption that the second order derivative of the estimated signal should vanish. This, in turn, produces a genuinely smooth estimate, which may be an acceptable approximation of the audio signal for high sampling rates. However, the approach is sensitive to noise, in which case it is outperformed by Consistent IHT.

2.3 The SPADE algorithms

Interestingly, none of the presented approaches considers the sparse analysis data model. While overseen before, cosparsity may be well-adapted for the declipping scenario, where the estimate is heavily constrained by Ξ in its native domain. Therefore, and given the current state-of-the-art, we set three goals:

1. Competitive declipping performance,
2. Computational efficiency,
3. Versatility: use sparse synthesis or sparse analysis prior on an equal footing.

We previously mentioned that, for the declipping inverse problem, the empirical evidence is not in favor of ℓ_1 convex relaxation. Hence, the goal is to build an algorithmic framework based on non-convex heuristics, that can be straightforwardly parametrized for use in both the synthesis and the analysis setting. To allow for possible real-time implementation, the algorithms need to support block-based processing.

Inspired by simplicity and computational efficiency of Consistent IHT, it would be interesting to exploit the same idea in the cosparse setting. Unfortunately, while the synthesis ℓ_0-projection corresponds to simple hard-thresholding, the cosparse ℓ_0-projection

$$P_\Sigma(\mathbf{v}) = \arg\min_{\mathbf{x}\in\Sigma} \|\mathbf{x}-\mathbf{v}\|_2, \quad \Sigma = \{\mathbf{x} \mid \|\mathbf{A}\mathbf{x}\|_0 = \mathsf{k}\}$$

is proven to be NP-hard [237]. Therefore, we take another route, and seek the estimates which are only *approximatelly* sparse or cosparse. The heuristics should approximate the solution of the following synthesis- and analysis-regularized inverse problems[2]:

$$\operatorname*{minimize}_{\mathbf{x},\mathbf{z}} \|\mathbf{z}\|_0 + \chi_\Xi(\mathbf{x}) + \chi_{\ell_2\le\varepsilon}(\mathbf{x}-\mathbf{D}\mathbf{z}) \tag{2.5}$$

$$\operatorname*{minimize}_{\mathbf{x},\mathbf{z}} \|\mathbf{z}\|_0 + \chi_\Xi(\mathbf{x}) + \chi_{\ell_2\le\varepsilon}(\mathbf{A}\mathbf{x}-\mathbf{z}). \tag{2.6}$$

The characteristic function χ_Ξ of the constraint set Ξ forces the estimate $\mathbf{x}$ to satisfy (2.2). The additional penalty $\chi_{\ell_2\le\varepsilon}$ is a *coupling* functional. Its role is to enable the end-user to explicitly bound the distance between the estimate and its sparse approximation. These are difficult optimization problems: besides inherited NP-hardness, the two problems are also non-convex and non-smooth.

We can represent (2.5) and (2.6) in an equivalent form, using the characteristic function on the cardinality of $\mathbf{z}$ and an integer-valued unknown k:

$$\operatorname*{minimize}_{\mathbf{x},\mathbf{z},\mathsf{k}} \chi_{\ell_0\le\mathsf{k}}(\mathbf{z}) + \chi_\Xi(\mathbf{x}) + f_c(\mathbf{x},\mathbf{z}) \tag{2.7}$$

where $f_c(\mathbf{x},\mathbf{z})$ is the appropriate coupling functional. For a fixed k, problem (2.7) can be seen as a variant of the *regressor selection* problem, which is (locally) solvable by ADMM [39]:

Synthesis version	Analysis version
$\bar{\mathbf{z}}^{(\mathsf{i}+1)} = \mathfrak{H}_\mathsf{k}(\hat{\mathbf{z}}^{(\mathsf{i})} + \mathbf{u}^{(\mathsf{i})})$	$\bar{\mathbf{z}}^{(\mathsf{i}+1)} = \mathfrak{H}_\mathsf{k}(\mathbf{A}\hat{\mathbf{x}}^{(\mathsf{i})} + \mathbf{u}^{(\mathsf{i})})$
$\hat{\mathbf{z}}^{(\mathsf{i}+1)} = \arg\min_{\mathbf{z}} \|\mathbf{z}-\bar{\mathbf{z}}^{(\mathsf{i}+1)}+\mathbf{u}^{(\mathsf{i})}\|_2^2$	$\hat{\mathbf{x}}^{(\mathsf{i}+1)} = \arg\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x}-\bar{\mathbf{z}}^{(\mathsf{i}+1)}+\mathbf{u}^{(\mathsf{i})}\|_2^2$
subject to $\mathbf{D}\mathbf{z}\in\Xi$	subject to $\mathbf{x}\in\Xi$
$\mathbf{u}^{(\mathsf{i}+1)} = \mathbf{u}^{(\mathsf{i})} + \hat{\mathbf{z}}^{(\mathsf{i}+1)} - \bar{\mathbf{z}}^{(\mathsf{i}+1)}$	$\mathbf{u}^{(\mathsf{i}+1)} = \mathbf{u}^{(\mathsf{i})} + \mathbf{A}\hat{\mathbf{x}}^{(\mathsf{i}+1)} - \bar{\mathbf{z}}^{(\mathsf{i}+1)}$.

(2.8)

Unlike the standard regressor selection algorithm, for which the ADMM multiplier [39] needs to be carefully chosen to avoid divergence, the above formulation is independent of its value.

In practice, it is difficult to guess the optimal value of k beforehand. An adaptive estimation strategy is to periodically increase k (starting from some small value), perform several runs of (2.8) for a given k and repeat the procedure until the constraint embodied by f_c is satisfied. This corresponds to *sparsity relaxation*: as k gets larger, the estimated $\mathbf{z}$ becomes less sparse - the same principle as the one applied in Consistent IHT.

[2] Observe that if $\mathbf{D}$ and $\mathbf{A}$ are unitary matrices, the two problems become identical.

The proposed algorithm, dubbed *SParse Audio DEclipper (SPADE)*, comes in two flavors. The pseudocodes for the synthesis version ("*S-SPADE*") and for the analysis version ("*A-SPADE*") are given in Algorithm 1 and Algorithm 2.

Algorithm 1 *S-SPADE*

Require: $\mathbf{D}, \mathbf{y}, \mathbf{M}_r, \mathbf{M}_c^+, \mathbf{M}_c^-, g, r, \varepsilon$

1: $\hat{\mathbf{z}}^{(0)} = \mathbf{D}^H\mathbf{y}, \mathbf{u}^{(0)} = \mathbf{0}, i = 1, k = g$
2: $\bar{\mathbf{z}}^{(i)} = \mathfrak{H}_k\left(\hat{\mathbf{z}}^{(i-1)} + \mathbf{u}^{(i-1)}\right)$
3: $\hat{\mathbf{z}}^{(i)} = \arg\min_{\mathbf{z}} \|\mathbf{z} - \bar{\mathbf{z}}^{(i)} + \mathbf{u}^{(i-1)}\|_2^2$
   s.t. $\mathbf{x} = \mathbf{D}\mathbf{z} \in \Xi$
4: **if** $\|\mathbf{D}\left(\hat{\mathbf{z}}^{(i)} - \bar{\mathbf{z}}^{(i)}\right)\|_2 \le \varepsilon$ **then**
5: terminate
6: **else**
7: $\mathbf{u}^{(i)} = \mathbf{u}^{(i-1)} + \hat{\mathbf{z}}^{(i)} - \bar{\mathbf{z}}^{(i)}$
8: $i \leftarrow i + 1$
9: **if** $i \bmod r = 0$ **then**
10: $k \leftarrow k + g$
11: **end if**
12: go to 2
13: **end if**
14: **return** $\hat{\mathbf{x}} = \mathbf{D}\hat{\mathbf{z}}^{(i)}$

Algorithm 2 *A-SPADE*

Require: $\mathbf{A}, \mathbf{y}, \mathbf{M}_r, \mathbf{M}_c^+, \mathbf{M}_c^-, g, r, \varepsilon$

1: $\hat{\mathbf{x}}^{(0)} = \mathbf{y}, \mathbf{u}^{(0)} = \mathbf{0}, i = 1, k = g$
2: $\bar{\mathbf{z}}^{(i)} = \mathfrak{H}_k\left(\mathbf{A}\hat{\mathbf{x}}^{(i-1)} + \mathbf{u}^{(i-1)}\right)$
3: $\hat{\mathbf{x}}^{(i)} = \arg\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \bar{\mathbf{z}}^{(i)} + \mathbf{u}^{(i-1)}\|_2^2$
   s.t. $\mathbf{x} \in \Xi$
4: **if** $\|\mathbf{A}\hat{\mathbf{x}}^{(i)} - \bar{\mathbf{z}}^{(i)}\|_2 \le \varepsilon$ **then**
5: terminate
6: **else**
7: $\mathbf{u}^{(i)} = \mathbf{u}^{(i-1)} + \mathbf{A}\hat{\mathbf{x}}^{(i)} - \bar{\mathbf{z}}^{(i)}$
8: $i \leftarrow i + 1$
9: **if** $i \bmod r = 0$ **then**
10: $k \leftarrow k + g$
11: **end if**
12: go to 2
13: **end if**
14: **return** $\hat{\mathbf{x}} = \hat{\mathbf{x}}^{(i)}$

The relaxation rate and the "greediness" (relaxation stepsize) are controlled by the integer-valued parameters $r > 0$ and $g > 0$, while the parameter $\varepsilon > 0$ is the stopping threshold.

Lemma 1. *The SPADE algorithms terminate in no more than* $i = \lceil nr/g + 1 \rceil$ *iterations.*

Proof. Once $k \ge n$, the hard thresholding operation $\mathfrak{H}_k$ becomes an identity mapping. Then, the minimizer of the constrained least squares step 3 is $\hat{\mathbf{z}}^{(i-1)}$ (respectively, $\hat{\mathbf{x}}^{(i-1)}$) and the distance measure in the step 4 is equal to $\|\mathbf{u}^{(i-1)}\|_2$. But, in the subsequent iteration, $\mathbf{u}^{(i-1)} = \mathbf{0}$ and the algorithm terminates. □

This bound is quite pessimistic: in practice, we observed that the algorithm terminates much sooner, which suggest that there might be a sharper upper bound on the iteration count.

2.3.1 Computational aspects

The general form of the SPADE algorithms does not impose restrictions on the choice of the dictionary nor the analysis operator. From a practical perspective, however, it is important that the complexity per iteration is kept low. The dominant cost of SPADE is in the evaluation of the linearly constrained least squares minimizer step, whose computational complexity can be generally high. Fortunately, for some choices of **D** and **A** this cost is dramatically reduced.

Namely, if the matrix $\mathbf{A}^H$ forms a *tight frame* ($\mathbf{A}^H\mathbf{A} = \zeta\mathbf{I}$, $\|\mathbf{A}\mathbf{v}\|_2 = \zeta\|\mathbf{v}\|_2$), it is easy to verify that the step 3 of A-SPADE reduces to:

$$\begin{aligned}\mathbf{x}^{(i)} &= P_\Xi\left(\frac{1}{\zeta}\mathbf{A}^H(\bar{\mathbf{z}}^{(i)} - \mathbf{u}^{(i-1)})\right), \text{ where:}\\ \Xi &= \left\{\mathbf{x} \,|\, \begin{bmatrix} -\mathbf{M}_c^+ \\ \mathbf{M}_c^- \end{bmatrix}\mathbf{x} \le \begin{bmatrix} -\mathbf{M}_c^+ \\ \mathbf{M}_c^- \end{bmatrix}\mathbf{y} \text{ and } \mathbf{M}_r\mathbf{x} = \mathbf{M}_r\mathbf{y}\right\}.\end{aligned} \tag{2.9}$$

The projection $P_\Xi(\cdot)$ is straightforward and corresponds to component-wise mappings[3], as mentioned in subsection A.3. Thus, the per iteration cost of the algorithm is reduced to the cost of evaluating matrix-vector products.

Unfortunately, for S-SPADE this simplification is not possible and the constrained minimization in step 3 needs to be computed iteratively. However, by exploiting the tight frame property of $\mathbf{D} = \mathbf{A}^H$ and the Woodbury matrix identity, one can build an efficient algorithm that solves this optimization problem with low complexity. For instance, another ADMM can be nested inside S-SPADE, as follows:

$$\begin{aligned}\hat{\mathbf{z}}^{(j+1)} &= \left(\mathbf{I} - \frac{\gamma}{1+\zeta\gamma}\mathbf{D}^H\mathbf{D}\right)\left(\bar{\mathbf{z}}^{(i)} - \mathbf{u}^{(i-1)} + \gamma\mathbf{D}^H\left(\mathbf{w}^{(j)} - \mathbf{u}_w^{(j)}\right)\right)\\ \mathbf{w}^{(j+1)} &= P_\Xi\left(\mathbf{D}\hat{\mathbf{z}}^{(j+1)} + \mathbf{u}_w^{(j)}\right),\\ \mathbf{u}_w^{(j+1)} &= \mathbf{u}_w^{(j)} + \mathbf{D}\hat{\mathbf{z}}^{(j+1)} - \mathbf{w}^{(j+1)},\end{aligned} \tag{2.10}$$

where γ, $\mathbf{w}$ and $\mathbf{u}_w$ are the *inner* ADMM multiplier, auxiliary and dual variable, respectively. The per-iteration complexity of this nested algorithm is again reduced to the cost of several matrix-vector products, but the overall complexity of S-SPADE can still be significantly higher than of the A-SPADE algorithm.

Finally, the computational complexity can be further reduced if the matrix-vector products with $\mathbf{D}$ and $\mathbf{A}$ can be computed with less than quadratic cost. Some transforms that support both tight frame property and fast product computation are also favorable in our audio (co)sparse context. Such well-known transforms are Discrete Fourier Transform (DFT), (Modified) Discrete Cosine Transform (M)DCT, (Modified) Discrete Sine Transform (M)DST and Discrete Wavelet Transform (DWT), for instance.

2.3.2 Complex transforms

Precaution should be taken when complex transforms (such as DFT) are used for the dictionary / the analysis operator. The reason is obvious - we are estimating real signals and the algorithms should be aware of it. Formally, this requirement is embodied into Ξ, but the numerical procedure for the least squares minimization needs to be accordingly adapted.

[3] Since the matrices $\mathbf{M}_r$, $\mathbf{M}_c^+$ and $\mathbf{M}_c^-$ are restriction operators.

Explicitly, in the synthesis case, we would solve the following optimization problem:

$$\underset{\mathbf{z}}{\text{minimize}} \|\mathbf{z}-\bar{\mathbf{z}}^{(\mathrm{i})}+\mathbf{u}^{(\mathrm{i}-1)}\|_2^2 \quad \text{s.t.} \quad \mathbf{x}=\mathbf{D}\mathbf{z}\in\Xi, \quad \Re(\mathbf{D})\Im(\mathbf{z})=-\Im(\mathbf{D})\Re(\mathbf{z}), \tag{2.11}$$

where $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and the imaginary part of the argument, respectively.

In the analysis case, the optimization problem writes as follows:

$$\underset{\mathbf{x}}{\text{minimize}} \|\Re(\mathbf{A})\mathbf{x}-\Re(\bar{\mathbf{z}}^{(\mathrm{i})}+\mathbf{u}^{(\mathrm{i}-1)})\|_2^2+\|\Im(\mathbf{A})\mathbf{x}-\Im(\bar{\mathbf{z}}^{(\mathrm{i})}+\mathbf{u}^{(\mathrm{i}-1)})\|_2^2 \quad \text{s.t.} \quad \mathbf{x}\in\Xi. \tag{2.12}$$

Both problems can be straightforwardly solved by proximal splitting, and the tight frame structure can still be exploited.

2.4 Experiments

The experiments are aimed to highlight differences in audio enhancement performance between S-SPADE and A-SPADE, and implicitly, the sparse and cosparse data models. It is noteworthy that in the formally equivalent setting ($\mathbf{A}=\mathbf{D}^{-1}$), the two algorithms become identical. As a sanity-check, we include this setting in the experiments. The relaxation parameters are set to $\mathsf{r}=1$ and $\mathsf{g}=1$, and the stopping threshold is $\varepsilon=0.1$. Additionally, we include Consistent IHT and social sparsity declipping algorithms as representatives of state-of-the-art. The former is known to be very computationally efficient, while the latter should exhibit good declipping performance.

As mentioned before, this work is not aimed towards investigating the appropriateness of various time-frequency transforms in the context of audio recovery, which is why we choose traditional Short Time Fourier Transform (STFT) for all experiments. We use sliding square-rooted Hamming window of size 1024 samples with 75% overlap. The redundancy level of the involved frames (corresponding to *per-chunk* inverse DFT for the dictionary and forward DFT for the analysis operator) is 1 (no redundancy), 2 and 4. The social sparsity declipper, based on Gabor dictionary, requires batch processing of the whole signal. We adjusted the temporal shift, the window and the number of frequency bins in accordance with previously mentioned STFT settings [4].

2.4.1 Numerical performance

Here we investigate the numerical performance of the concerned algorithms. Audio examples consist of 10 music excerpts taken from RWC database [115], with different tonal and vocal content. The excerpts are of approximately similar duration (~ 10s), and are sampled at 16kHz with 16bit encoding. The inputs are generated by artificially clipping the audio excerpts at five levels, ranging from severe ($\mathrm{SDR}_{\mathbf{y}}=1$dB) towards mild ($\mathrm{SDR}_{\mathbf{y}}=10$dB).

[4] We use the implementation kindly provided by the authors.

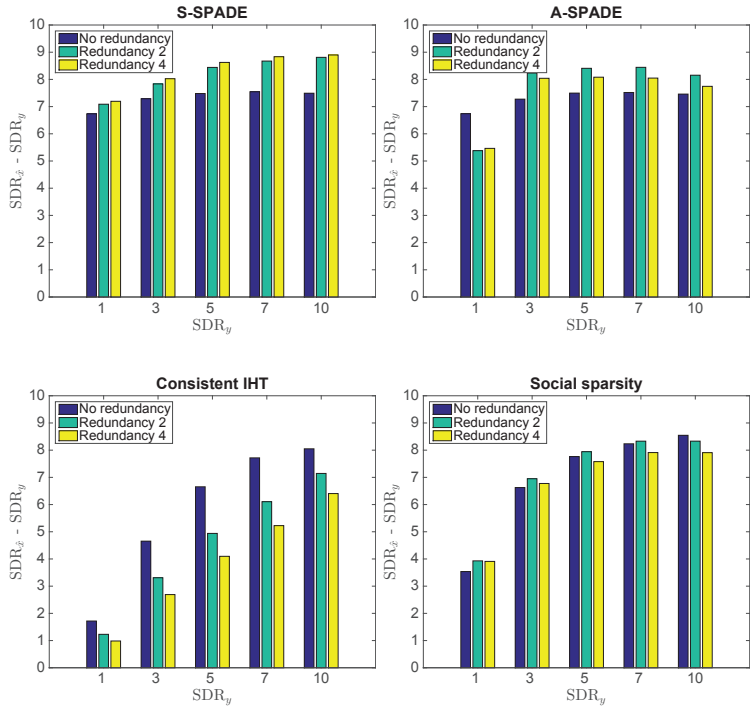


Figure 2.2 – Declipping performance of the four algorithms in terms of the SDR improvement.

Signal recovery As a recovery performance measure in these experiments, we use a simple difference between Signal-to-Distortion Ratios (SDR) of clipped ($\mathrm{SDR}_{\mathbf{y}}$) and processed ($\mathrm{SDR}_{\hat{\mathbf{x}}}$) signals:

$$\mathrm{SDR}_{\mathbf{y}} = 20\log_{10}\frac{\|\left[\begin{smallmatrix}\mathbf{M}_c^+\\ \mathbf{M}_c^-\end{smallmatrix}\right]\mathbf{x}\|_2}{\|\left[\begin{smallmatrix}\mathbf{M}_c^+\\ \mathbf{M}_c^-\end{smallmatrix}\right]\mathbf{x}-\left[\begin{smallmatrix}\mathbf{M}_c^+\\ \mathbf{M}_c^-\end{smallmatrix}\right]\mathbf{y}\|_2}, \quad \mathrm{SDR}_{\hat{\mathbf{x}}} = 20\log_{10}\frac{\|\left[\begin{smallmatrix}\mathbf{M}_c^+\\ \mathbf{M}_c^-\end{smallmatrix}\right]\mathbf{x}\|_2}{\|\left[\begin{smallmatrix}\mathbf{M}_c^+\\ \mathbf{M}_c^-\end{smallmatrix}\right]\mathbf{x}-\left[\begin{smallmatrix}\mathbf{M}_c^+\\ \mathbf{M}_c^-\end{smallmatrix}\right]\hat{\mathbf{x}}\|_2}.$$

Hence, only the samples corresponding to clipped indices are taken into account. Concerning *SPADE*, this choice makes no difference, since the remainder of the estimate $\hat{\mathbf{x}}$ perfectly fits the observations. However, it may favor the other two algorithms that do not share this feature.

According to the results presented in figures 2.2 and 2.3, the *SPADE* algorithms yield highest improvement in SDR among the four considered approaches, mostly pronounced when clipping is severe. As assumed, S-SPADE and A-SPADE achieve similar results in a non-redundant setting, but when the overcomplete frames are considered, the synthesis version performs somewhat better. Moreover, S-SPADE is the only algorithm whose performance consistently improves with redundancy. Interestingly, the overall best results for the analysis version are obtained for the twice-redundant frame, while the performance slightly drops for the redundancy four. This is probably due to the absolute choice of the parameter ε, and suggests that in the analysis setting, this value should be replaced by a relative threshold.

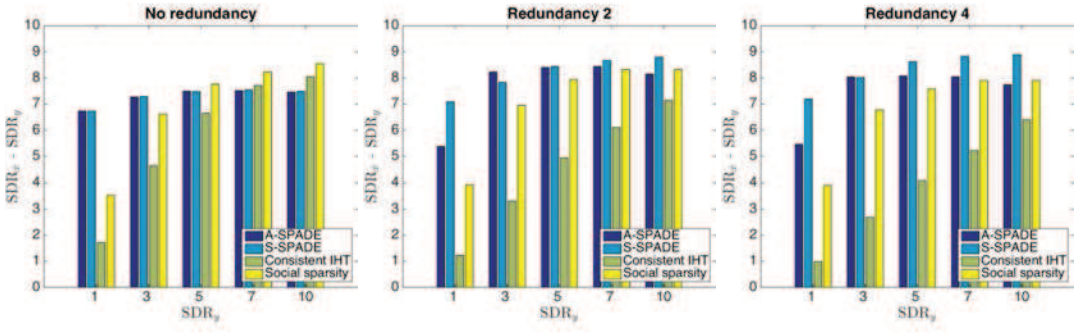


Figure 2.3 – SDR improvement *vs* redundancy for all algorithms.

Processing time We decided not to present all processing time results, due to the way social sparsity declipper is implemented: first, its stopping criterion is based on the iteration count, and second, for its heavy computations, it uses a time-frequency toolbox whose backend is coded in C programming language (and compiled), as opposed to the other algorithms which are fully implemented in Matlab®. We may only remark that even this accelerated version of the code was still somewhat slower than A-SPADE and Consistent IHT, which require (on the average) 3min and 7min, respectively, to declip the audio in the non-redundant case, compared to about 10min for the social sparsity algorithm in the same setting.

However, we are interested in the computational costs of A-SPADE and S-SPADE, as these two algorithms are our proxies for the analysis and synthesis data models. Table 2.1 shows a huge difference in processing time between the two algorithms, with S-SPADE being extremely costly, due to the nested iterative minimization procedure (2.10). This is not very surprising, since the synthesis version usually needs to perform orders of magnitude more matrix-vector multiplications (in total) than the analysis one. Although there might be a more resourceful way to implement the costly S-SPADE projection step, it cannot be as efficient as the closed form solution implemented in A-SPADE. On the other hand, the computational cost of A-SPADE grows faster than the cost of S-SPADE, with respect to the redundancy parameter (although their absolute difference is still highly in favor of the analysis algorithm). This might be another indicator that A-SPADE should take into account the redundancy factor, in order to avoid wasteful iterations and, possibly, improve signal recovery performance.

Redundancy	Data model	1dB	3dB	5dB	7dB	10dB
1	Analysis	1	3	4	4	5
	Synthesis	265	471	641	746	783
2	Analysis	5	8	9	11	11
	Synthesis	328	539	698	796	864
4	Analysis	16	26	32	35	37
	Synthesis	502	786	961	1067	1125

Table 2.1 – Processing times in minutes for the A-SPADE (analysis) and S-SPADE (synthesis) algorithms.

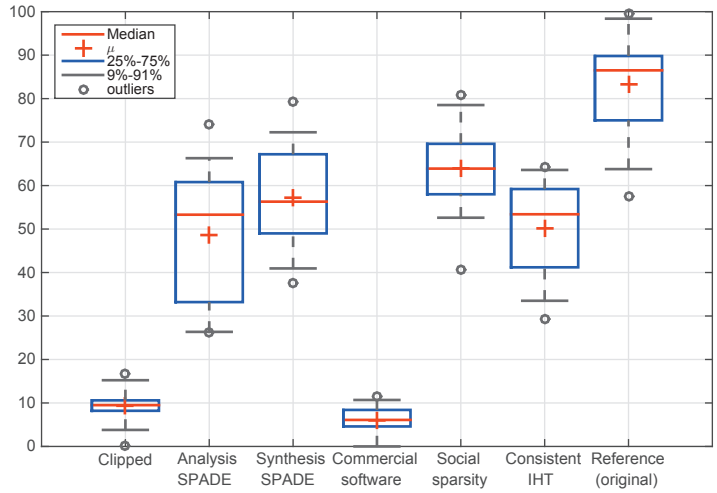


Figure 2.4 – Perceptual evaluation results based on the MUSHRA test method.

2.4.2 Perceptual evaluation

In this series of experiments we are interested in the perceptual quality performance, evaluated by human listeners. For this purpose, we use the *MUltiple Stimuli with Hidden Reference and Anchor (MUSHRA)* evaluation framework with a web-based interface (*i.e. BeaqleJS* [149]). As noted in [247], MUSHRA provides relevant results when the reference (original) audio can be reliably identified from the processed (in our case, declipped) signals. Therefore, we restrict the evaluation to severe clipping only ($\mathrm{SDR}_{\mathbf{y}} = 3\mathrm{dB}$), and choose, for each algorithm, the setting in which it achieves highest numerical recovery performance (*i.e.* 4-redundant for S-SPADE, 2-redundant for A-SPADE and social sparsity, and non-redundant for Consistent IHT). In total, the evaluation group consist of 14 expert and non-expert listeners, who were asked to grade the quality of an audio track on a scale from 0 ("bad") to 100 ("excellent"). Each participant evaluates five audio tracks, chosen randomly from the set of ten tracks we used for numerical evaluation. In addition to the output of each algorithm, included are the hidden reference and clipped tracks, as well as the audio processed by a professional audio restoration software.

The evaluation results are presented in figure 2.4. The social sparsity declipper obtains the highest median score, followed by S -SPADE, while A-SPADE and Consistent IHT share the third place. The score difference among these four algorithms is small, *i.e.* about 10 between the first and the fourth in ranking. On the other hand, the commercial software's output is graded worse than the clipped signal itself. The good listening test performance of the social sparsity algorithm verifies that accounting for the refined structure in the time-frequency plane improves perceptual quality, thus serving as a good prior model for audio signals.

2.5 Summary and contributions

This chapter was about the declipping inverse problem, addressed by the sparse synthesis and sparse analysis regularizations. We presented a novel and flexible declipping algorithm, that can easily accommodate sparse (S-SPADE) or cosparse (A-SPADE) prior, and as such has been used to compare the recovery performance of the two data models.

The empirical results are slightly in favor of the sparse synthesis data model. However, the analysis version does not fall far behind, which makes it attractive for practical applications. Indeed, due to the natural way of imposing clipping consistency constraints, it can be implemented in an extremely efficient way, even allowing for real-time signal processing. We envision that the performance of A-SPADE can be enhanced by more appropriate choice of stopping criteria and parameterization. Numerical benchmark and perceptual evaluation of real audio verify that the two versions perform competitively against considered state-of-the-art algorithms in the field, but may be further improved by incorporating structured (*e.g.* social) (co)sparsity priors.

In the next chapter we will discuss a class of inverse problems inspired and dictated by physics, where the synthesis and analysis model become nominally equivalent. Despite the model equivalence, as we have seen in the non-redundant setting for the declipping inverse problem, computational complexity of the two sparse and cosparse approaches can be very different. We will discuss this algorithmic aspect in greater detail in the following material.

3 Physics-driven inverse problems

We started this thesis by introducing inverse problems illustrated by some physical examples. Indeed, this type of inverse problems is of a very large scope and very often encountered in practice. At the same time, many of these problems are quite challenging due to their ill-posed nature. In a real-world setting, this behavior is closely linked with the fact that there are physical and practical limitations in terms of the amount of measurements we can take. In other words, we are capable of obtaining only discrete observations. On the other hand, physical phenomena are essentially continuous and inherently infinite-dimensional. In turn, the measurement operators will not fulfill injectivity and/or stability criterions necessary to ensure well-posedness. Therefore, if there is any hope of finding solutions, such problems call for regularization. Fortunately, there is usually a side information that may be used to devise a correctness class. This chapter concerns *certain* inverse problems arising in physical context, for which the correctness class can be interpreted in terms of a set of maximally sparse or cosparse solutions (possibly with additional constraints). Since these problems are directly related to physical laws and phenomena, we term this class *physics-driven inverse problems*. The *nominally equivalent* sparse analysis and sparse synthesis regularizations applied to problems of this class are encompassed in the *physics-driven (co)sparse*, or shorter, the physics-driven framework.

Physics-driven inverse problems should be recognized as an instance of array signal processing problems, which we recall in the first section, along with the main contributions of this chapter. The second section is a brief introduction to linear partial differential equations and the associated *Green's functions*. In the same section we turn our attention to Poisson's equation, which is essential for the EEG inverse problem, but also lays foundation to more involved equations. One of these is the wave equation, which is discussed in the remainder of the section. In the third section of the chapter, we draw connections between such physical models and the sparse analysis/synthesis regularizations. The fourth section is about a practical method to solve the optimization problems arising from these regularizations. The final, fifth section is dedicated to discussion on computational complexities of the analysis and synthesis regularizations in the physics-driven context.

3.1 Array signal processing

Array signal processing is concerned with treatment of (usually electromagnetic or acoustic) data acquired from more than one spatially distributed sensors. It has found numerous uses in radar applications, radio astronomy, navigation, wireless communications, biomedical and sonar signal processing, to name a few. The advantages of this approach over single channel signal processing come from the natural observation that the wave-like physical phenomena exist in up to three spatial and one temporal dimension. When only one fixed sensor is available, we are "sensing" only the temporal evolution of a signal at a given sensor location. As a consequence, in order to infer some information from the signal, one usually needs to impose stronger prior assumptions on the signal model. Contrary, the sensor array measurements also allow for observing the spatial character of the signal, which may help in building more accurate signal models.

However, this additional data comes at a cost - first in terms of the physical equipment (a network of sensors - an "*array*"), but also in terms of the complexity of applied algorithms. One obvious cause of complexity is the increased amount of data that has to be processed, but this is not the only reason. Extending standard single-channel methods to signal processing on arrays is not straightforward, due to the spatial sampling constraints. Indeed, sampling in spatial domain is much cruder than in the temporal domain, and standard signal processing theory warns us that naive treatment of this data leads to spatial aliasing. These factors impose difficulties in designing algorithms for array signal processing, but the prize is given as the potential to address inverse problems which are extremely challenging, if not impossible, to solve using only one sensor.

If we adopt Occam's reasoning, and assume that Nature indeed prefers economic descriptions, the low-complexity regularization may prevent the spatial aliasing phenomenon. Inspired by this principle, we provide an intuitive framework that puts under the same umbrella certain physical problems ruled by linear partial differential equations and the sparse regularization concept.

3.1.1 Contributions

This chapter is largely based on the framework paper [139], which aims at unifying physics-driven inverse problems regularized by the sparse and cosparse data models. More than that, we highlight that the two regularizations are equivalent in the physics-driven case, but that the optimization problems generated by these two are different from computational point of view. Indeed, provided that the continuous domain problems are discretized using locally supported methods (such as finite differences or finite element methods), we will see that the sparse analysis regularization is a preferable choice in practice. Additionally, we introduce Weighted SDMM, a simple ADMM-based optimization algorithm for solving convex-regularized physics-driven inverse problems.

3.2 Linear partial differential equations and the Green's functions

Signals of our interest are physical quantities obeying certain known physical laws. As mentioned, these represent a huge class: for example, sound propagates according to the acoustic wave equation, Magnetic Resonance Imaging (MRI) is based on Bloch's equations, Maxwell's equations are at the foundation of wireless communications etc. We limit our scope to signals which can be modeled by *linear Partial Differential Equations (PDEs)*. In the context of ill-posed inverse problems, the knowledge that a signal satisfies a linear PDE is a strong prior information that could be useful to perform regularization.

Linear PDEs take the following form:

$$\sum_{|\mathbf{d}|\leq\zeta} a_{\mathbf{d}}(\boldsymbol{\omega}) D^{\mathbf{d}} x(\boldsymbol{\omega}) = z(\boldsymbol{\omega}), \quad \boldsymbol{\omega} \in \Omega \tag{3.1}$$

where $a_{\mathbf{d}}$, x and z are functionals of the parameter $\boldsymbol{\omega}$ (e.g., space and/or time) in the domain Ω, and $\mathbf{d}$ is the *multi-index* variable with $|\mathbf{d}| = \mathsf{d}_1 + \ldots + \mathsf{d}_\mathsf{l}$, and $\mathsf{d}_\mathsf{i} \in \mathbb{N}_0$. Note that, in general, Ω is not a vector space, but rather a *differentiable manifold* - a topological space that only locally behaves as a Euclidean space. Therefore, the domain variable $\boldsymbol{\omega}$ is simply a *tuple* of coordinates in Ω. However, for our purposes, even the domain Ω is simplified to a set of a vector space.

For a given $\mathbf{d} = (\mathsf{d}_1, \ldots, \mathsf{d}_\mathsf{l})$, $D^{\mathbf{k}} x(\boldsymbol{\omega})$ denotes the $\mathbf{d}^{\text{th}}$ partial differential of x with respect to $\boldsymbol{\omega}$, defined as:

$$D^{\mathbf{d}} x(\boldsymbol{\omega}) = \frac{\partial^{|\mathbf{k}|} x}{\partial \omega_1^{\mathsf{d}_1} \partial \omega_2^{\mathsf{d}_2} \ldots \partial \omega_\mathsf{l}^{\mathsf{d}_\mathsf{l}}}.$$

3.2.1 Boundary conditions

In order to satisfy Hadamard's well-posedness requirements, a PDE is accompanied with appropriate *boundary* and/or *initial conditions*. These depend on the type of PDE, and the physical problem we aim at modeling. Boundary conditions dictate how should the solution behave at the boundaries $\partial\Omega$ of the domain. The most common ones are:

Dirichlet : assigns a value at the boundary, *i.e.* $\forall \boldsymbol{\omega} \in \partial\Omega\colon\ x(\boldsymbol{\omega}) = b_D(\boldsymbol{\omega})$, where $b_D(\boldsymbol{\omega})$ is known.

Neumann : assigns a *flux* at the boundary, *i.e.* $\forall \boldsymbol{\omega} \in \partial\Omega\colon\ \nabla x(\boldsymbol{\omega}) \cdot \mathbf{n}(\boldsymbol{\omega}) = b_N(\boldsymbol{\omega})$, where $b_N(\boldsymbol{\omega})$ is known and $\mathbf{n}$ is an outward normal vector to the boundary $\partial\Omega$.

Mixed : a subset of the boundary $\partial\Omega$ is modeled by Dirichlet, and the complementary subset is modeled by Neumann condition.

Robin : assigns a linear combination of Dirichlet and Neumann conditions at each

point of the boundary, *i.e.* $\forall \boldsymbol{\omega} \in \partial\Omega\colon\ b_1(\boldsymbol{\omega})x(\boldsymbol{\omega}) + b_2(\boldsymbol{\omega})\nabla x(\boldsymbol{\omega}) \cdot \mathbf{n}(\boldsymbol{\omega}) = b_R(\boldsymbol{\omega})$, where $b_R(\boldsymbol{\omega}), b_1(\boldsymbol{\omega})$ and $b_2(\boldsymbol{\omega})$ are all known.

Cauchy : imposes *both* Dirichlet and Neumann data at the boundary.

Periodic : for some geometries, these conditions enforce $x(\boldsymbol{\omega}_\mathrm{i}) = x(\boldsymbol{\omega}_\mathrm{j})$, where $\boldsymbol{\omega}_\mathrm{i} \in \partial\Omega_\mathrm{i}$, $\boldsymbol{\omega}_\mathrm{j} \in \partial\Omega_\mathrm{j}$, such that $\{\partial\Omega_\mathrm{i}, \partial\Omega_\mathrm{j}\} \subseteq \partial\Omega$.

A partial differential equation with associated boundary conditions is commonly known as *boundary value problem*. It is noteworthy that for an arbitrary combination of $a_\mathbf{k}$, z and boundary conditions, the solution x of (3.1) is not guaranteed to exist, or to be unique (these questions are still subject of active research in the field of partial differential equations). Even if a unique solution exists, it may not be possible to derive it analytically.

However, if PDE is modeling a time-dependent quantity $x(\boldsymbol{\omega}) = x(\mathbf{r}, \mathrm{t})$, on the product domain $\Omega = \Gamma \times [0, \tau]$, the Cauchy-Kovalevskaya theorem [99] states that the *initial conditions* suffice to ensure uniqueness of the solution[1]. The initial conditions can be seen as one-sided (at $\mathrm{t} = 0$) Dirichlet conditions for the first $|\mathbf{k}| - 1$ derivatives of a PDE:

$$\partial_\mathrm{t}^\mathrm{k} x(\mathbf{r}, 0) = i_\mathrm{t}^\mathrm{k}(\mathbf{r}), \forall (\mathbf{r} \in r, \mathrm{k} \leq |\mathbf{k}| - 1). \tag{3.2}$$

This is called an *inital value problem*. In the same way, if both boundary and initial conditions are prescribed, we have an *initial boundary value problem*. Hereafter, we will use the fact that t is just another domain variable, and informally address initial (boundary) value problems as a subtype of boundary value problems.

3.2.2 Operator form

Differentiation is a linear operation, therefore we can represent (3.1) compactly in the linear operator form:

$$Lx = z, \tag{3.3}$$

where $L = \sum_{|\mathbf{k}| \leq \zeta} a_\mathbf{k}(\boldsymbol{\omega}) D^\mathbf{k}\cdot$, $x := x(\boldsymbol{\omega})$ and $z := z(\boldsymbol{\omega})$.

Boundary conditions we mentioned earlier are also linear with respect to x, therefore we can equip L with appropriate boundary conditions $Bx_{\partial\Omega} = z_{\partial\Omega}$ such that the newly generated operator $A := (L, B)$ defines a well-posed problem:

$$Ax = z \quad \Leftrightarrow \quad Lx = z, \quad Bx_{\partial\Omega} = z_{\partial\Omega}, \tag{3.4}$$

where x and z on the left side, by abuse of notation[2], also encompass $x_{\partial\Omega}$ and $z_{\partial\Omega}$.

[1] Valid when all $a_\mathbf{k}(\boldsymbol{\omega})$ are real analytic functions [160].

[2] We will occasionally use x and z defined only on the interior of the domain, which is always indicated in the text.

Now, we restrict our attention to *self-adjoint* operators A. This means that, for any $u, v \in \mathtt{dom}(A)$, the following holds:

$$\langle Au, v\rangle = \langle u, Av\rangle, \tag{3.5}$$

i.e. the *adjoint operator* A^* is identical to the operator A. Bearing in mind that A incorporates boundary conditions, the adjoint A^* is equal to A if and only if their domains and boundary conditions are the same. If the two operators differ only in boundary terms, the operator A is called *formally self adjoint*. In this thesis we consider only self-adjoint operators, however, generalization towards formally self-adjoint operators is possible, provided that we have sufficient boundary information.

An "inverse" operation to differentiation is integration. Assuming that there exists a function $g(\boldsymbol{\omega}, \mathbf{w})$ such that the following holds:

$$x(\boldsymbol{\omega}) = \int_\Omega z(\mathbf{w}) g(\boldsymbol{\omega}, \mathbf{w}) d\mathbf{w} + \text{boundary terms}\ ^3, \tag{3.6}$$

we have the integral representation of the solution $x(\boldsymbol{\omega})$ of (3.4). The function $g(\boldsymbol{\omega}, \mathbf{w})$ is known as the *Green's function* or the *fundamental solution* of a PDE. The Green's functions are constructed by solving the following boundary value problem:

$$\begin{aligned} Lg(\boldsymbol{\omega}, \mathbf{w}) &= \delta(\boldsymbol{\omega} - \mathbf{w}), \quad \mathbf{w} \in \Omega \setminus \partial\Omega, \\ Bg(\boldsymbol{\omega}, \mathbf{w}) &= 0, \quad \mathbf{w} \in \partial\Omega. \end{aligned} \tag{3.7}$$

Here $\delta(\cdot)$ denotes *Dirac delta* distribution:

$$\delta(\boldsymbol{\omega}_0) = \begin{cases} \infty & \text{when } \boldsymbol{\omega} = \boldsymbol{\omega}_0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.8}$$

In signal processing language, the Green's functions correspond to impulse responses of a linear operator L with imposed homogeneous boundary conditions. There are some ambiguities in the literature concerning the terminology: in some cases, boundary/initial conditions are omitted, hence $g(\boldsymbol{\omega})$ are fundamental solutions of a linear PDE (3.1) alone, and therefore not unique. In this work, we consider "*the*" fundamental solutions, which are uniquely defined by the problem (3.7).

Since integration is again a linear operation, we compactly represent the integral (3.6) in operator form, as follows:

$$x = Dz. \tag{3.9}$$

[3] The "boundary terms" depend on the particular problem and for self-adjoint A they vanish.

Provided that $A = (L, B)$ is self-adjoint, the operators D and A can be seen as the inverses of each other. This is due to:

$$x(\boldsymbol{\omega}) = \int_\Omega x(\mathbf{w})\delta(\boldsymbol{\omega}-\mathbf{w})d\mathbf{w} = \int_\Omega x(\mathbf{w})Ag(\boldsymbol{\omega},\mathbf{w})d\mathbf{w} = \int_\Omega Ax(\mathbf{w})g(\boldsymbol{\omega},\mathbf{w})d\mathbf{w} = \int_\Omega z(\mathbf{w})g(\boldsymbol{\omega},\mathbf{w})d\mathbf{w}, \quad (3.10)$$

where the third equality is the consequence of self-adjointness of A. Extending this approach to linear operators which are formally self-adjoint is possible by accounting for the boundary terms. In the case where the operator is not even formally self-adjoint, defining the integral representation may be more difficult.

In the following two subsections we give a brief introduction to two PDEs of our interest: Poisson's and the linear wave equation.

3.2.3 Poisson's equation

Poisson's equation is one of the most common partial differential equations in physics and engineering. It is widely used in electrostatics, mechanics and thermodynamics, where it models steady-state phenomena. For any $\boldsymbol{\omega} \in \Omega \setminus \partial\Omega$ it is defined as:

$$\triangle x(\boldsymbol{\omega}) = z(\boldsymbol{\omega}), \quad (3.11)$$

where $L = \triangle$ is known as the *Laplace operator* (when defined on a *Riemannian manifold*[4], it is known as the *Laplace-Beltrami* operator). For simplicity, we assume that $\Omega \subseteq \mathbb{R}^{\mathsf{d}}$, $\mathtt{dim}(\Omega) = \mathsf{d}$, which yields:

$$\triangle x(\boldsymbol{\omega}) = \triangle_{\boldsymbol{\omega}} x = \nabla \cdot \nabla x = \frac{\partial^2 x}{\partial \omega_1^2} + \frac{\partial^2 x}{\partial \omega_2^2} + \ldots + \frac{\partial^2 x}{\partial \omega_{\mathsf{d}}^2}, \quad (3.12)$$

i.e. Laplace operator is the divergence of the gradient acting on x.

The Laplace operator is formally self-adjoint, and when accompanied with appropriate boundary conditions B it becomes a "fully" self-adjoint operator (yielding the operator A). Particularly, when homogeneous Dirichlet, Neumann or mixed boundary conditions are imposed, the operator A is self-adjoint. This is easily observed by exploiting *Green's second identity*:

$$\int_\Omega (u\triangle v - v\triangle u)d\boldsymbol{\omega} = \int_{\partial\Omega} (u\nabla v - v\nabla u)\cdot \mathbf{n}\, d\underline{\boldsymbol{\omega}} \quad (3.13)$$

where $\underline{\boldsymbol{\omega}}$ is an integration variable over $\partial\Omega$ and $\mathbf{n}$ is the outward normal vector to the boundary $\partial\Omega$. For those three boundary conditions, the right hand side vanishes, and the operator A is self-adjoint. Hence, if the solution of (3.7) is available, we can "invert" Poisson's equation. An

[4] A Riemannian manifold is a differentiable manifold whose *tangent spaces* are endowed with an inner product [158].

explicit solution of the Green's function for Poisson's equation is available in certain settings, as we will see later.

For $z(\boldsymbol{\omega}) = 0$, the expression (3.11) is known as Laplace's equation. Therefore, its solutions constitute the null space of Poisson's equation - these are trivial only if certain boundary conditions are imposed. Uniqueness results for homogeneous Dirichlet, Neumann and mixed boundary conditions are well-known and easy to derive. Assume, for instance, that x_1 and x_2 are both solving the boundary value problem induced by (3.11) and B. Then, their difference $\tilde{x} = x_1 - x_2$ is a solution of Laplace's equation:

$$\triangle \tilde{x} = \triangle (x_1 - x_2) = z - z = 0. \tag{3.14}$$

Green's first identity states:

$$\int_{\Omega} (\tilde{x} \triangle \tilde{x} + \nabla \tilde{x} \cdot \nabla \tilde{x})\, d\boldsymbol{\omega} = \int_{\partial\Omega} \tilde{x} \nabla \tilde{x} \cdot \mathbf{n} d\underline{\boldsymbol{\omega}}.$$

Hence, if either of the three boundary conditions is present, the right side of this equality is zero. Moreover, the first term on the left side vanishes due to (3.14), meaning that $\nabla \tilde{x} = \mathbf{0}$, *i.e.* $x_1 - x_2 = \text{const}$ everywhere in Ω. In the homogeneous Dirichlet (and mixed boundary) setting, this implies $x_1 = x_2$. In the "pure" Neumann case, the solution is unique up to an additive constant.

3.2.4 The wave equation

The wave equation is another fundamental linear partial differential equation, arising in fields such as electromagnetics, fluid dynamics and relativity theory, where it is used to model certain dynamic phenomena. Here, we define it on the *spacetime* product space $\boldsymbol{\omega} := (\mathbf{r}, \mathrm{t}) \in \Omega = \Gamma \times (\mathrm{t}_1, \mathrm{t}_2)$ (possibly, $\mathrm{t}_1 = -\infty, \mathrm{t}_2 = \infty$), as follows:

$$\Box_{\boldsymbol{\omega}} x = \triangle_{\mathbf{r}} x(\mathbf{r}, \mathrm{t}) - \frac{1}{c(\mathbf{r}, \mathrm{t})^2} \frac{\partial^2 x(\mathbf{r}, \mathrm{t})}{\partial \mathrm{t}^2} = z(\mathbf{r}, \mathrm{t}), \tag{3.15}$$

where $c(\mathbf{r}, \mathrm{t})$ denotes the *speed of propagation*. The domain Ω is generally represented by a so-called *Lorentzian manifold*[5]. Again, for simplicity, we assume more restricted space, where $\Omega = \Gamma \times \mathbb{R}$, $\Gamma \subseteq \mathbb{R}^{\mathsf{d}}$, $\mathtt{dim}(\Gamma) = \mathsf{d}$ (thus, $\mathtt{dim}(\Omega) = \mathsf{d} + 1$). The operator $L = \Box$ is known as *D'Alembert operator*, and is also formally self-adjoint.

Since this is an initial value problem, it is necessary to impose appropriate Cauchy conditions to ensure uniqueness:

$$x(\mathbf{r}, \mathrm{t}_1) = i(\mathbf{r}), \quad \frac{\partial x(\mathbf{r}, \mathrm{t}_1)}{\partial \mathrm{t}} = i_{\mathrm{t}}(\mathbf{r}). \tag{3.16}$$

[5] A generalization of Riemannian manifold, for which the *metric tensor* is not positive definite, since the temporal and spatial dimensions have opposite signs [158]

Additionally, one may have boundary conditions $Bx(\mathbf{r},\mathrm{t})$, for $\mathbf{r}\in\partial r$, which are compatible with the initial ones at $(\mathbf{r},\mathrm{t}_1)$. In our case, initial and boundary conditions are homogeneous: $i(\mathbf{r}) = i_{\mathrm{t}}(\mathbf{r}) = Bx(\mathbf{r},\mathrm{t}) = 0$.

For several interesting homogeneous initial and boundary conditions, the encompassing operator A is self-adjoint. To illustrate this, we simplify matters by assuming that $c(\mathbf{r},\mathrm{t}) = \mathrm{c} > 0$ is constant throughout the domain Ω. By applying the change of variables $\tilde{\boldsymbol{\omega}} = \left(\mathbf{r}, \frac{\mathrm{it}}{c}\right) \in \tilde{\Omega}$, where $\mathrm{i}^2 = -1$, we can identify $\Box_{\boldsymbol{\omega}}$ with $\triangle_{\tilde{\boldsymbol{\omega}}}$. Then, we can again use second Green's identity (3.13) which gives:

$$\int_{\partial\tilde{\Omega}} (u\nabla v - v\nabla u)\cdot\mathbf{n}\, d\underline{\tilde{\boldsymbol{\omega}}} = \lim_{\mathrm{t}\to\mathrm{t}_1}\int_{\partial r}\left(u\left(\mathbf{r},\frac{\mathrm{it}}{\mathrm{c}}\right)\nabla v\left(\mathbf{r},\frac{\mathrm{it}}{\mathrm{c}}\right) - v\left(\mathbf{r},\frac{\mathrm{it}}{\mathrm{c}}\right)\nabla u\left(\mathbf{r},\frac{\mathrm{it}}{\mathrm{c}}\right)\right)\cdot\mathbf{n}\, d\mathbf{r} + \lim_{\mathrm{t}\to\mathrm{t}_2}\int_{\partial r}\left(u\left(\mathbf{r},\frac{\mathrm{it}}{\mathrm{c}}\right)\nabla v\left(\mathbf{r},\frac{\mathrm{it}}{\mathrm{c}}\right) - v\left(\mathbf{r},\frac{\mathrm{it}}{\mathrm{c}}\right)\nabla u\left(\mathbf{r},\frac{\mathrm{it}}{\mathrm{c}}\right)\right)\cdot\mathbf{n}\, d\mathbf{r}. \quad (3.17)$$

Again, with the homogeneous Dirichlet, Neumann and mixed boundary conditions, the expression is equal to zero and A is self-adjoint.

In addition, we will sometimes use the so-called *Mur's absorbing boundary condition*, which is stated as follows:

$$\frac{\partial x}{\partial \mathrm{t}} + c\xi\nabla x\cdot\mathbf{n} = 0, \quad (3.18)$$

where ξ denotes the "specific impedance" coefficient. For large $c\xi$, this condition approximates homogeneous Neumann boundary condition, whereas for the small values it yields $x(\mathbf{r},\mathrm{t}) \approx x(\mathbf{r},\mathrm{t}+\mathrm{t}')$. Since the initial conditions are also homogeneous, this implies that the equation (3.17) is again approximately equal to zero. Hence, in all these cases there exist an inverse integral operator to (3.15), defined by (3.7).

An important feature of the wave equation is *finite speed of propagation*, embodied in the coefficient $c(\mathbf{r},\mathrm{t})$. An illustrative example is the so-called *Cauchy problem*, for which the initial data is prescribed by (3.16), $\Gamma = \mathbb{R}^{\mathrm{d}}$ and $z = 0$. The explicit solutions are well-known for the case where $c(\mathbf{r},\mathrm{t}) = \mathrm{c}$ is constant. For instance, when $\mathrm{d} = 3$, the solution is

$$x(\mathbf{r},\mathrm{t}) = \frac{1}{4\pi}\frac{\partial}{\partial \mathrm{t}}\left(\mathrm{t}\int_{\|\boldsymbol{\mu}\|_2=1} i(\mathbf{r}+\mathrm{ct}\boldsymbol{\mu})\, d\sigma(\boldsymbol{\mu})\right) + \frac{\mathrm{t}}{4\pi}\int_{\|\boldsymbol{\mu}\|_2=1} i_{\mathrm{t}}(\mathbf{r}+\mathrm{ct}\boldsymbol{\mu})\, d\sigma(\boldsymbol{\mu}), \quad (3.19)$$

where $\sigma(\cdot)$ is a *spherical measure*[6]. This shows that the information travels from the initial conditions to $x(\mathbf{r},\mathrm{t})$ at the speed c. Moreover, the solution only depends on the data on the sphere of radius ct, which is known as *Huygens principle*. Intuitively, information propagates as a sharp wave front, leaving no trace behind. This holds true for any odd spatial dimension d - for even dimensions, the information still propagates at speed c, but the solution will depend

[6] Informally, spherical measure can be seen as the ratio between the area of the part of the sphere and the total area of the sphere.

on the "ball" of radius ct (the wave front has a "tail" [7]).

Finally, if solutions of the wave equation are harmonic, *i.e.* they can be written in the form $x(\mathbf{r},\mathrm{t}) = \Re(x_{\mathbf{r}}(\mathbf{r})e^{-\mathrm{i}\omega\mathrm{t}})$, where ω denotes the angular frequency, the wave equation can be reduced to time-independent form:

$$\triangle x_{\mathbf{r}}(\mathbf{r}) - \frac{\omega^2}{\mathrm{c}^2} x_{\mathbf{r}}(\mathbf{r}) = z_{\mathbf{r}}, \tag{3.20}$$

also known as the *Helmholtz equation*. The Helmholtz equation can be easily derived by applying Fourier transform to the wave equation (hence, the time-domain solution is obtained by applying the inverse Fourier transform to the solution of the Helmholtz equation). In this work, however, we consider only the standard linear wave equation in time, given by (3.15).

3.3 Regularization of physics-driven inverse problems

Let us now formally define a physics-driven inverse problem. Assume that we are measuring a physical quantity $x(\boldsymbol{\omega})$ only in a *part* of its domain Ω. We know that x (approximately) satisfies a partial differential equation (3.4) with prescribed boundary conditions, however we do not know the forcing term z. Moreover, x has an integral representation (3.6) provided by the associated Green's functions (3.7) (assuming that we have somehow computed the Green's functions already). Our goal is to recover, or estimate x on the entire domain Ω such that the solution complies with the measurements. Likewise, we may recover, or estimate, the forcing term z and evaluate x using (3.6). Without additional information, this abstract problem is, generally, severely ill-posed. This is simply a consequence of the fact that the number of degrees of freedom in the problem is too large (theoretically, infinite).

The crucial fact that we aim at exploiting is that the $z(\boldsymbol{\omega})$-term is usually sparse in some physical dimension(s), *i.e.* $Ax(\boldsymbol{\omega})$ is *mostly* equal to zero within Ω. The subdomain $\underline{\Omega} \subset \Omega$ for which z is non-zero is usually a *source*, *sink* or some other type of singularity. In other words the volume occupied by singularities is considerably smaller than the overall volume of the domain. This can be interpreted as continuous-domain sparsity of $z(\boldsymbol{\omega})$ or cosparsity of $x(\boldsymbol{\omega})$. Furthermore, the integral operator D in (3.9) is reminiscent of the synthesis dictionary, while the differential operator A in (3.4) resembles the analysis operator. Thereby we term this concept a *physics-driven regularization*.

The linear observation operator M is usually dictated by the measuring device, hence only partially under our control. For the physics-driven problems we consider here, the instruments directly measure the physical quantity x, *i.e.* $Mx = y$, where the functional y is a realization of the random variable Y, defined as follows:

[7] Actually, Huygens principle *approximately* holds in even dimensions, due to different integral kernel.

$$\mathrm{Y}(\boldsymbol{\omega}) = \begin{cases} x(\boldsymbol{\omega}) + \mathrm{E} & \text{for } \boldsymbol{\omega} \in \Upsilon \subset \Omega \setminus \underline{\Omega}, \\ 0 & \text{otherwise.} \end{cases}$$

Here, E represents a random variable, modeling the *measurement error*. The error is due to, *e.g.* lowpass filtering, aliasing and the instrumentation noise. The set Υ is assumed to be known beforehand (for instance, set of electrode positions in EEG).

Since the aim is to exploit low-complexity regularizations, which are commonly defined in a discrete setting, we need to *discretize* all the involved quantities. There are many discretization techniques to achieve this, with different advantages and drawbacks (in appendices B.1 and B.2 we provide two examples of discretization using the *Finite Difference Method (FDM)*). First, the continuous domain Ω is replaced by a set of discrete coordinates of dimension n (for the sake of simplicity, we also denote the discretized domain by Ω). The discretized differential and integral operators, A and D, are represented in matrix forms as $\mathbf{A} \in \mathbb{R}^{\mathsf{n}\times\mathsf{n}}$ and $\mathbf{D} \in \mathbb{R}^{\mathsf{n}\times\mathsf{n}}$, respectively. Regardless of the employed discretization method, we expect that $\mathbf{D} \approx \mathbf{A}^{-1}$. This should be no surprise, since the Green's functions embodied in D are obtained by discretizing the impulse responses of the operator A. The discrete observation data $y \to \mathbf{y} \in \mathbb{R}^{\mathsf{m}}$ is obtained by downsampling a discretized signal $x \to \mathbf{x} \in \mathbb{R}^{\mathsf{n}}$ by means of a row-reduced identity matrix $\mathbf{M} \in \mathbb{R}^{\mathsf{m}\times\mathsf{n}}$. An additive noise model is still assumed. Likewise, the right hand side z of (3.4) is discretized into $\mathbf{z} \in \mathbb{R}^{\mathsf{n}}$, and $\mathbf{Ax} = \mathbf{z}$ (analogously, $\mathbf{Dz} = \mathbf{x}$) holds.

Cosparse regularization (*implicitly* encouraging homogeneous boundary/initial conditions) reads as follows:

$$\underset{\mathbf{x}}{\text{minimize}}\ \|\mathbf{Ax}\|_0 + f_d(\mathbf{Mx} - \mathbf{y}). \tag{3.21}$$

Here, $f_d(\cdot)$ denotes a measure of data-fidelity in the discrete context (*e.g.* the sum of square differences). The goal of analogous sparse regularization is to recover the discretized right hand side $\mathbf{z}$ by solving the optimization problem

$$\underset{\mathbf{z}}{\text{minimize}}\ \|\mathbf{z}\|_0 + f_d(\mathbf{MDz} - \mathbf{y}), \tag{3.22}$$

where $f_d(\cdot)$ is the same penalty functional as in (3.21).

As mentioned in Chapter 1, subsection 1.2.2, minimization of the ℓ_0 "norm" is intractable. Because of that, we are relaxing the two problems noted above, in the sense that ℓ_0 is replaced by a *convex* sparsity promoting penalty f_r (such as the ℓ_1 norm). The choice of f_d is also problem-dependent - for example, it can be the characteristic function of an affine set:

$$\chi_{\{\mathbf{v}|\mathbf{Mv}=\mathbf{y}\}}(\mathbf{v}) = \begin{cases} 0, & \text{if } \mathbf{Mv} = \mathbf{y}, \\ +\infty, & \text{otherwise.} \end{cases} \tag{3.23}$$

Note that, due to simple structure of the matrix $\mathbf{M}$, one could easily build a (sparse) null space basis $\texttt{null}(\mathbf{M}) = \mathbf{I} - \mathbf{M}^\mathsf{T}\mathbf{M}$, *i.e.* its columns are the "missing" rows of the row-reduced identity matrix $\mathbf{M}$. Thus, the null space method, mentioned in section A.3, can be exploited.

In many settings, properties of a part of the region $\Omega_1 \subset \{\Omega \setminus \underline{\Omega}\}$ are known beforehand (*e.g.* $\Omega_1 = \partial\Omega$ with homogeneous boundary conditions). Consequently, the rows of the discretized analysis operator $\mathbf{A}$ can be split into $\mathbf{A}_{\Omega_1}$ and $\mathbf{A}_{\Omega_1^c}$. Taking this into account one can envision a separable problem of the form $f_r(\mathbf{A}_{\Omega_1}\mathbf{x}) + f_c(\mathbf{A}_{\Omega_1^c}\mathbf{x})$. The $f_c(\cdot)$ penalty can be, for instance, another characteristic function $\chi_{\mathbf{A}_{\Omega_1^c}\mathbf{x}=\mathbf{0}}$. Accordingly, in the synthesis context, this leads to an equivalent problem of the form $f_r(\mathbf{z}_{\Omega_1}) + f_c(\mathbf{z}_{\Omega_1^c})$, where $\mathbf{z}_{\Omega_1}$ and $\mathbf{z}_{\Omega_1^c}$ denote the corresponding subvectors of $\mathbf{z}$.

Other variants can be envisioned to encode other types of prior knowledge at different levels of precision. In the same manner, the framework can be extended to account for multiple constraints, by taking f_c to be the sum of convex functionals $f_c = \sum_{\mathrm{i}=1}^{\mathrm{f}} f_{c_\mathrm{i}}$. However, we assume that there exist a *feasible point* $\hat{\mathbf{x}}$ (accordingly, $\hat{\mathbf{z}}$), such that all imposed constraints are satisfied.

Once equipped with penalties that reflect available prior knowledge, the optimization problems corresponding to sparse analysis and sparse synthesis regularization read as:

$$\underset{\mathbf{x}}{\text{minimize}}\; f_r(\mathbf{A}\mathbf{x}) + f_d(\mathbf{M}\mathbf{x} - \mathbf{y}) + f_c(\mathbf{C}\mathbf{x} - \mathbf{c}). \tag{3.24}$$

$$\underset{\mathbf{z}}{\text{minimize}}\; f_r(\mathbf{z}) + f_d(\mathbf{M}\mathbf{D}\mathbf{z} - \mathbf{y}) + f_c(\mathbf{C}\mathbf{D}\mathbf{z} - \mathbf{c}), \tag{3.25}$$

Here f_r is an objective, while f_d and f_c are the (extended-valued[8]) penalty functionals for the measurements and additional problem constraints, respectively. Since $\mathbf{A}\mathbf{x} = \mathbf{z}$, and $\mathbf{D} = \mathbf{A}^{-1}$, **the two problems are nominally equivalent**, and one can straightforwardly use the solution of one of them to recover the solution of another. Finally, since all penalty functionals are convex, and the feasible set is assumed non-empty, it is theoretically possible to find global minimizers of the two optimization problems.

3.4 SDMM for physics-driven (co)sparse inverse problems

Now we discuss one way to practically address convex optimization problems (3.24) and (3.25) arising from the physics-driven framework. Namely, we will apply the modified SDMM [67] algorithm, introduced in section A.2 of appendix A, to these problems. SDMM is a first-order optimization algorithm for solving convex problems of the form

$$\underset{\mathbf{x},\mathbf{z}_\mathrm{i}}{\text{minimize}} \sum_{\mathrm{i}=1}^{\mathrm{f}} f_\mathrm{i}(\mathbf{z}_\mathrm{i}) \text{ subject to } \mathbf{H}_\mathrm{i}\mathbf{x} - \mathbf{h}_\mathrm{i} = \mathbf{z}_\mathrm{i}, \tag{3.26}$$

by iterating the update steps given in (A.11).

[8] Allowed to take $+\infty$ values to encode hard constraints.

This formulation makes the algorithm suitable for solving the regularized physics-driven problems (3.24) and (3.25). For the analysis-regularized problem (3.24), we set

- $\mathbf{H}_1 = \mathbf{A}$ (physics: encoding PDE),
- $\mathbf{H}_2 = \mathbf{M}$ (measurement),
- $\mathbf{H}_3 = \mathbf{C}$ (additional constraints).

For the synthesis version (3.25), the problem is parametrized by

- $\mathbf{H}_1 = \mathbf{I}$ (the estimate is sparse[9]),
- $\mathbf{H}_2 = \mathbf{MD}$ (physics and measurement: subsampled Green's function basis),
- $\mathbf{H}_3 = \mathbf{CD}$ (additional constraints).

In both cases we have $\mathbf{h}_1 = \mathbf{0}$, $\mathbf{h}_2 = \mathbf{y}$ and $\mathbf{h}_3 = \mathbf{c}$, and the functionals f_i are replaced by f_r, f_d and f_c. The matrix $\mathbf{C}$ usually encodes constraints in diffuse or sparse domain (in the latter case the product $\mathbf{CD}$ would be a row-reduced identity matrix).

3.4.1 Weighted SDMM

The functionals f_i encode both an objective and constraints. However, the least squares step treats all $\mathbf{z}_i$ equally, meaning that $\mathbf{x}$ is not guaranteed to satisfy the constraints. Moreover, in practice, $\mathbf{x}$ is often far from being feasible. To alleviate this problem, an intuitive solution is to set different weights to different blocks $\left[\mathbf{H}_i\mathbf{x} + \mathbf{u}_i^{(k)} - \mathbf{z}_i^{(k+1)}\right]$ of the sum of squares in (A.11). This weighting can be seen as choosing different SDMM multipliers ρ_i for different functionals $f_i(\cdot)$:

$$\mathbf{z}_i^{(k+1)} = \operatorname{prox}_{\frac{1}{\rho_i} f_i}\left(\mathbf{H}_i\mathbf{x}^{(k)} - \mathbf{h}_i + \mathbf{u}_i^{(k)}\right), \tag{3.27}$$

$$\mathbf{x}^{(k+1)} = \arg\min_{\mathbf{x}} \sum_{i=1}^{f} \frac{\rho_i}{2} \|\mathbf{H}_i\mathbf{x} - \mathbf{h}_i + \mathbf{u}_i^{(k)} - \mathbf{z}_i^{(k+1)}\|_2^2, \tag{3.28}$$

$$\mathbf{u}_i^{(k+1)} = \mathbf{u}_i^{(k)} + \mathbf{H}_i\mathbf{x}^{(k+1)} - \mathbf{h}_i - \mathbf{z}_i^{(k+1)}. \tag{3.29}$$

To derive this, consider the following splitting (with $\rho = 1$):

$$\begin{aligned} &\underset{\mathbf{x},\mathbf{z}_i}{\text{minimize}} \sum_{i=1}^{f} f_i\left(\frac{1}{\sqrt{\rho_i}}\mathbf{z}_i\right), \\ &\text{subject to } \sqrt{\rho_i}\,(\mathbf{H}_i\mathbf{x} - \mathbf{h}_i) = \mathbf{z}_i. \end{aligned} \tag{3.30}$$

[9] $\mathbf{I}$ denotes the identity matrix.

Then, the SDMM iterates are defined as follows:

$$\begin{aligned}
\mathbf{z}_i^{(k+1)} &= \arg\min_{\mathbf{z}_i} \left(f_i\left(\frac{1}{\sqrt{\rho_i}}\mathbf{z}_i\right) + \frac{1}{2}\left\| \sqrt{\rho_i}\left(\mathbf{H}_i\mathbf{x}^{(k)} - \mathbf{h}_i\right) - \mathbf{z}_i + \mathbf{u}_i^{(k)} \right\|_2^2 \right), \\
\mathbf{x}^{(k+1)} &= \arg\min_{\mathbf{x}} \sum_{i=1}^{f} \frac{1}{2}\left\| \sqrt{\rho_i}\left(\mathbf{H}_i\mathbf{x} - \mathbf{h}_i\right) + \mathbf{u}_i^{(k)} - \mathbf{z}_i^{(k+1)} \right\|_2^2, \\
\mathbf{u}_i^{(k+1)} &= \mathbf{u}_i^{(k)} + \sqrt{\rho_i}\left(\mathbf{H}_i\mathbf{x}^{(k+1)} - \mathbf{h}_i\right) - \mathbf{z}_i^{(k+1)}.
\end{aligned} \tag{3.31}$$

By abuse of notation $\mathbf{z}_i := \frac{1}{\sqrt{\rho_i}}\mathbf{z}_i$, $\mathbf{z}_i^{(k)} := \frac{1}{\sqrt{\rho_i}}\mathbf{z}_i^{(k)}$ and $\mathbf{u}_i^{(k)} := \frac{1}{\sqrt{\rho_i}}\mathbf{u}_i^{(k)}$, we arrive at the expressions (3.27) - (3.29).

Empirically, to quickly attain a feasible point and preserve convergence speed, it seems appropriate to adjust the multipliers ρ_1, ρ_2 and ρ_3 relative to the penalty parameters. Our strategy is to first fix the value $\rho_1 = \rho$, and then set i) $\rho_2 = \max(\rho, \rho/\varepsilon)$ and $\rho_3 = \max(\rho, \rho/\sigma)$ if f_2 and f_3 are the indicator functions bounding a norm of their arguments by ε and σ, respectively; or ii) $\rho_2 = \max(\rho, \rho\sqrt{\varepsilon})$ and $\rho_3 = \max(\rho, \rho\sqrt{\sigma})$ if f_2 and f_3 are norm-squared penalties weighted by ε and σ, respectively. Other types of penalties are allowed, but may require different weighting heuristics.

3.4.2 The linear least squares update

For problems of modest scale, the least squares minimization step in (3.28) can (and should) be performed exactly, by means of a direct method, *i.e.* matrix inversion. For computational efficiency, this requires relying on matrix factorization such as the Cholesky decomposition. An important observation is that Cholesky decomposition is band-preserving, meaning that, if all $\mathbf{H}_i$ are banded, the Cholesky factor of the coefficient matrix will be also banded [29, Theorem 1.5.1]. Further, a desirable property is to obtain Cholesky factors essentially as sparse as the factorized matrix. Many efficient algorithms heuristically achieve this goal (such as the sparse Cholesky decomposition [63] used in our computations). However, the number of non-zero elements of the factorized matrix is a lower bound on the number of non-zero elements of its Cholesky decomposition [74, Theorem 4.2], and only sparse matrices would benefit from this factorization.

For large scale problems one needs to resort to iterative algorithms and approximate the solution of (3.28). An important advantage of ADMM is that it ensures convergence even with inexact computations of intermediate steps, as long as the accumulated error is finite [90]. Moreover, these algorithms can be usually initialized (*warm-started*) using the estimate from the previous ADMM iteration, which can have a huge influence on the overall speed of convergence. The downside of the weighted SDMM is that conditioning of the weighted matrix $\mathbf{H}$ is usually worse than in the unweighted setting, which is why applying the standard conjugate gradient method to the normal equations $\mathbf{H}^T\mathbf{H}\mathbf{x} = \mathbf{H}^T\left(\mathbf{z}^{(k+1)} + \mathbf{h} - \mathbf{u}^{(k)}\right)$ should be avoided. Instead, we suggest using the Least Squares Minimal Residual (LSMR) method [104],

which is less sensitive to matrix conditioning. Assuming no a priori knowledge on the structure of $\mathbf{H}$, one may use diagonal (right) preconditioner (recommended by the authors), whose elements are reciprocal to the ℓ_2-norms of the columns of $\mathbf{H}$. Even though there exist more efficient preconditioners (such as incomplete Cholesky / LU factorizations), two advantages are provided by this diagonal preconditioner: i), there are no issues with stability, as with the incomplete preconditioners, and ii), it can be efficiently computed in the function handle implementation of the synthesis problem (for which only $\mathbf{MD}$ exists in the matrix form).

3.5 Computational complexity

Having $\mathbf{Ax} = \mathbf{z}$, it can be easily shown that the above-described SDMM algorithm yields *numerically identical* solutions for the synthesis and the analysis problems, as long as all evaluations in (3.27), (3.28) and (3.29) are exact (this corresponds to the usage of direct methods, described in the previous subsection). However, as detailed below, the overall cost of the analysis minimization is driven by that of the multiplication with $\mathbf{A}$ and its transpose, which is $O(n)$ thanks to the sparsity of the analysis operator $\mathbf{A}$ (or $O(bn)$, where b is the band of $\mathbf{A}$, for direct computation of (3.28)). This is in stark contrast with synthesis minimization, whose cost is dominated by much heavier $O(mn)$ multiplications with the dense matrix $\mathbf{MD}$ and its transpose (analogously, $O(n^2)$ if the direct methods are used). The density of the dictionary $\mathbf{D}$ is not surprising - it stems from the fact that the physical quantity modeled by $\mathbf{x}$ is spreading in the domain of interest (otherwise, we would not be able to obtain remote measurements). As a result, and as will be confirmed experimentally in chapters 4 and 6, *the analysis minimization is computationally much more efficient.*

3.5.1 Initialization costs

Generating the analysis operator $\mathbf{A} \in \mathbb{R}^{n\times n}$ in matrix form is problem dependent, but usually of moderate cost. The cost of building the transfer matrix $\mathbf{M}$ is negligible, since it often corresponds to simple acquisition models (*e.g.* the row-reduced identity). To efficiently compute *the reduced dictionary* $\mathbf{G} = \mathbf{MD} \in \mathbb{R}^{m\times n}$, which satisfies $\mathbf{A}^\top\mathbf{G}^\top = \mathbf{M}^\top$, one needs to solve m linear systems $\mathbf{A}^\top\mathbf{g}_j = \mathbf{m}_j$ of order n, where $\mathbf{g}_j^\top$ and $\mathbf{m}_j^\top \in \mathbb{R}^n$ are the rows of $\mathbf{G}$ and $\mathbf{M}$, respectively. Thus, it adds at least[10] $O(mn)$ operations on the price of computing $\mathbf{A}$ and $\mathbf{M}$, unless an analytical expression of Green's functions is available.

If the direct method is used to solve the linear least squares step, classical algebraic manipulations show that we first need to compute the coefficient matrix $\mathbf{H_A} = \rho_1\mathbf{A}^\top\mathbf{A} + \rho_2\mathbf{M}^\top\mathbf{M}$, in the analysis case, or $\mathbf{H_S} = \rho_1\mathbf{I} + \rho_2(\mathbf{MD})^\top\mathbf{MD}$, in the synthesis case. Due to the sparse structure of $\mathbf{A}$ and $\mathbf{M}$, the former can be computed in $O(n)$, while the latter requires $O(n^2m^2)$ operations. The Cholesky factorization requires $O(n^3)$ operations in general, but this is significantly reduced for sparse matrices [74]. However, it is known [263] that the minimum fill-in problem –finding

[10] Assuming the favorable scenario where the linear system can be solved in $O(n)$ operations.

the best sparsity-preserving permutation– is an NP-complete problem, thus all available algorithms are (usually very efficient) heuristics. Therefore, it is very difficult to derive tight bounds on initialization complexity.

3.5.2 Iteration costs

Each iteration of SDMM involves three types of operations.

Component-wise scalar operations: evaluation of proximal operators and update of scaled Lagrangian multipliers $\mathbf{u}_i$ given $\mathbf{H}_i\mathbf{x}^{(k)}$ and $\mathbf{H}_i\mathbf{x}^{(k+1)}$. These have O(n) complexity with respect to the problem size n, since they involve only component-wise thresholding and vector norm computations (see section A.3).

Matrix-vector products: computation of $\mathbf{H}_i\mathbf{x}^{(j)}$. For analysis, the matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ are both sparse with O(n) nonzero entries, hence these matrix-vector products also have an O(n) complexity. This is also the case for $\mathbf{H}_3$ in the considered scenarios. In contrast, for synthesis, the matrix $\mathbf{H}_2 = \mathbf{MD}$ is dense, reflecting the discretized Green's functions. The matrix-vector product with this matrix is of O(mn), and it dominates the cost of all other matrix-vector products.

Least squares: the solution of problem (3.28). When an iterative solver is used to address the least squares step, we assume that a properly preconditioned and warm-started iterative method would terminate to sufficient accuracy in considerably less than n iterations. The overall computational complexity is governed by the cost of matrix-vector products $\mathbf{H}_i\mathbf{v}$ and $(\mathbf{H}_i)^\mathsf{T}\mathbf{w}$ for some intermediate vectors $\mathbf{v}$, $\mathbf{w}$, which, as just seen, have very different complexities in the analysis and synthesis settings.

When a direct method is used to evaluate the linear least squares step the complexity analysis is more delicate. Since the matrix $\mathbf{H}_\mathrm{A}$ is usually banded for both acoustic and EEG problems, we know that regular Cholesky decomposition will usually produce factors, which are much sparser in the analysis than in the synthesis case. However, due to the mentioned NP-hardness of the minimum fill-in problem, it is impossible to exactly evaluate the sparsity of the yielded *sparse* Cholesky factorization $\mathbf{P}^\mathsf{T}\mathbf{H}_\mathrm{A}\mathbf{P}$ ($\mathbf{P}$ is a permutation matrix), and estimate the computational complexity. Yet, it is reasonable to assume that the permuted coefficient matrix would yield even sparser Cholesky factor than $\mathbf{L}_\mathrm{A}$. Therefore, one may expect the analysis model to be computationally much more efficient. This result was checked through simulations.

3.5.3 Memory aspects

Another view on computational scalability is through memory requirements. For the synthesis model, assuming the general case where the analytical expression of the Green functions is not available, the least requirement is storing the (m × n) matrix $\mathbf{MD}$ in memory (to avoid computational overhead, it is usually necessary to also store $(\mathbf{MD})^\mathsf{T}$). Hence, the minimum storage

requirement for the synthesis case is $O(mn)$. This cost can quickly become prohibitive, even for modern computers with large amounts of RAM. On the other hand, memory requirement for storing the analysis operator is $O(n)$.

Evaluating the storage requirements of sparse Cholesky factorization is not viable, due to the NP-hardness of the fill-in problem. It is, however, presumed (and experimentally checked in chapter 4) that it requires substantially less than $O(n^2)$ memory units, which is the requirement for the regular (synthesis) Cholesky factor.

3.5.4 Structured matrices

In some cases, a special matrix structure can be exploited in order to further reduce storage size and computational effort. Our goal in this work is not to exploit such matrix structures, but to emphasize that in the analysis case, one can "naively" manipulate the matrix (as long as the discretization has local support) and still gain in terms of computational and storage resources. This is not necessarily the case with the synthesis approach, and even if possible, requires specialized techniques in order to exploit particular matrix structure. For example, in the case where $\mathbf{D} \in \mathbb{R}^{n\times n}$ is a Toeplitz matrix, this means first embedding it in a circulant form, and then applying Fast Fourier Transform at cost $O(n \log n)$ [111] (even in this case the analysis approach is cheaper, since it requires $O(n)$ calculations).

Finally, note that while such specific matrix structures in $\mathbf{A}$ and $\mathbf{D}$ may be useful when iterative algorithms are used to solve the least squares step in SDMM, these become useless for the direct computation of linear systems with matrices $\mathbf{H}_A$ and $\mathbf{H}_S$ (since forming the normal equations usually disturbs the structure). For all these reasons, in the subsequent experiments, we disregard any additional structures of the involved matrices except their sparsity patterns.

3.5.5 Conditioning

Additionally, for the proposed weighting scheme, the matrix $\mathbf{H}_A$ is usually better conditioned than $\mathbf{H}_S$. The rationale comes from the fact that the applied multipliers $\rho_1 \leq \rho_2$ usually assign a large weight to the diagonal elements of the matrix $\mathbf{H}_A$. And since $\mathbf{H}_A = \mathbf{A}^T \mathbf{H}_S \mathbf{A}$, one can see this as *preconditioned* synthesis approach.

Moreover, in the analysis case and common constraints, such as when f_d is a characteristic function of an affine or the ℓ_2 norm-constrained set ($\mathbf{Mx} = \mathbf{y}$ or $\|\mathbf{Mx} - \mathbf{y}\|_2 \leq \varepsilon$), one can replace $\mathbf{H}_2 = \mathbf{M}$ with $\mathbf{H}_2 = \mathbf{I}$ and still maintain a simple evaluation[11] of (3.27). However, the problem (3.28) is now much better conditioned, since $\mathbf{H}_2 = \mathbf{I}$ acts as Tikhonov regularization, and the proposed weighting often yields a diagonally dominant coefficient matrix $\mathbf{H}_A$. In principle, the same "trick" could be used in the synthesis variant as well, provided that the matrix $\mathbf{H}_2 = \mathbf{MD}$ forms a tight frame (which is always the case for $\mathbf{H}_2 = \mathbf{M}$ for the analysis approach).

[11] Notice that in this case the linear mapping $\mathbf{M}$ is incorporated *within* the proximal operator.

3.6 Summary

In this chapter we introduced the general framework for sparse regularizations of physics-driven inverse problems.

We have discussed the physical phenomena modeled by either linear partial differential equations, or the integral equations through the appropriate Green's functions. It was emphasized that when a PDE is expressed by a self-adjoint operator, the corresponding Green's functions are unique and the integral form can be seen as the inverse operator of the concerned PDE. Two important cases were discussed: Poisson's equation and the linear wave equation, both of which are self-adjoint systems when accompanied with common homogeneous boundary conditions.

Furthermore, when there is a side knowledge in terms of a low number of field-generating singularities, a natural connection between the two (differential and integral) representations and sparse regularizations can be established. Models defined by linear PDEs are linked to the sparse analysis data model, whilst models defined by the Green's functions are related to the sparse synthesis data model. Since the representations are equivalent, the two data models are also equivalent, but their numerical properties are different. Due to the inherited sparsity of the discretized PDE (embodied in the analysis operator), the analysis approach scales much more favorably than the equivalent problem regularized by the sparse synthesis model. Finally, we have presented a first order optimization algorithm for solving the problems yielded by the according convex relaxations. This ADMM variant, termed Weighted SDMM, can be straightforwardly applied to both analysis- and synthesis-regularized problems.

In the upcoming chapters we will see how the physics-driven framework can be applied to concrete problems, namely acoustic and brain source localization. The Weighted SDMM algorithm will be used for the numerical verification of predicted numerical differences between the analysis and synthesis regularizations.

4 (Co)sparse acoustic source localization

Acoustic or sound source localization is the problem of determining the position of one or more sources of sound based solely on microphone recordings. The knowledge of sound source locations is a very valuable piece of information, that can be exploited in multiple contexts. In speech and sound enhancement, a source position estimate is used to perform noise reduction, by means of *beamforming* techniques [264, 107, 64, 23, 245]. The noise, in this case, is considered as any unwanted source of sound (for example, degradations attributed to reverberation). Another application of acoustic localization is tracking of sound sources [3, 252, 174, 65], which may be used to automatize camera steering when accompanied with other signal modalities [250]. Robotics is another area of engineering that exploits sound source localization techniques [243, 184, 185, 76], as well as seismic [177, 62] and medical [233, 213] imaging, among many others.

The acoustic localization problem we are interested in is the following:

> *Given an array of* m *omnidirectional microphones, with a known geometry, determine locations of* k *point monopole sound sources, within an enclosed spatial environment.*

For various reasons, acoustic source localization is a challenging task, and, as such, has provoked significant research efforts. There exist a plethora of methods that address this problem more or less successfully. In this Chapter, we propose a new one, based on physics-driven regularization, and argue its advantages and shortcomings.

The content of the chapter is as follows: in the first section we briefly recall the physics of idealized sound propagation in air. The second Section discusses some well-known and widely used algorithmic approaches to solve the acoustic localization problem. In the third Section, we formulate the problem as an ill-posed physics-driven inverse problem stabilized by sparse regularizations. In the fourth, final Section, we provide comprehensive experimental results based on numerical simulations. The written material is partially based on publications [140, 139].

4.1 Physical model of sound propagation

In this section, we provide a very brief and simplified introduction to some fundamental concepts of air acoustics, based on these references: [152, 153, 43, 100, 211].

4.1.1 The acoustic wave equation

Sound is produced by the vibration of small "volume particles" of the propagating medium. A volume particle is composed of sufficiently many infinitesimal particles (*e.g.* molecules), such that their individual, irregular thermal-induced motions are somewhat averaged-out (this includes, *e.g.* Brownian motion). Sound waves are categorized into transversal (which are orthogonal to the direction of propagation) and longitudinal (parallel to the direction of propagation), such as presented in figure 4.1. However, in gases (which are the only propagating medium of interest in this work), the transversal component of sound waves is negligible, at least away from the boundary. Therefore, the physical model of sound propagation in air is a special case of more general sound propagation models.

Sound wave amplitude in fluids is usually expressed through *particle displacement* $\mathbf{s}$, *particle velocity* $\mathbf{v}$, *sound pressure* p and *density* ρ. They are all functions of position $\mathbf{r}$ and time t (therefore, their domain variable is $\boldsymbol{\omega} \in \Gamma \times \mathbb{R} \subset \mathbb{R}^n \times \mathbb{R}$). These quantities are mutually related,

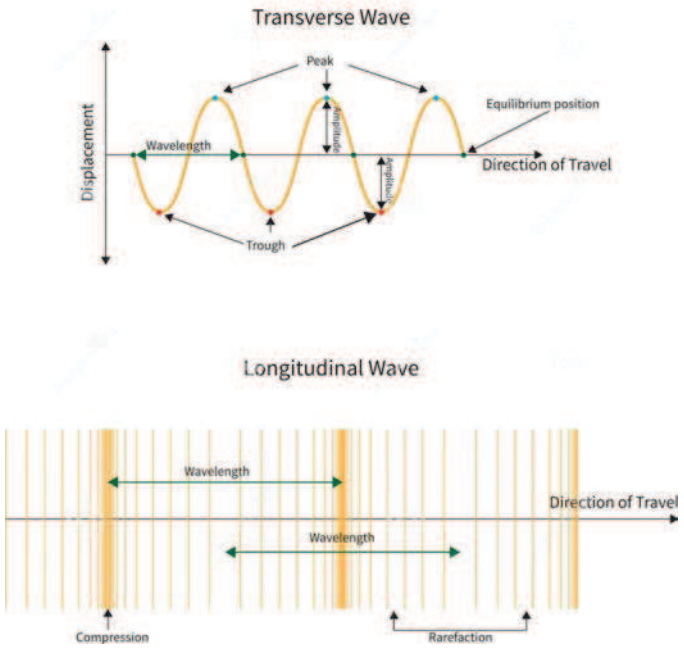


Figure 4.1 – Transverse and longitudinal sound waves[1].

[1] Image downloaded from www.dreamstime.com.

such that:

$$\mathbf{v} = \frac{\partial \mathbf{s}}{\partial \mathrm{t}} \tag{4.1}$$

$$\nabla_{\mathbf{r}} p = -\rho_0 \frac{\partial \mathbf{v}}{\partial \mathrm{t}} \tag{4.2}$$

$$\rho_0 \nabla \cdot \mathbf{v} = -\frac{\partial \rho}{\partial \mathrm{t}}, \tag{4.3}$$

where equation (4.2) is known as *conservation of linear momentum*, while equation (4.3) is known as *conservation of mass*.

The quantity ρ_0 represents the medium density in the equilibrium state - when the field is not excited. This value is computed from the *Equation of state of ideal gas*:

$$p_0 = \rho_0 \mathrm{R} \mathrm{T}_0, \tag{4.4}$$

where p_0 represents the pressure at temperature T_0, while the factor R is equal to $287\,\mathrm{J\,kg^{-1}\,K^{-1}}$ in air. For example, at $\mathrm{T}_0 = 273\mathrm{K}$ and $p_0 = 100\mathrm{kPa}$, the reference temperature and the reference pressure, respectively, $\rho_0 \approx 1.2754\mathrm{kg/m^3}$. This equation also indicates the connection among the variations of temperature, pressure and density (the increase in temperature increases pressure, and vice-versa).

Unless the acoustic event is extremely strong (*e.g.* explosions), the absolute change in acoustic pressure and density is small compared to their equilibrium values ($|p_\mathrm{t} - p_0| = |\tilde{p}| \ll p_0$, $|\rho_\mathrm{t} - \rho_0| = |\tilde{\rho}| \ll \rho_0$). With some other approximations and linearizations, one arrives at the expression for sound pressure:

$$p = \left(\frac{d p_\mathrm{t}}{d \rho_\mathrm{t}}\right)_{\rho_0} \rho = \mathrm{c}^2 \rho, \tag{4.5}$$

where c^2 is assumed constant as long as the magnitude of the acoustic event is not too high. We assume that $c(\mathbf{r}, \mathrm{t})$ is a *slowly* varying function of space and/or time, therefore, the isotropic[2] condition still approximately holds. This assumption is physically relevant - for instance, the air-conditioning and heating devices introduce temperature gradients, and thus, the spatio-temporal change in sound speed, due to (4.4) and diffusion. If assumed constant, it can be estimated by the following formula:

$$\mathrm{c} = (331.3 + 0.606\,\mathrm{T})\,\mathrm{m/s}, \tag{4.6}$$

where T is the temperature in Celsius. Given the relation (4.5), one can substitute ρ in (4.3) yielding $\rho_0 \nabla \cdot \mathbf{v} = -\frac{1}{\mathrm{c}^2} \frac{\partial p}{\partial \mathrm{t}}$. From conservation of momentum, we obtain $\nabla \cdot \nabla_{\mathbf{r}} p = \triangle p = -\frac{\partial}{\partial \mathrm{t}} \left(\rho_0 \nabla \cdot \mathbf{v}\right)$. Since we already have the expression under parenthesis in terms of pressure, we

[2] $c(\mathbf{r}, \mathrm{t}) = \mathrm{c}$ is a constant.

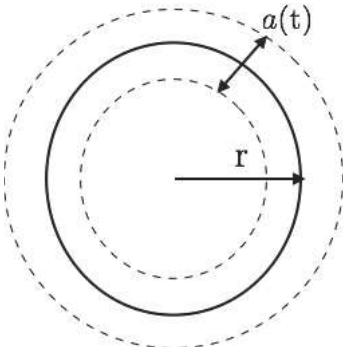


Figure 4.2 – The pulsating sphere ($a(\mathrm{t}) \ll \mathrm{r}$).

arrive at *the homogeneous acoustic wave equation*:

$$\triangle p - \frac{1}{\mathrm{c}^2}\frac{\partial^2 p}{\partial \mathrm{t}^2} = 0. \tag{4.7}$$

It models the sound pressure, which is a physically measurable quantity.

Equation (4.7) is, of course, identical to (3.15) with $x := p$, for which we have shown[3] that c is indeed the propagating speed, *i.e. the speed of sound*, in this case.

The same derivation[4] can be applied in the case when a sound source is present in the medium, by incorporating additional terms in the conservation of momentum and mass equations. Generally, this leads to the following *inhomogeneous acoustic wave equation*:

$$\triangle p - \frac{1}{\mathrm{c}^2}\frac{\partial^2 p}{\partial \mathrm{t}^2} = z + \nabla \cdot \mathbf{f}, \tag{4.8}$$

where $z := z(\mathbf{r}, \mathrm{t})$ is either $z = -\frac{\partial^2 m(\mathbf{r},\mathrm{t})}{\partial^2 \mathrm{t}}$ ($m(\mathbf{r}, \mathrm{t})$) is a function of injected (or removed) mass per unit volume) or $z = -\rho_0 \frac{\partial q(\mathbf{r},\mathrm{t})}{\partial \mathrm{t}}$ ($q(\mathbf{r}, \mathrm{t})$ is the function of velocity of volume displacement per unit volume). The term $\mathbf{f}$ is an *external force* per unit volume.

There are various types of sound sources, approximated by different mathematical models. For our purposes, we constrain the choice on a simple kind: the *point monopole source*. This type of sound source is modeled by $z(\mathbf{r}, \mathrm{t}) = z_{\mathbf{r}_0}(\mathrm{t})\delta(\mathbf{r}_0 - \mathbf{r})$ (no external forces $\mathbf{f}$ are present). For a hypothetical direct-radiation[5] monopole loudspeaker, $z_{\mathbf{r}_0}(\mathrm{t})$ is a function of volume displacement. When $z_{\mathbf{r}_0}(\mathrm{t})$ term is a function of mass injection (or removal), a monopole source can model *e.g.*, a combustion process. The source acts as an infinitesimal singularity in space - it can be seen as the limit case of the *pulsating spherical source* (figure 4.2), for $\mathrm{r} \to 0$. Monopole and spherical sources radiate equally in all spatial directions, and their induced sound wave fronts are spherical (in odd dimensions), or ball-like (in even dimensions).

[3] for the $\Gamma = \mathbb{R}^{\mathrm{n}}$.

[4] Except for *turbulent flow* sources, which depend on non-linearities [100].

[5] Without an acoustic horn.

4.1.2 Boundary conditions

The wave equation, or D'Alembert operator, constitutes a well posed problem if accompanied with Cauchy initial conditions, as noted in the previous chapter (subsection 3.2.4). In the practical setting the system is also causal, *i.e.* the time axis starts at 0, hence our temporal domain is $t \in [0, +\infty)$. Furthermore, as a convenient simplification, we will assume that the field was initially at rest, thus the compound operator A is self-adjoint. In other words:

$$p(\mathbf{r}, 0) = 0 \quad \text{and} \quad \left.\frac{\partial p(\mathbf{r}, t)}{\partial t}\right|_{t=0} = 0. \tag{4.9}$$

Finally, we are concerned with sound propagation in enclosures, thus boundary conditions have to be imposed. We envision Dirichlet, Neumann, Robin, mixed or Mur's absorbing boundary condition, defined in section 3.2, p31. They are all considered homogeneous, to ensure self-adjointness of A. The physical interpretation of these conditions is as follows:

Dirichlet: setting $p(\mathbf{r}, t) = 0$, for $\mathbf{r} \in \partial\Gamma$, corresponds to *soft wall* approximation. The reflected wave has the same amplitude as the incident wave, but the opposite phase.

Neumann: setting $\nabla_{\mathbf{r}} p(\mathbf{r}, t) \cdot \mathbf{n} = 0$, for $\mathbf{r} \in \partial\Gamma$, corresponds to *hard wall* approximation. The reflected wave has the same amplitude and phase as the incident wave.

Mixed: depending on the imposed boundary condition at the point, this corresponds to either soft or hard wall approximation.

Robin: setting $p(\mathbf{r}, t) + \alpha \nabla_{\mathbf{r}} p(\mathbf{r}, t) \cdot \mathbf{n} = 0$ makes the phase of the reflected wave dependent on the *impedance* coefficient α.

Mur's absorbing: the parameter ξ in (3.18) is termed *specific acoustic impedance* and controls the reflection coefficient, as follows [147]:

$$R(\theta) = \frac{\xi \cos\theta - 1}{\xi \cos\theta + 1}, \tag{4.10}$$

where θ is the incidence angle. This implies that for the large values of ξ, the reflected wave is almost identical to the incident wave (Neumann condition), while for the values of ξ close to 1 this condition models the absorbing boundary *for the normal incidence waves*[6]. For $\xi = 0$, Mur's boundary condition is equivalent to constant Dirichlet's boundary condition, which is equal to zero given the initial conditions (hence, it approximates soft wall). In general, coefficient ξ is frequency-dependent, but, for the sake of simplicity, we take it to be constant along frequencies.

[6] This condition is, therefore, perfect absorber in one-dimensional space. However, it is not a good approximation of *e.g.* anechoic chambers.

4.2 Traditional sound source localization methods

After this condensed recall of the physics of sound propagation, we return to sound source localization problem. Widely used approaches to tackle this problem can be roughly divided into two groups: "TDOA-based" and "beamforming" methods. We briefly describe both concepts in the text to follow. In the acoustic source localization case, the number of microphones is denoted by m, while the number of sound sources is denoted by k, to avoid introducing new notation. However, these do not correspond to the number of measurements (which is mt) nor the sparsity level in this case.

4.2.1 TDOA-based methods

TDOA is an acronym for *Time Difference of Arrival*, which is the offset between source signal arrival times for a pair of microphones in an array.

Note first that, for some applications, the acoustic localization problem can be substantially relaxed: instead of looking for the actual coordinates, we ask only for the direction of the source. This corresponds to estimating the *azimuth* (in 2D space) and the *elevation* angle (in 3D), as presented in figure 4.3a. In that case the microphones are assumed to be in the *far-field* region[7], meaning that the sound waves reach microphones in the form of a planar wave front, as shown in the figure. Then, by knowing the sound speed c and exploiting TDOA, one can easily estimate the source direction using two microphones, modulo ambiguities (*e.g.* the side of arrival).

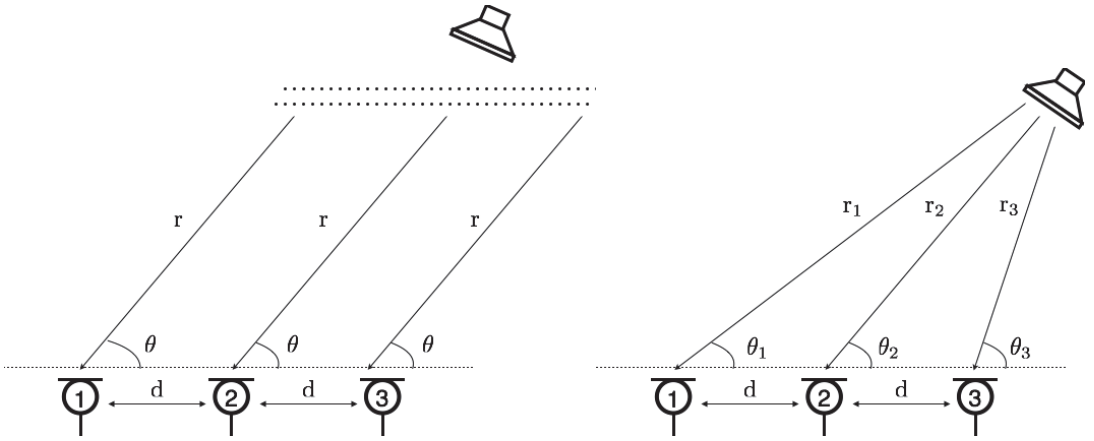
Sometimes it is also necessary to estimate the distance, and completely characterize the location of the source (*e.g.* for robot navigation). This is possible only if the microphone array is within the *near-field* of the source (figure 4.3b). Standard approach is based on *multilateration*. Theoretically, for a known speed of sound c and a given geometry of the (minimum) $\mathsf{m} = \mathsf{d}+1$ microphone array[8], it is sufficient to determine TDOAs between a reference and d other microphones, to estimate all other parameters [22]. The simplified 2D case, presented in figure 4.3b, is based on solving the following system of equations (τ_{12} and τ_{13} are TDOAs between microphones 1 and 2, and microphones 1 and 3, respectively):

$$\begin{aligned} \mathrm{r}_2 - \mathrm{r}_1 &= \mathrm{c}\tau_{12} & {\mathrm{r}_2}^2 &= {\mathrm{r}_1}^2 + \mathrm{d}^2 + 2\mathrm{r}_1\mathrm{d}\cos\theta_1 \\ \mathrm{r}_3 - \mathrm{r}_1 &= \mathrm{c}\tau_{13} & {\mathrm{r}_3}^2 &= {\mathrm{r}_1}^2 + 4\mathrm{d}^2 + 4\mathrm{r}_1\mathrm{d}\cos\theta_1, \end{aligned}$$

where the distances d between the microphones are assumed equal, for simplicity. This is easily derived by applying trigonometric rules to the geometry of the problem, and it straightforwardly generalizes to three dimensions and more microphones/sources.

[7] In practice, the far-field assumption holds if the distance between sources and a microphone array is much larger than the array aperture.

[8] Here d represents the number of spatial dimensions.



(a) Far-field case. (b) Near-field case.

Figure 4.3 – Source position relative to the microphone array.

Generally, knowledge of array geometry and range differences (*e.g.* $\mathrm{r_i}-\mathrm{r_j}$) is sufficient to recover source location. The problem is well-understood and there are closed form formulations that can be used to approximate the solution with high accuracy ([229] and references therein). Therefore, in both far-field and near-field cases, the problem simplifies to estimating TDOAs for the microphone pairs. Unfortunately, this is a non-trivial task due to the presence of noise, reverberation and/or masking (unwanted) sound sources.

Standard way of estimating TDOAs is by maximizing the cross correlation of microphone pairs, *e.g.* for pressure signals recorded on microphones 1 and 2:

$$\underset{\tau}{\text{maximize}}\; r_{12}(\tau) = \underset{\tau}{\text{maximize}} \int_{-\infty}^{\infty} p_1(\mathrm{t})\, p_2(\mathrm{t}+\tau)\, d\mathrm{t}. \tag{4.11}$$

However, this estimate is highly sensitive to noise and reverberation. To deal with these problems, *Generalized Cross Correlation (GCC)* is used instead. GCC is computed as a weighted cross correlation of signals in frequency domain. The estimate is then computed by means of inverse Fourier transform:

$$r_{12}^{\mathrm{GCC}}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} w(\omega)\phi_{12}(\omega) \exp(\mathrm{j}\omega\tau)\, d\omega, \tag{4.12}$$

where $\phi_{12}(\omega)$ is the *cross spectrum*, which is the Fourier transform of $r_{12}(\tau)$. In practice, the integrals are approximated by finite sums, *i.e.* by the (inverse) Discrete Fourier Transform. The choice of weighting function $w(\omega)$ is of crucial importance. For instance, if the main source of degradation is the additive Gaussian noise, an SNR-weighted version of GCC can be used [145]. However, in practice, the highest degradation is due to reverberation, and the weighting that ameliorates its effect is so-called *GCC-PHAT* (from *PHAse Transform* [22, 80, 145]). The GCC-PHAT simply uses $w(\omega) = \frac{1}{|\phi_{12}(\omega)|}$, which deemphasizes contributions of individual frequency components. This choice of weighting makes the method more robust

with respect to reverberation, but at the expense of increased sensitivity to noise (since all frequencies equally contribute to estimation).

In the case where $k > 1$ sources are active, a straightforward way to estimate TDOAs of all of them is observing the k largest peaks of GCC-PHAT. However, as the number of sources increases, the estimation quality decreases, since one source can be interpreted as noise when estimating TDOA of another [22]. The performance is further worsened if the source signals are correlated. Still, a large-scale experimental evaluation [30] of several TDOA estimation methods has shown that variants of GCC-PHAT are competitive or outperform other common methods in the multisource, two-microphone setting.

When more than two microphones are available, they can be used to improve the TDOA estimate. This is done by (explicitly or implicitly) exploiting the mutual relations among delay variables (for instance, in the example problem in figure ... , $\tau_{13} = \tau_{12} + \tau_{23}$). Based on this observation, several approaches have been proposed [21, 137, 80, 120]. Most of these methods assume free space propagation model. Having this assumption, one can conclude that signal reaching any microphone in an array will be the sum of delayed and attenuated versions of source signals $z_i(t) := z(\mathbf{r}_i, t)$, possibly with an additive noise. Now, the time delay τ is defined as TDOA between a reference microphone and another microphone in the array (call it the *reference pair*). Then, one can define a function $f_j(\tau)$, based on the array geometry, that maps τ to the actual TDOA between the reference and the j^{th} microphone:

$$p_j(t) = \sum_{i=1}^{s} a_{i,j} z_i\left(t - \tau_{0,j} - f_j(\tau)\right) + E_j(t). \tag{4.13}$$

Here $a_{i,j}$ is the attenuation coefficient between the i^{th} source and j^{th} microphone, $\tau_{0,i}$ is the delay between the reference microphone and the i^{th} source, and $E_j(t)$ is the additive noise at j^{th} microphone.

This formulation allows for building a spatial correlation matrix, parametrized by τ, which lies at the heart of these methods. For instance, *Multichannel Cross-Correlation Coefficient (MCCC) algorithm* [21] is based on the spatial correlation matrix defined as

$$\mathbf{R}(\tau) = \mathbb{E}_t\left[\mathbf{p}\mathbf{p}^T\right] \quad \text{where } \mathbf{p}(t,\tau) = \left[p_1(t)\ p_2(t + f_2(\tau))\ p_3(t + f_3(\tau))\ \ldots\ p_m(t + f_2(\tau))\right]^T. \tag{4.14}$$

The goal of MCCC algorithm is to maximize the MCCC value, which can be seen as generalization of cross-correlation coefficient to multichannel case [61, 22]. It can be shown that the estimate that maximizes MCCC is found by solving:

$$\hat{\tau} = \arg\min_{\tau} \det(\mathbf{R}(\tau)), \tag{4.15}$$

where $\det(\mathbf{R})$ denotes the determinant of the positive semidefinite matrix $\mathbf{R}$. Whitening the signal, before cost computation, is analogous to PHAT weighting for the GCC method [61]. Likewise, the approach can be easily extended to multisource scenario.

Another group of direct methods that exploits spatial correlation matrices is the family of eigenvector-based techniques (often called *high resolution methods*). Among these, the most prominent members are the adaptations of *MUltiple SIgnal Classification (MUSIC)* [223] and *Estimation of Signal Parameters via Rotation Invariance Techniques (ESPRIT)* [214] algorithms. Unfortunately, they are very sensitive to reverberation effects and cannot be straightforwardly extended to multiple source setting [80, 22], thus they are not interesting for our indoor sound source localization problem. However, they can be useful in the brain source localization problem, and we will get back to them in chapter 6.

The common trait of all mentioned approaches is that they essentially assume the free field propagation model. Therefore, even though their performance may be improved by applying a convenient weighting method (*e.g.* PHAT), they are, by their nature, not well adapted to sound source localization in reverberant rooms. An algorithm termed *Adaptive Eigenvalue Decomposition (AED)* [20] was designed with this issue in mind. It attempts to blindly estimate channel impulse responses for the two microphone array, by exploiting the fact that the vector of impulse responses is in the null space of a certain correlation matrix. The extension to more than two microphones is possible [128], although it is not straightforward. The multichannel version is more robust since it alleviates the issue when channel filters share common zeros, and generally performs better than GCC/MCCC approaches in reverberant environments. Unfortunately, the algorithm cannot be easily applied to multisource setting, and because of that, it is not considered in this work. Some current research efforts are devoted to generalizing the approach to multisource scenario, *e.g.* an ICA-based method proposed in [44].

4.2.2 Beamforming methods

A *beamformer* is a multisensor array system that assigns temporal delays to signals, before combining them into output, in order to focus on some specific location in space. Beamforming methods are often used in combination with TDOA techniques, to improve SNR of the input signals (thus, they take the temporal delay estimate as an input). However, they can be also used independently, where they browse the predefined search space (*i.e.* the *energy map*) for the location(s) of maximal radiated energy.

In that sense, they are still closely related to TDOA methods, since they need to estimate temporal delays by maximizing the observed energy (or the *Steered Response Power (SRP)* [80]). These approaches can be seen as variants of *delay-and-sum* beamformer, which simply delays the input signals in order to compensate for propagation delays to each sensor:

$$b_{\mathbf{r}}(t) = \sum_{i=1}^{m} w_i p_i\left(t + f_i(\mathbf{r})\right). \tag{4.16}$$

Here the steering function $f_i(\mathbf{r})$ maps the location $\mathbf{r}$ to the delay of the i^{th} microphone, while the weights w_i are only used to shape the beam and are omitted in the remaining discussion.

Maximizing SRP corresponds to:

$$\underset{\mathbf{r}}{\text{maximize}}\ \mathbb{E}[b_{\mathbf{r}}^2] = \sum_{i=1}^{m}\sum_{j=1}^{m}\mathbb{E}[p_i(t+f_i(\mathbf{r}))\,p_j(t+f_j(\mathbf{r}))] = \sum_{i=1}^{m}\sum_{j=1}^{m} r_{i,j}(f_j(\mathbf{r}))-f_i(\mathbf{r}))]). \tag{4.17}$$

The correlations $r_{j,k}$ are computed in the frequency domain, which strongly resembles GCC approaches. Indeed, before evaluating correlations, one can apply frequency weighting to robustify estimation in adverse environments. Despite some initial attempts (*e.g.* [255]), the optimization (4.17) is usually not performed directly, due to the pronounced non-convexity of the cost function. Instead, the search space is discretized and the correlations $r_{i,j}$ are computed for all *discrete* locations $\mathbf{r}$. In the second stage, the cost function is evaluated for every $\mathbf{r}$, yielding the estimated location. The approach is easily extendable to multisource case, by declaring k highest-energy locations to be the source location estimates.

The most successful beamforming method uses PHAT frequency weighting, and is termed SRP-PHAT or *Global Coherence Field (GCF)* [195, 80]. In the case of $m = 2$ microphones, SRP-PHAT is equivalent to GCC-PHAT method. For an array with more than two sensors, SRP-PHAT accounts for all pairwise cross-correlations, but in a different manner than MCCC algorithm. Empirical studies [82] have shown that the two approaches are competitive, but with SRP-PHAT being more robust to highly reverberant environments and microphone calibration errors. Lastly, computational complexity of SRP-PHAT is lower than that of MCCC, due to the necessity to evaluate matrix determinants (4.15) in the latter approach [197]. Several improvements of the SRP-PHAT algorithm, in terms of reducing its computational complexity, have been suggested in [84, 83, 66].

4.3 Sound source localization by wavefield extrapolation

We have seen that indoor localization is challenging, particularly when significant reverberation and noise is present. Thus, different approaches have been proposed to somehow mitigate these issues, but their performance is usually a trade-off between accuracy, sensitivity to noise and sensitivity to reverberation. A preferable algorithm should be able to accurately localize multiple sound sources, maintain robustness to reverberation (*i.e.* the acoustic multipath) and benefit from the multichannel recordings.

The traditional "reverberation-aware" approaches aim at reducing reverberation effects: PHAT-based by suppressing the "reverberation noise", and AED by improving the estimation of a direct path component of the signal. Thus, reverberation is traditionally considered as unwanted phenomenon. On the other hand, knowledge of the propagation channel is successfully exploited in wireless communications. The famous *rake* receiver [36] exploits multipath propagation to increase SNR by applying *matched filter* to several intentionally delayed copies of the received signal. Matched filter simply correlates an input signal with an estimate of the channel impulse response, which is analogous to time-reversal [103] filtering. By delaying the input signal, rake targets individual multipath components in the mixture, which are later

constructively combined in order to improve the estimate of the original signal. Recent work, presented in [85], is one example of exploiting the rake concept for acoustic multipath.

Recognizing that reverberation is not necessarily an enemy, new localization methods based on somewhat extravagant idea arose: determine sound source locations by estimating the acoustic pressure field they produced. Since the field is known at only few measurement locations, the goal is to extrapolate the sound at the remaining space, or, equivalently, to recover the source term inducing this field. This becomes the acoustic inverse source problem. For the purpose of plain sound source localization, recovering the entire field may seem overly demanding. Yet, there are several advantages to this approach, which we preview, for now:

1. Signal model incorporates realistic acoustic channels between sources and microphones, and is thus robust to reverberation.
2. The model is not directly dependent on the strength of the direct component in the impulse response, which enables interesting applications (more on this in chapter 5).
3. Provided that the estimation is successful, localization should be highly accurate.
4. The "byproduct" is the acoustic field estimate *for all* discrete spatio-temporal coordinates, which is a distinct feature in its own right.

4.3.1 Prior art

Acoustic wavefield estimation has been exhaustively investigated in the field of acoustic tomography. For example, in seismic signal processing it is known as *wavefield inversion* [167, 79]. The goal there is to infer the underlying anisotropic structure of an object, by estimation of the attenuation, sound speed or scattering properties in different subsurface layers. The idea is to radiate a sound at some position on the surface and then measure the echo signal induced by this action. This problem is complementary to the inverse source problem, which is the topic of our interest, where we assume that the domain is given beforehand, but the sound sources are unknown.

Sound source localization through wavefield extrapolation and low-complexity regularization was first introduced by Malioutov et al. in [172]. They assumed a free-field propagation model, which allowed them to explicitly compute the associated Green's functions. The narrowband sound sources were estimated by applying sparse synthesis or low-rank regularizations. A wideband extension was proposed in [171], which is, however, a two-stage approach that implicitly depends on solving the narrowband problem.

"Indoor" localization[9] was first performed by Dokmanić and Vetterli [87, 86] in frequency domain. They used the Green's functions dictionary numerically computed by solving the Helmholtz equation with Neumann boundary conditions, by the *Finite Element Method*

[9] By "indoor" we mean a non-free space setting, even though all approaches consider 2D localization.

(FEM). The wideband scenario was tackled as jointly sparse problem, to which, in order to reduce computational cost, a modification of the OMP algorithm was applied. However, as argued in [58], this approach is critically dependent on the choice of frequencies, and can fail if modal frequencies are used. Le Roux et al. [156] proposed the CoSaMP algorithm for solving the sparse synthesis problem in the same spirit. In his doctorate thesis [86], Dokmanić remarks that convex relaxation performs better than the proposed greedy approach, but the computational complexity of the sparse synthesis regularization prohibits its use.

Building upon their result [57, 55] on approximating solutions of Helmholtz equation by plane waves, Chardon et al. proposed a narrowband sparse synthesis method to localize sound sources without explicit knowledge of boundary conditions [56]. Sound sources need to be within a space enclosed by convex hull of sensors, and the method requires more measurements compared to the case where the boundary is known beforehand.

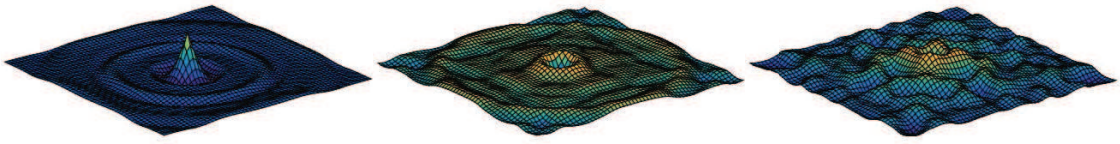
All these methods are based on the sparse synthesis prior. The first approach that exploited cosparsity was proposed by Nam et al. in [188], where the analysis operator was derived by discretizing the acoustic wave equation in time domain. Then, a greedy algorithm was used to estimate source positions corresponding to jointly sparse vectors. The authors speculated possible numerical advantages of the analysis approach compared to synthesis, due to the inherited sparsity of the analysis operator.

4.3.2 Contributions

In the conference paper [140] we propose convex relaxation for solving the acoustic inverse source problem, for the purpose of localizing sources, using the sparse analysis prior. It was shown empirically that convex relaxation outperforms the greedy approach proposed in [188] (later on, although without explicit reference to cosparsity, Antonello et al. revisited convexity in their own work [10]). In the framework paper [139], significant space is devoted to computational unevenness of the sparse synthesis and sparse analysis regularization of the acoustic wave equation in practice. Indeed, using the sparse analysis approach we managed to simulate the problem in three spatial dimensions (whilst all other approaches are limited to 2D), in the physically realistic setting. Furthermore, the robustness of the physics-driven regularization for sound source localization was discussed. The remainder of the present chapter elaborates and extends this work by including comprehensive numerical experiments for various problem parameterizations.

4.3.3 Physics-driven (co)sparse acoustic source localization

To fit the problem in the physics-driven framework, we use the regularization procedure presented in chapter 3. Recall that the necessary ingredients to apply the physics-driven approach are the physical model, the existence of sparse singularities and the measurement system that measures the diffuse physical quantity implicitly related to sparse singularities.



$\mathbf{p}_{t_1}$ $\mathbf{p}_{t_2}$ $\mathbf{p}_{t_3}$

Figure 4.4 – Example of discretized 2D pressure field: $\mathbf{p} = \left[\dots \mathbf{p}_{t_1}^T \dots \mathbf{p}_{t_2}^T \dots \mathbf{p}_{t_3}^T \dots \right]^T$.

The physical model

Modeling the sound wave propagation in air has been discussed in subsection 4.1.1. It is given in the form of the inhomogeneous acoustic wave equation (4.8), along with the homogeneous initial and boundary conditions:

$$\begin{aligned} \triangle p - \frac{1}{c^2}\frac{\partial^2 p}{\partial \mathrm{t}^2} &= z, \text{ for } (\mathbf{r}, \mathrm{t}) \in \{\Gamma \setminus \partial\Gamma\} \times (0, \tau) \\ p = 0, \quad \frac{\partial p}{\partial \mathrm{t}} &= 0, \text{ for } \forall \mathbf{r},\ \mathrm{t} = 0 \\ B(p) &= 0, \text{ for } \mathbf{r} \in \partial\Gamma,\ \forall \mathrm{t}. \end{aligned} \tag{4.18}$$

Here $B(p)$ denotes the applied boundary conditions, modeled by Robin or Mur's absorbing condition described in subsection 4.1.2, for the sake of generality (we can recover Dirichlet, Neumann or mixed conditions by setting appropriate values to α and ξ). The functional $z := z(\mathbf{r}, \mathrm{t})$ is defined as follows:

$$z(\mathbf{r}, \mathrm{t}) = \sum_{\mathrm{j}=1}^{\mathrm{k}} z_{\mathrm{j}}(\mathrm{t})\delta(\mathbf{r} - \mathbf{r}_{\mathrm{j}}) = \sum_{\mathrm{j}=1}^{\mathrm{k}} -\frac{\partial^2 m_{\mathrm{j}}(\mathrm{t})}{\partial \mathrm{t}^2}\delta(\mathbf{r} - \mathbf{r}_{\mathrm{j}}), \tag{4.19}$$

where $m_{\mathrm{j}}(t)$ is the mass injection function described in (4.8). We assume that the temporal mean of $z(\mathbf{r}, \mathrm{t})$ is zero, *i.e.*:

$$\int_0^T z(\mathbf{r}, \mathrm{t})\, d\mathrm{t} = 0, \tag{4.20}$$

in order to be able to correctly model so-called *soft sources* [226] (explained in appendix B.1).

This way we formulated a well-posed problem, which is, as now usual, compactly represented by the forward operator $A(p) = z$. In fact, some strong assumptions are made here: in the room acoustic context, we assume that the geometry, the boundary (*e.g.* wall) structure, and the propagation speed are known beforehand. We will see in chapter 5 that some of these assumptions can be relaxed.

Moving into discrete setting, we now apply the Finite Difference Time Domain (FDTD) Standard Leap Frog (SLF) method [159], which corresponds to second-order centered finite differ-

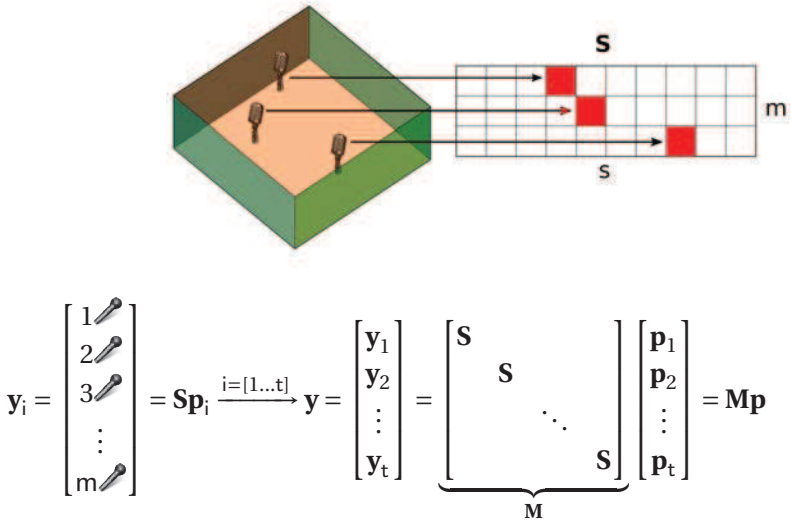


Figure 4.5 – Structures of the matrices **S** (top) and **M** (bottom).

ences in space and time (described in appendix B.1). Let s be the number of points used for discretizing space and t the number of temporal discretization points (time samples). This yields a non-singular system of difference equations of the form

$$\mathbf{A}\mathbf{p} = \mathbf{z}, \tag{4.21}$$

with the square invertible coefficient matrix $\mathbf{A} \in \mathbb{R}^{st \times st}$. Analogously, the discretized spatio-temporal pressure field $\mathbf{p} \in \mathbb{R}^{st}$ and the discretized spatio-temporal source component $\mathbf{z} \in \mathbb{R}^{st}$ are built by vectorization and sequential concatenation of t corresponding s-dimensional vector fields (as illustrated in figure 4.4 for the vector **p**). The matrix operator **A** is a banded lower triangular matrix. Moreover, the matrix **A** is very sparse, as it can have only a very limited number of non-zeros per row (*e.g.* maximum seven in the 2D case).

The Green's functions associated with the forward model (4.18) are obtained by setting $z = \delta(\mathbf{r})\delta(t)$. In the discrete case, this corresponds to solving the linear system $\mathbf{A}\mathbf{g} = \mathbf{i}$, where **i** is a column of the identity matrix $\mathbf{I} \in \mathbb{R}^{st \times st}$. In other words, we can build the Green's functions dictionary by computing the inverse $\mathbf{D} = \mathbf{A}^{-1}$. It is important to note that one cannot derive an analytical expression of the Green's functions in the general case of arbitrary combination of initial and boundary conditions[10]. Thus, in many cases, the solutions are indeed approximated by a numerical method.

The measurement system

The observation data is collected by measuring the sound pressure, which is nothing else but recording sound by m microphones. Assuming that the analogue signal was processed by a

[10] Not to be confused with the well-posedness of the forward PDE problem (4.18) itself.

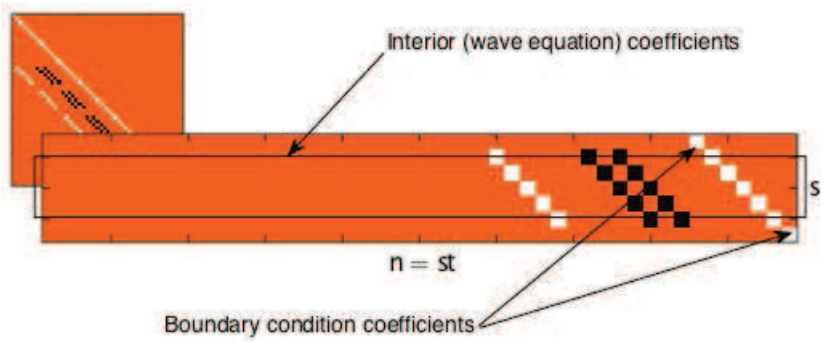


Figure 4.6 – Structure of the matrix **A** for a demonstrative 1D problem with homogeneous Dirichlet boundary conditions: the zoomed region represents the last s rows.

built-in low-pass filter, the measurement system simplifies to a spatio-temporal subsampling matrix $\mathbf{M} \in \mathbb{R}^{mt \times st}$, as illustrated in figure 4.5. The matrix **M** is block-diagonal, with each block being a spatial subsampling (restriction) operator, *i.e.* a row-reduced identity matrix $\mathbf{S} \in \mathbb{R}^{m \times s}$.

Spatial sparsity

Finally, the last requirement is fulfilled by assuming that the sound sources are spatially sparse. In the discrete setting, this means that the number of point sources k is much smaller than the "size" s of a discretized spatial domain. Now, the question is whether, given the measurements, the sparsity assumption alone is sufficient to perfectly recover the underlying signal.

The analysis and synthesis problems are equivalent in our framework, therefore it is sufficient to discuss only one approach. Since the structure of the matrices **A** and **M** in FDTD-SLF discretization is predictable, it is easier to discuss the sparse analysis setting. Concerning the matrix **A** given in figure 4.6, a spatial reconfiguration would only induce the change in position of the rows corresponding to boundary conditions (assuming the boundary type is fixed). Regarding the measurement matrix **M**, microphone positions are directly mapped to the rows of the identity matrix, as presented in figure 4.5. The matrix **M** is constrained by the fact that we cannot sample the boundary nor the source positions. Further, if there are k fixed sound sources in space, the easiest way to formulate spatio-temporal source support is by considering the matrix $\mathbf{Z} \in \mathbb{R}^{s \times t}$, whose k rows can contain non-zero elements. Then, the support is defined as the indices of non-zero elements of sequentially concatenated columns of the matrix **Z**. Obviously, the support of each source will appear periodically, with period s. Again, there is a natural constraint on the support set: sources cannot be placed on the boundary, nor in the microphone positions.

Recalling the necessary uniqueness condition for the sparse analysis recovery (1.15), in chapter 1, we know that the concatenated matrix $\begin{bmatrix} \mathbf{A}_\Lambda \\ \mathbf{M} \end{bmatrix}$ needs to be of full column rank. Let us check if this is indeed the fact. Since $\mathbf{A} \in \mathbb{R}^{n \times n}$ is square-invertible, any row-reduced submatrix $\mathbf{A}_\Lambda \in \mathbb{R}^{(n-kt) \times n}$ will not be a full column rank matrix, by the fundamental theorem of linear

algebra. In order to stabilize the linear system, rows of the matrix $\mathbf{M}$ need to cover the null-space of $\mathbf{A}_\Lambda$, *i.e.* to complement the basis vectors consisting of rows of $\mathbf{A}_\Lambda$. We focus on the *last* s rows of the matrix $\mathbf{A}$, illustrated as the zoomed region in figure 4.6. Some of these rows correspond to source positions - let Υ denote the index set of these rows within $\mathbf{A}$ (naturally, $|\Upsilon| = k$). Since these do not belong to the cosupport, they will not appear in the matrix $\mathbf{A}_\Lambda$. Now, consider the matrix $\mathbf{N}_\tau$ formed by extracting rows indexed by Υ from the identity matrix $\mathbf{I} \in \mathbb{R}^{n\times n}$. By inspecting the structure of the analysis operator in figure 4.6, it becomes clear that each row of $\mathbf{N}_\tau$ is orthogonal to any other[11] row of $\mathbf{A}$, and therefore, to any row of $\mathbf{A}_\Lambda$ (*i.e.* $\mathbf{N}_\tau^T \subset \mathtt{null}(\mathbf{A}_\Lambda)$). Since $\mathbf{N}_\tau$ is a row-reduced identity matrix, it is easy to compare it with the measurement matrix $\mathbf{M}$. Unfortunately, it becomes obvious that $\mathbf{N}_\tau^T \subset \mathtt{null}(\mathbf{M})$, since these matrices are formed by complementary rows of $\mathbf{I}$, due to the problem constraints (we are not allowed to place microphones at source positions). This means that we cannot hope to recover the source signal at $t = \tau$ (in discrete domain, this corresponds to time instant t), unless additional information is assumed. It is expected, and only a consequence of the finite propagation speed prescribed by the wave equation: the information sent by sources cannot reach microphones instantaneously. We term this phenomenon the *acoustic event horizon*.

Therefore, in addition to the initial and boundary conditions, we need to impose the *terminal condition*, *i.e.* we need to characterize the solution at $t = \tau$ beforehand. We do this by assuming that the terminal conditions are also homogeneous, *i.e.* the acoustic sources are turned off before the end of data acquisition (another possibility is to completely disregard an estimate at $t = \tau$). While the terminal condition is a necessary requirement for the accurate wavefield recovery, for the source localization, as we will see later, this assumption can be dropped.

Unfortunately, the inverse source problem is proven to be ill-posed, in general [78, 31]. That is, the boundary, initial and terminal conditions, along with measurements, are not sufficient to achieve perfect signal recovery. It has been proven that certain, so-called *non-radiating sources* can produce a field which is supported only within the source region, *i.e.* they radiate no energy outside the spacetime occupied by themselves. This also holds for the inverse scattering problems, where these pathological singularities are known as *non-scattering potentials*. However, even though non-radiating sources can be easily "created" by any function $z_{\mathrm{nr}} \in H^2(\Omega)$ supported in a compact region, but otherwise arbitrary [78], there has been no evidence to date of their physical existence [79]. Moreover, we never observed this behavior during our empirical investigation. Nevertheless, this suggests that inferring positive theoretical results, based on the compressed sensing theory, is not straightforward, and that one needs to assume a more restricted class of signals than only spatially sparse.

Localization

Given $\hat{\mathbf{p}}$, an estimate of the pressure field, or equivalently, $\hat{\mathbf{z}}$, an estimate of the source term, the localization task becomes straightforward. After identifying $\{\hat{\mathbf{z}}_1, \hat{\mathbf{z}}_2 \ldots \hat{\mathbf{z}}_j \ldots \hat{\mathbf{z}}_s\}$, the t-long

[11] "Any other" = not corresponding to the same row index.

subvectors of $\hat{\mathbf{z}}$ corresponding to each discrete spatial location in Γ, source locations are retrieved by setting a threshold on the minimum energy of $\hat{\mathbf{z}}_j$. Consequently, any spatial location $j \in [1, s]$ with $\|\hat{\mathbf{z}}_j\|_2$ higher than the threshold, is declared a sound source location. If the number of sound sources k is known beforehand, estimating the threshold can be avoided. Instead, one would consider k highest in magnitude spatial locations j to be sound source positions.

4.3.4 Convex regularization

To estimate $\hat{\mathbf{p}}$ and $\hat{\mathbf{z}}$, we need to solve *one* of the following two regularized physics-driven inverse problems:

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} f_r(\mathbf{A}\mathbf{p}) + f_d(\mathbf{M}\mathbf{p} - \mathbf{y}) + f_{c_1}(\mathbf{A}_0\mathbf{p}) + f_{c_2}(\mathbf{A}_\tau\mathbf{p}) + f_{c_3}(\mathbf{A}_{\partial\Gamma}\mathbf{p}) \tag{4.22}$$

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z}} f_r(\mathbf{z}) + f_d(\mathbf{M}\mathbf{D}\mathbf{z} - \mathbf{y}) + f_{c_1}(\mathbf{z}_0) + f_{c_2}(\mathbf{z}_\tau) + f_{c_3}(\mathbf{z}_{\partial\Gamma}), \tag{4.23}$$

where the matrices $\mathbf{A}_0$, $\mathbf{A}_\tau$ and $\mathbf{A}_{\partial\Gamma}$ are formed by extracting rows of $\mathbf{A}$ corresponding to initial, terminal and boundary conditions, respectively. Analogously, the vectors $\mathbf{z}_0$, $\mathbf{z}_\tau$, $\mathbf{z}_{\partial\Gamma}$ are formed by extracting corresponding elements of $\mathbf{z}$.

The choice of convex penalties f_r, f_d, f_{c_1}, f_{c_2} and f_{c_3} should reflect the particular problem setting. To simplify matters, we will only use one type of functional for all penalties f_d, f_{c_i} - the indicator function of the ℓ_2 norm of a vector:

$$\chi_{\ell_2 \le \varepsilon}(\mathbf{v}) := \begin{cases} 0 & \|\mathbf{v}\|_2 \le \varepsilon, \\ +\infty & \text{otherwise.} \end{cases} \tag{4.24}$$

This allows us to use the linear equality constraint, by setting ε to a very low value. The choice of functional is generally suboptimal, as it promotes isotropic distribution of the residual vector entries. Hence, the assumption is not entirely correct, *e.g.* due to correlations between the true signal and the finite difference approximation error embedded into the residual $\mathbf{M}\mathbf{p} - \mathbf{y}$. The associated $\text{prox}_{\ell_2 \le \varepsilon}(\cdot)$ operator admits a closed form, given in appendix A.3.

The ℓ_1 norm Concerning the penalty f_r, the most common convex relaxation of the non-convex ℓ_0 objective is the ℓ_1-norm,

$$\|\mathbf{z}\|_1 = \sum_i |z_i| \tag{4.25}$$

which is known to promote sparse solutions, as discussed in chapter 1.

Group $\ell_{2,1}$ norms In addition to the ℓ_1 norm, we will also consider the *group* $\ell_{2,1}$ norm, which is defined as the ℓ_1 norm of a vector $[z_1\ z_2\ \ldots z_g]^T$, where z_i denotes the ℓ_2 norm of i^{th}

group of elements of **z**. Therefore, this type of norm is more general, as it includes the standard ℓ_1 norm as a special case (the "groups" are just singletons). It encodes the known structure in the signal estimate, and should therefore provide more accurate results. The proximal operators associated with some types of these norms admit closed-form expressions, such as the ones we describe below (their proximal operators are given in appendix A.3).

The joint $\ell_{2,1}$ norm If local spatial stationarity of the sources is assumed (say, for sufficiently short acquisition time), the sources retain fixed positions in space. Then we favor solutions for which all temporal slices of the sparse estimate have the same support (also known as *jointly sparse* vectors). This is promoted via the *joint*[12] $\ell_{2,1}$-norm:

$$\|\mathbf{z}\|_{2,1} = \sum_{\mathrm{i}} \sqrt{\sum_{\mathrm{j}} |z_{\mathrm{i,j}}|^2}, \tag{4.26}$$

where $z_{\mathrm{i,j}}$ denotes the $(\mathrm{i},\mathrm{j})^{\mathrm{th}}$ element obtained by transforming the vector **z** into a matrix **Z** whose columns are jointly sparse subvectors.

Hierarchical $\ell_{2,1}$ norms An attractive class of (overlapping) group $\ell_{2,1}$ norms are *hierarchical* $\ell_{2,1}$ norms [132, 14]. The easiest way to visualize this hierarchy is as a tree-like structure. Whenever a "parent" (the group which is higher in hierarchy) is not selected, a "child" (the group lower in hierarchy) is also not selected. For our purposes, an interesting case is a special type of hierarchal $\ell_{2,1}$-norms where groups are either singletons or disjoint subsets of elements [132]. This objective function should encourage solutions with a small number of active disjoint groups and which are overall sparse. In our case, where the disjoint groups are defined as temporally jointly sparse vectors (as in (4.26)), it is evaluated as a sum $\|\mathbf{z}\|_{2,1} + \|\mathbf{z}\|_1$. This norm may perform well in realistic settings, for instance, if a source emits speech signal, which usually contains silent intervals. When placed into the objective function it should promote solutions which are spatially and temporally sparse.

4.3.5 Computational complexity of the cosparse *vs* sparse regularization

In section 3.5 we argued that computational complexity of the analysis regularization should be considerably lower than in the synthesis case, provided that discretization method is locally supported. This is especially pronounced in models of time-dependent phenomena, such as the wave equation. For the FDTD-SLF discretization, given the sparse and banded structure of the matrix **A**, our intuition tells us that the corresponding inverse is rarely a sparse matrix (generally, for $\mathbf{A}^{-1}$ to be also banded, occurs only if both matrices can be factorized into a product of block diagonal matrices [230]). Indeed, this is true: even though it is also a lower triangular matrix, the dictionary **D** cannot be sparse - it becomes obvious if one rewrites the

[12] *Joint* indicates that individual groups ("columns") would have identical support, but the groups themselves are actually disjoint (there are no common elements).

Problem size s × t	(19 × 19) × 61	(30 × 30) × 97	(48 × 48) × 155	(76 × 76) × 246	(95 × 95) × 307
Synthesis (GB)	0.1	0.6	4.1	26	63
Analysis (GB)	0.001	0.005	0.02	0.07	0.2

Table 4.1 – Memory requirements relative to the problem size, with m = 10 microphones.

discretization provided in appendix B.1 in the causal (explicit) form. Then, the columns of the dictionary $\mathbf{D}$ are simply the truncated impulse responses of an infinite impulse response filter.

The cost of matrix-vector products in the analysis case is of order $O(n) = O(st)$, as opposed to $O(mst^2)$ in the synthesis case. Further, factorization of the coefficient matrices $\mathbf{H}_A$ and $\mathbf{H}_S$, used in the linear least squares update (3.28) can be compared. Since the bandwidth[13] of the matrix $\mathbf{A}$ is of $O(s)$, when Cholesky factorization is used for solving the least squares step of Weighted SDMM, we have $\mathbf{H}_A = \mathbf{L}_A\mathbf{L}_A^T$ with $\mathtt{nnz}(\mathbf{L}_A) = O(s^2t)$. In the synthesis case, the coefficient matrix is not banded, thus $\mathbf{H}_S = \mathbf{L}_S\mathbf{L}_S^T$, with $\mathtt{nnz}(\mathbf{L}_S) = O(s^2t^2)$. Sparsity of the sparse Cholesky factors cannot be predicted beforehand, as discussed before.

The computational complexity of the synthesis method is unreasonably high for regularizing the acoustics physics-driven inverse problem, which is experimentally verified in the following Section. Moreover, the memory requirements presented in table 4.1, for the example setting with m = 10 microphones, indicate that the storage cost for the synthesis regularization increases fast with the problem dimension, and can, therefore, become extremely high when realistic physical domains are considered.

4.4 Simulations

The experiments are divided into six groups, in order to investigate different aspects of the acoustics physics-driven problem:

1. First, we test the fundamental hypothesis that *sparse analysis and sparse synthesis indeed achieve identical solutions*.
2. Secondly, we provide *comparison* with the greedy approach proposed in [188].
3. In the third group, we *drop the terminal condition* and verify that the localization performance is not affected.
4. Fourth group is dedicated to investigation of *scaling capabilities* of the two synthesis and analysis approaches, by varying the problem size and number of microphones.
5. The fifth subsection is aimed at comprehensive *performance evaluation* under various forward model parameterizations.

[13] The bandwidth of the matrix $\mathbf{A}$ is the number b such that $\mathbf{A}_{i,j} = 0$ for all $|i - j| > b$.

6. All previous experiments are purely numerical, and we postpone experiments with a more physical interpretation to final, sixth group of experiments, where we test the *robustness* of the localization approach with respect to modeling errors.

Algorithm parameterization. Unless otherwise specified, the constraint parameter of (4.24) is set to $\varepsilon = 0$, to enforce equality constraints. For the stopping criteria, we use (A.12), with the relative accuracy $\mu = 10^{-2}$ in (A.13), or the maximum number of SDMM iterations (5000). The SDMM multiplier ρ_1 is set to $\rho_1 = \rho = 10$ and the remaining ones are computed using the heuristics explained in Subsection 3.4.1. In the experiments where the LSMR algorithm is used to estimate the solution, we set its stopping criterion to $\|\mathbf{H}^\mathsf{T}(\mathbf{H}\mathbf{v}-\mathbf{h})\|_2/\|\mathbf{h}\|_2 \leq 10^{-4}\mu$ (given a least squares problem $\mathbf{v}^* = \arg\min_{\mathbf{v}} \|\mathbf{H}\mathbf{v}-\mathbf{h}\|_2^2$).

Data simulation and processing. The sampling model simulates recordings taken by fixed, but randomly placed microphones. The point sources are also randomly distributed in space, and their number is always lower than the number of microphones. This, along with randomization, is to ensure that the occurrence of localization ambiguities[14] is highly unlikely.

Assuming that the number of sound sources k is given, we determine the set of the sound source location estimates from $\mathbf{z} = \mathbf{A}\mathbf{x}$ (the analysis case) or directly from $\mathbf{z}$ (the synthesis case). Namely, we index the elements of $\mathbf{z}$ by $(\mathbf{r}, \mathrm{t})$, and then we simply declare the locations corresponding to the k highest values of $\|\mathbf{z}_{\mathbf{r},:}\|_2$ to be the estimated positions. Note that the knowledge of the number of sound sources in assumed for simplicity, otherwise a magnitude threshold could be estimated by standard precision/recall method.

Performance measures. The quality of localization is presented as an estimated error per source. It is computed as the *Root Mean Square Error (RMSE)* between pairs $(\hat{\mathbf{r}}_\mathrm{i}, \tilde{\mathbf{r}}_\mathrm{i})$, where $\hat{\mathbf{r}}_\mathrm{i}$ and $\tilde{\mathbf{r}}_\mathrm{i}$ denote the estimated and the true position of the i^{th} source, respectively. The pairs are chosen such that the overall error is minimal, by means of the Hungarian algorithm [151]. The quality of signal recovery is presented in terms of SNRs with respect to the estimation error:

$$\mathrm{SNR}_{\hat{\mathbf{p}}} = 20\log_{10}\frac{\|\mathbf{p}\|_2}{\|\mathbf{p}-\hat{\mathbf{p}}\|_2} \qquad \mathrm{SNR}_{\hat{\mathbf{z}}} = 20\log_{10}\frac{\|\mathbf{z}\|_2}{\|\mathbf{z}-\hat{\mathbf{z}}\|_2},$$

where $\hat{\mathbf{p}}$ and $\hat{\mathbf{z}}$ are the estimated pressure and source signal, respectively.

The simulation results are averaged over 50 realizations. Note that we resampled the sources and their location between these 50 experiments. The experiments were run on Intel® Xeon® 2.4GHz cores, equipped with 8GB RAM, in single-core/single-thread mode.

[14] For example: a microphone array placed on an axis of symmetry of a room.

4.4.1 Equivalence of the analysis and synthesis regularization

A first set of source localization experiments provides empirical evidence of the equivalence of the analysis and synthesis models for the physics-driven inverse problems. In particular, we compare their respective performance as a function of the number of microphones and white noise sources, while using the three chosen convex functionals f_r (the ℓ_1, joint $\ell_{2,1}$ and hierarchical $\ell_{2,1}$ norms).

To avoid a potential bias in the results if iterative methods are used to solve the least squares step of SDMM, we restrict the experiments to a small scale problem for which direct methods are still applicable. The domain is an artificial two dimensional spatial grid of size 15×15, simulated through 50 time samples. Boundary conditions are modeled by Robin condition, tuned by setting $\alpha = 100$ to approximate Neumann boundary condition. The results (figures 4.7, 4.8, 4.9 and 4.10) are given in the form of phase transition graphs, in terms of $\mathrm{SNR}_{\hat{\mathbf{p}}}$, $\mathrm{SNR}_{\hat{\mathbf{z}}}$ and the frequency (empirical probability) of *perfect* localization (error per source is zero).

The simulations concern cases when the emission time $\mathrm{t_e}$ of the sources is comparatively long ($\mathrm{t_e} = 45$ time samples), short ($\mathrm{t_e} = 20$) and very short ($\mathrm{t_e} = 5$). In all these settings, when the same objective f_r is used, figures 4.7, 4.8 and 4.9 verify point-to-point that both **the analysis and synthesis approach provide numerically identical results**. Regarding the emission duration, its effect is mostly pronounced in the last scenario ($\mathrm{t_e} \ll \tau$, figure 4.9), when the performance is adversely affected. Finally, we remark that **the structured norms outperform classical ℓ_1 minimization**, as predicted.

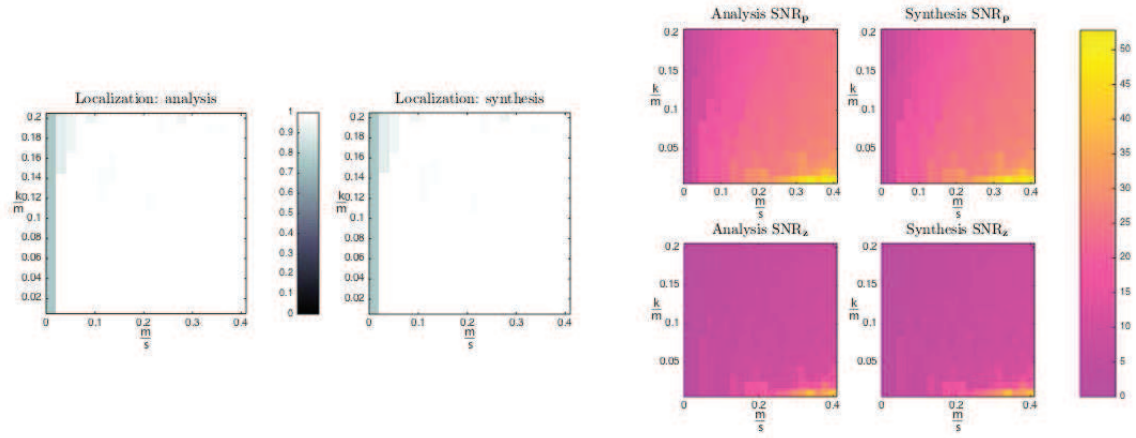
Therefore, we will use only the structured norms in the remaining experiments.

4.4.2 Comparison with GRASP

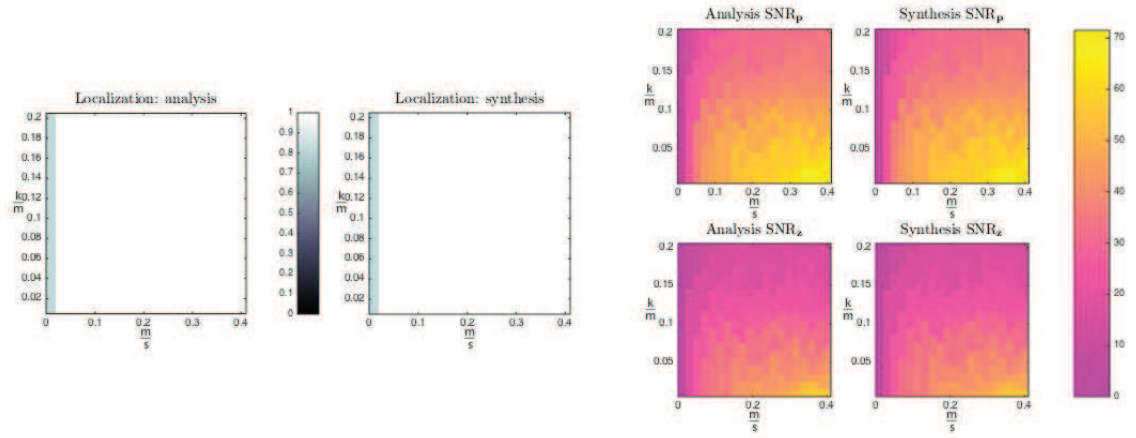
Now the *GReedy Analysis Structured Pursuit (GRASP)* algorithm [188][15] is tested on the same localization problems as in the previous subsection. Its performance for different emission durations is presented in figures 4.10a, 4.10b and 4.10c.

It is evident that **convex approaches outperform the GRASP algorithm** in terms of source localization, except perhaps the ℓ_1 minimization for the very short emission setting. The results in terms of SNR values are deceptive, as GRASP performs least-squares fitting to the estimated cosupport in each iteration. This means that, provided it has detected the correct spatial support (*i.e.* the localization is successful), its estimate will have high SNR. But the same technique can be applied to any convex approach, after completing the source localization step. Concerning the emission duration, performance of GRASP is stable with respect to changes of $\mathrm{t_e}$, due to its structure-aware cosupport estimation.

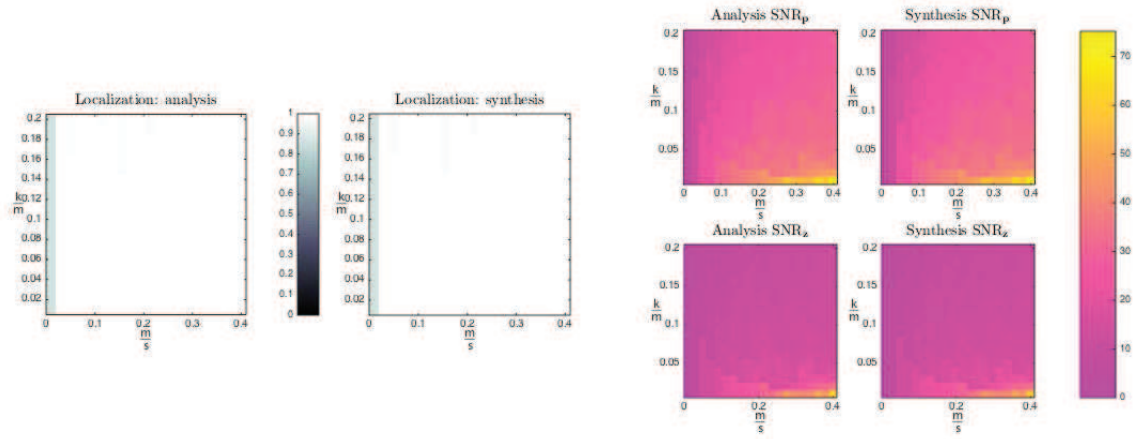
[15] The pseudocode of GRASP is provided in appendix D.



(a) The ℓ_1 norm: localization probability (left) and estimation SNR (right).

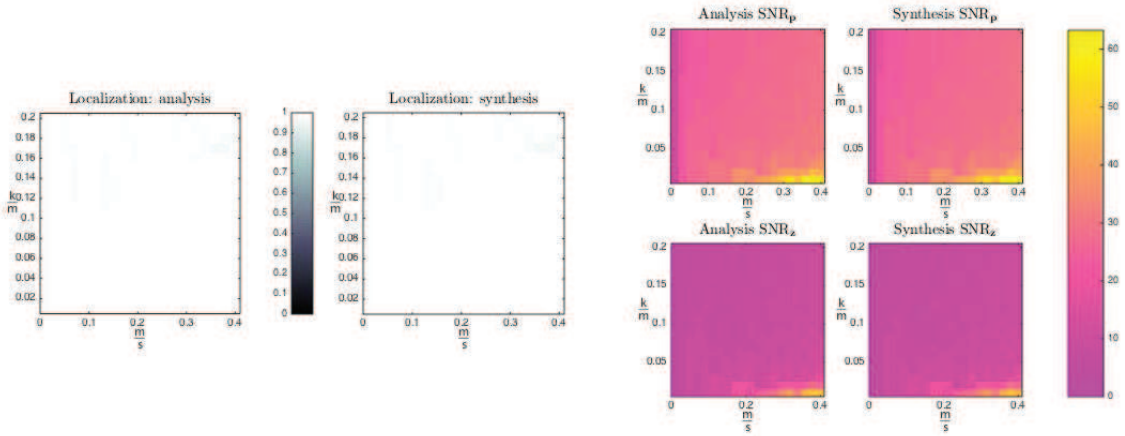


(b) The joint $\ell_{2,1}$ norm: localization probability (left) and estimation SNR (right).

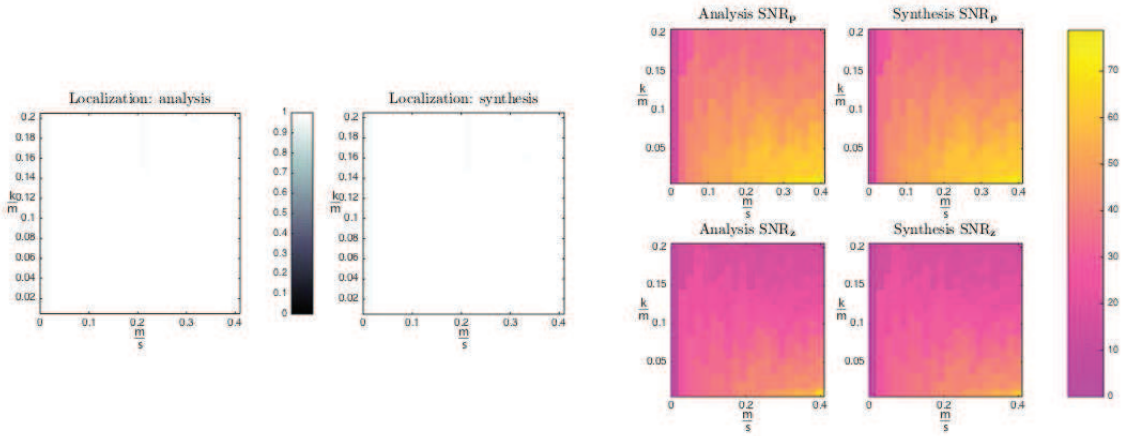


(c) The hierarchical joint $\ell_{2,1}$ norm: localization probability (left) and estimation SNR (right).

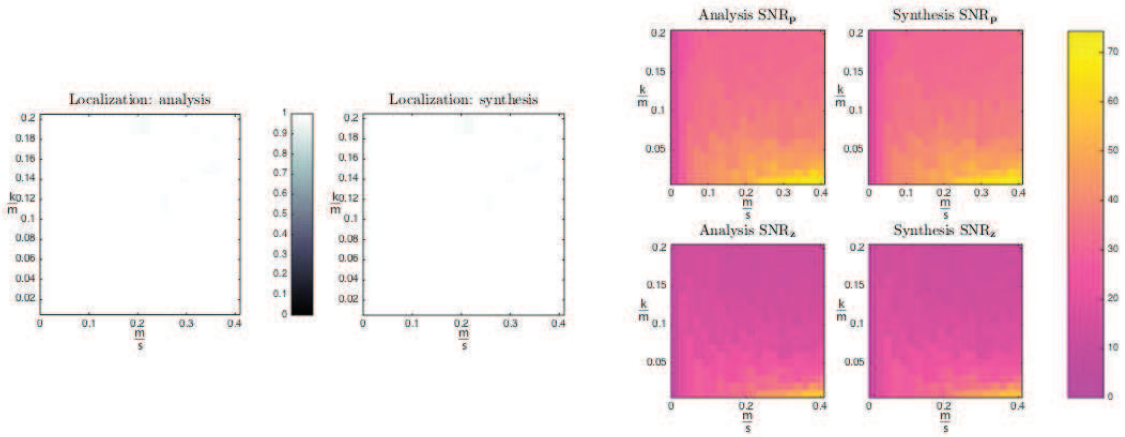
Figure 4.7 – Long source emission time ($t_e = 45$).



(a) The ℓ_1 norm: localization probability (left) and estimation SNR (right).

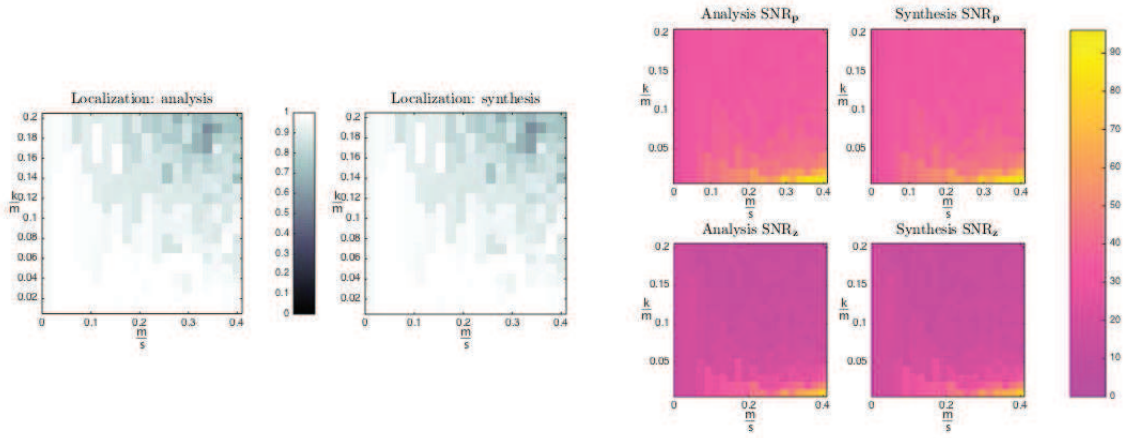


(b) The joint $\ell_{2,1}$ norm: localization probability (left) and estimation SNR (right).

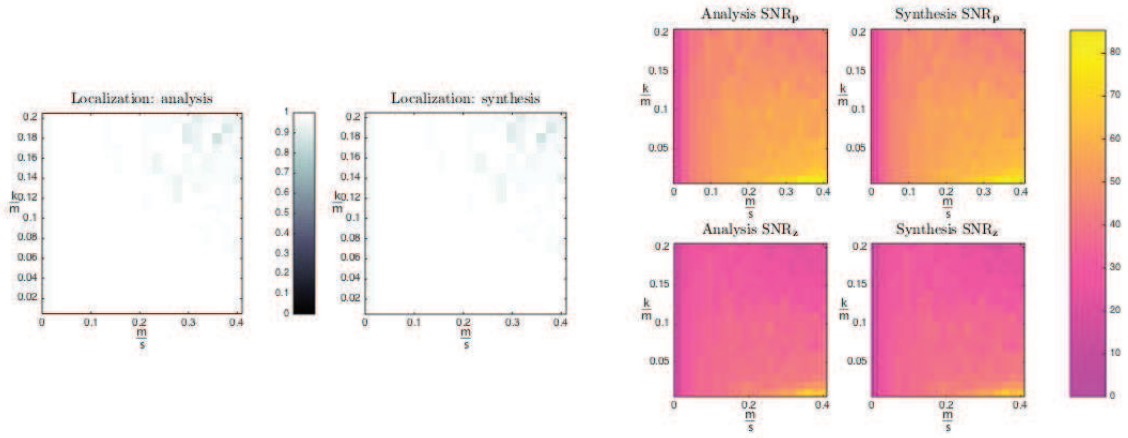


(c) The hierarchical joint $\ell_{2,1}$ norm: localization probability (left) and estimation SNR (right).

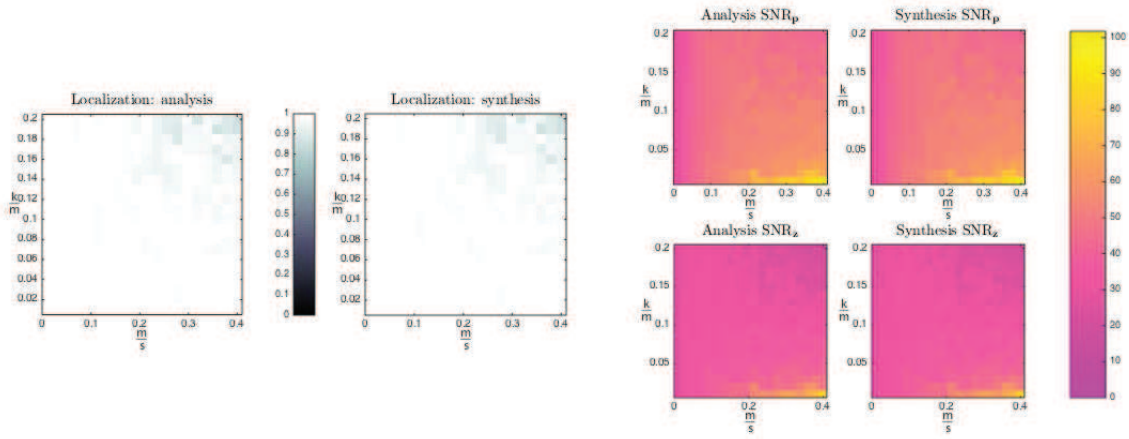
Figure 4.8 – Short source emission time ($t_e = 20$).



(a) The ℓ_1 norm: localization probability (left) and estimation SNR (right).

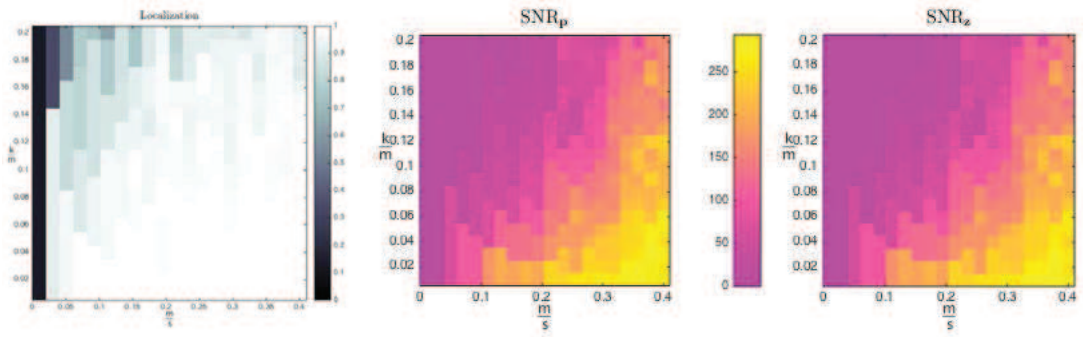


(b) The joint $\ell_{2,1}$ norm: localization probability (left) and estimation SNR (right).

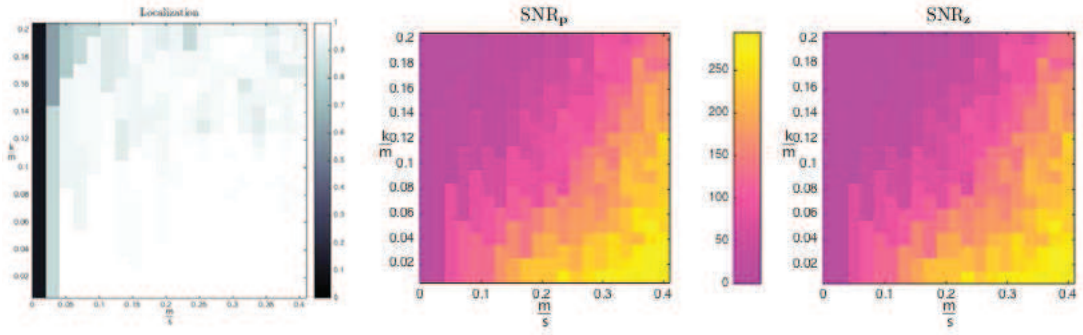


(c) The hierarchical joint $\ell_{2,1}$ norm: localization probability (left) and estimation SNR (right).

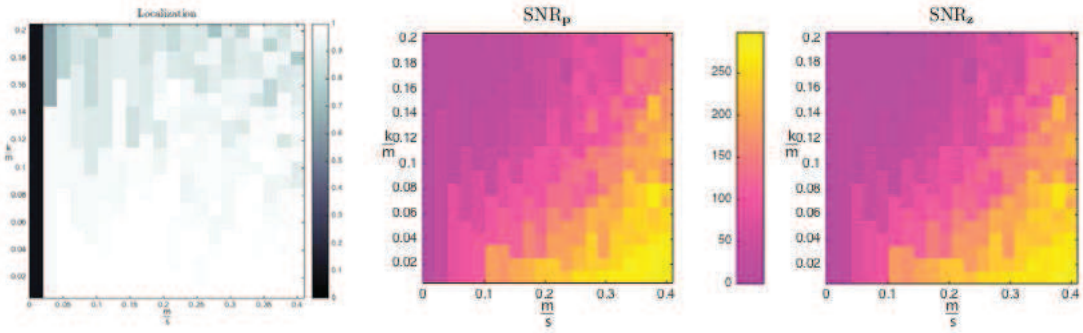
Figure 4.9 – Very short source emission time ($t_e = 5$).



(a) Localization probability (left) and estimation SNR (right) for $t_e = 45$ (long, compare with fig. 4.7).



(b) Localization probability (left) and estimation SNR (right) for $t_e = 20$ (short, compare with fig. 4.8).



(c) Localization probability (left) and estimation SNR (right) for $t_e = 5$ (very short, compare with fig. 4.9).

Figure 4.10 – GRASP performance for different emission durations.

4.4.3 Dropping the terminal condition

The terminal condition is quite a strong assumption for practical applications. Even though it is necessary if the aim is accurate recovery of the field, it might be relaxed if our primary goal is source localization. To test this setting, we simply drop the functional $f_{c_2}(\mathbf{A}_\tau \mathbf{p})$, and allow sources to emit beyond the acquisition time limit ($t_e > \tau = 0.425$s). Note that one cannot discard the initial conditions in the same way, since, in that case, even the forward model (4.21) becomes ill-posed.

The phase-transition graphs, given in figure 4.11 (experimental setting is equivalent as for figure 4.7), indicate that source localization is possible, despite violating the terminal condition requirement. Encouraged by this result, in the subsequent experiments, the terminal condition is not assumed any more. We do notice certain decrease in signal recovery performance compared to results in figure 4.7, in terms of source signal estimation ($\text{SNR}_\mathbf{z}$). Counterintuitively, we notice slight improvement in source localization performance (the far left region of the first graph in figure 4.11).

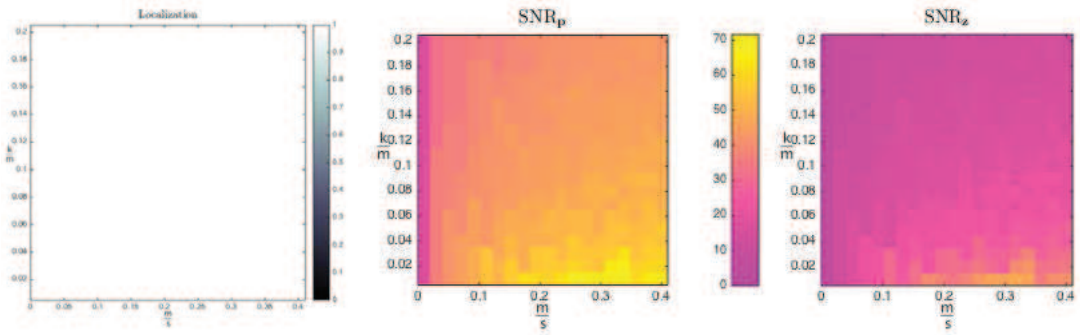
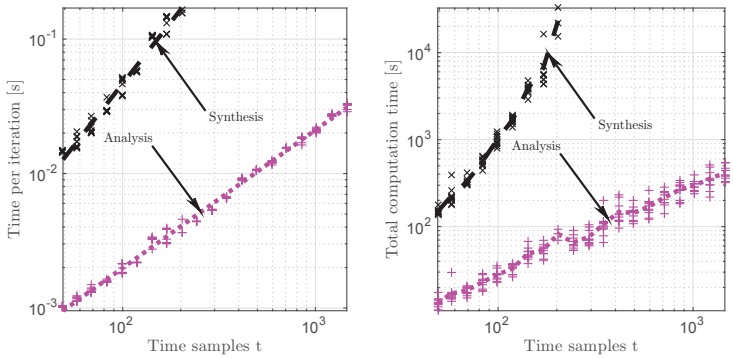


Figure 4.11 – Localization probability (left) and SNR (right) without terminal condition.

4.4.4 Scalability

A second series of source localization experiments compares the scalability potential of the two models for the acoustic source localization. Here we are interested in studying the computational cost as a function of both the problem size and the number of measurements. Using the results obtained by the experiments in the previous Subsection, we restrict the experimental setup to the regime where perfect localization is highly probable. The objective function is the hierarchical joint $\ell_{2,1}$ norm defined in subsection 4.3.4 and sound sources are modeled by white noise. Boundary condition is again the approximated Neumann condition, as in the previous Subsection.

In both analysis and synthesis case, we use the LSMR iterative method to approximately solve the least squares problem (3.28). This is necessary to avoid building and storing a fully dense coefficient matrix for the synthesis model (its storage cost would be of the order of 10^{11} bytes

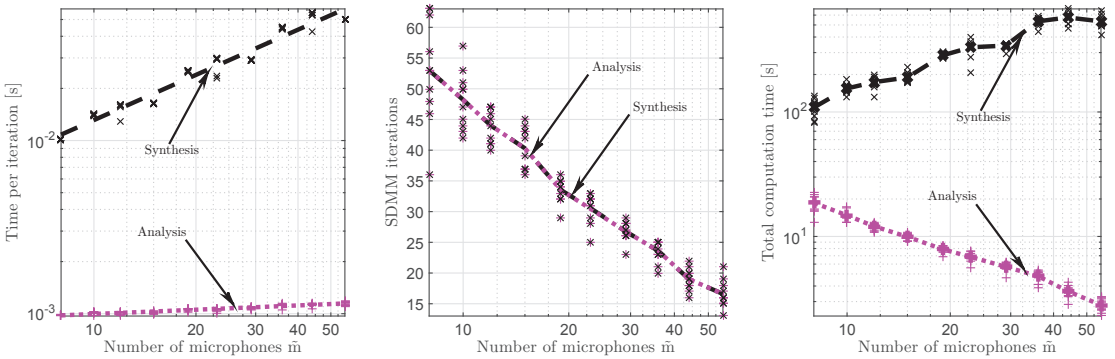


(a) Time per LSMR iteration *vs* t. (b) Total computation time *vs* t.

Figure 4.12 – Computation time relative to the problem size.

in double-precision floating point format). In order to ensure there is no bias towards any of the two models (the primal/dual residuals (A.13), defined in appendix A.2, may be influenced by norms of the involved matrices), an oracle stopping criterion is used: iterations stop when the objective function $f_r(\mathbf{z}^{(k)})$ falls below $\beta \cdot f_r(\mathbf{z})$ with $\beta = 1.1$ and $\mathbf{z}$ the ground truth signal.

Influence of the problem size We vary the number of time samples of the problem to verify that the two approaches scale differently with respect to temporal dimension, by considering 20 different values of t from 50 to 1455. The results on figure 4.12a confirm our predictions: the computational cost per iteration for the cosparse optimization problem grows linearly with t, while the cost of its synthesis counterpart is nearly quadratic. Moreover, the difference between the two models becomes striking when the total computation time is considered (figure 4.12b), since the synthesis-based problem exhibits cubic growth. Finally, we are unable to scale the synthesis problem above t = 203, due to significantly increased memory requirements (Table 4.1) and computation time.



(a) Time per LSMR iteration. (b) Number of SDMM iterations. (c) Total computation time.

Figure 4.13 – Computational cost *vs* number of microphones m.

Influence of the number of measurements Keeping the number of variables n fixed, we now vary the number of measurements in the acoustic scenario. Given the complexity analysis in section 3.5 and subsection 4.3.5, we expect the per-iteration complexity of the analysis model to be approximately independent of the number of microphones m, while the cost of the synthesis model should grow linearly with m.

The results shown on figures 4.13a and 4.13b confirm this behavior in terms of computational cost per inner iteration: in the synthesis case, it grows at almost linear rate, while being practically independent of m in the analysis case. However, the number of (outer) SDMM iterations decreases with m for both models. Overall, the total computation time increases in the synthesis case, but it *decreases* with the number of microphones in the analysis case. While perhaps a surprise, this is in line with recent theoretical studies [225] suggesting that the availability of more data may enable the acceleration of certain machine learning tasks. Here the acceleration is only revealed when adopting the analysis viewpoint rather than the synthesis one.

Scaling to 3D The optimization problems generated by the sparse analysis regularization scale much more favorably than in the case of the synthesis regularization, which allows us to test the approach in three-dimensional setting. Therefore, the spatial domain Γ is now modeled as a shoebox room of dimensions 2.5m × 2.5m × 2.5m, discretized by cubic voxels of dimension 0.25m × 0.25m × 0.25m, as shown in figure 4.14 (left). The acquisition time is set to t = 5s, and the sampling frequency is set to $f_s \approx 2.4$kHz such that it ensures marginal stability of SLF scheme (with speed of sound c = 343m/s). This discretization yields an optimization problem with roughly 1.2×10^7 variables, which is clearly out of reach for the synthesis approach. The number of microphones is set fixed to m = 20 and the number of sources is varied from k = 1 to k = 10. The emission duration of the sources is set to $t_e = 4$s, and the joint $\ell_{2,1}$ norm is used as objective.

In terms of source localization, we obtained **100% accuracy** for the given problem setup.

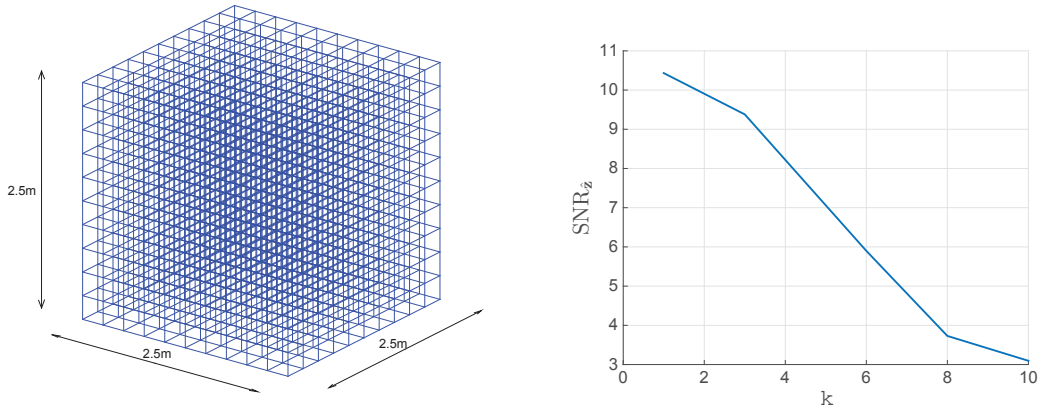


Figure 4.14 – 3D spatial discretization (left) and source recovery performance *vs* k (right).

Hence, we confirm that the localization approach is applicable, at least in this simplified three dimensional setting. The results presented in figure 4.14 (right) show the expected decrease in $\mathrm{SNR}_{\hat{\mathbf{z}}}$ when the number of sources increase. Finally, we remark that the computational time in this case is considerably high, indicating that more refined optimization approaches are needed when the number of variables is this large.

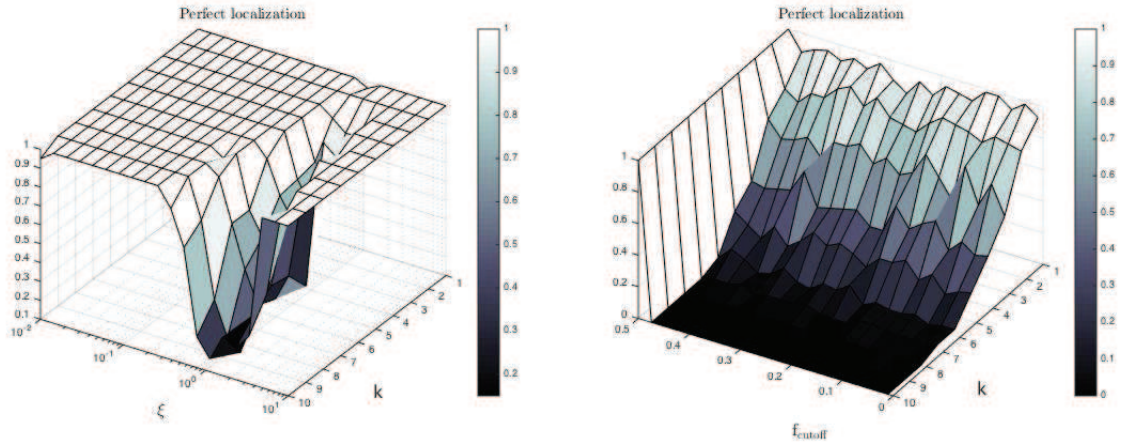
4.4.5 Model parameterization

After verifying that the synthesis and analysis model are indeed equivalent, but that only the latter is scalable to realistic dimensions, we no longer consider sparse synthesis regularization. In this subsection, we investigate the source localization performance as a function of model parameters. For this purpose, the experiments are performed in two-dimensional space of size 4.25m × 4.25m, and the acquisition time is set to 0.425s. With spatial discretization stepsize of 0.25m × 0.25m, and standard sound speed c = 343m/s, we end up with the modest scale problem involving ~ 250000 variables (for which sparse Cholesky factorization is still a viable least squares solver). All experiments are performed with the joint $\ell_{2,1}$ norm objective.

Absorption Neumann boundary condition is a valid approximation for good acoustic reflectors, such as walls or glass windows. However, in more realistic room acoustics, one needs to consider the absorption effects of the materials, such as soft floors (*e.g.* carpets) and tile ceilings. To investigate the performance as a function of absorption, we use Mur's boundary condition (3.18), and vary the specific acoustic impedance coefficient ξ between 0.01 (approximate Dirichlet condition, highly reverberant) and 10 (approximate Neumann, also highly reverberant). In between these, as mentioned in Subsection 4.1.2, when ξ is close to 1, the absorption of the scheme is the highest. In that case, the reflection coefficient in (4.10) becomes close to zero for normal incidence waves.

The results are presented in figure 4.15a. We used the same number of microphones (m = 20) for all experiments, while the number of white noise sources k is varied from 1 to 10. The results suggest that absorption has negative effect on localization performance. While this is perhaps counterintuitive from traditional point of view, where less reverberations are welcome, in physics-driven acoustic localization it is not the case. On contrary, since reverberation is implicitly embedded in the physical model, the approach actually benefits from redundant information obtained from this acoustic multipath.

Frequency A bandlimited signal model is a usual assumption in digital signal processing. To that end, we investigate localization performance as a function of source cutoff frequency. The sources are generated by lowpass filtered white noise, with normalized cutoff frequency ranging from 0.01 to 0.5 (*i.e.* no low-pass filtering applied). Again, the number of microphones is kept fixed to m = 20, and the number of sources k is varied. The boundaries are modeled by approximate Neumann condition.



(a) Specific acoustic impedance ξ.

(b) Upper bound on source frequency f_{cutoff}.

Figure 4.15 – Effects of model parameters on (empirical) perfect localization probability.

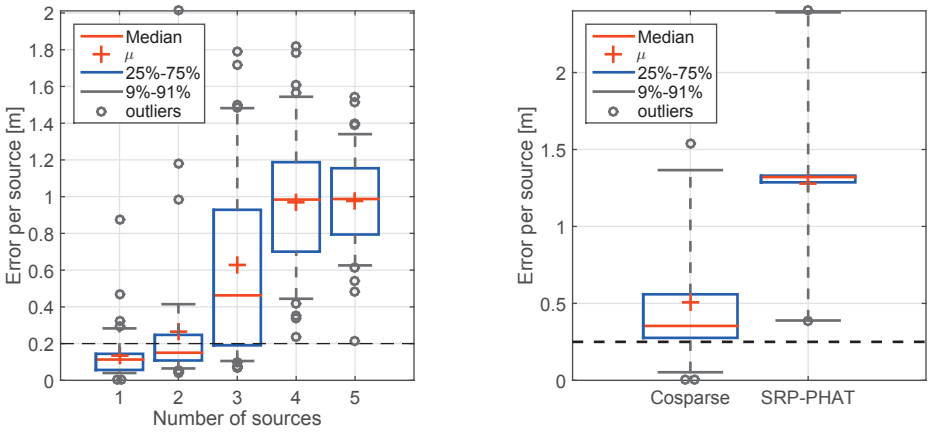
The performance graph, given in figure 4.15b, shows a dependency of perfect localization probability and localization error, with regards to the cutoff frequency and the number of sound sources k. The results lead to the conclusion that previous results, without source signal filtering, are unrealistically accurate (this is again observed by the abrupt "jump" in performance for $f_{cutoff} = 0.5$). This is related to so-called *inverse crime* phenomenon, discussed in the following subsection. Considering bandlimited sources, the performance seems to be stable with respect to cutoff frequency.

4.4.6 Robustness

In order to verify the robustness of an algorithm by numerical simulations, one should avoid committing the so-called *inverse crime* [135]. This is the scenario in which the measurement data (namely, the $\mathbf{y}$ vector in our case) is generated using the same *numerical* forward model which is later used to solve the inverse problem. That is, some modeling error should be introduced to simulate real-world conditions. Bear in mind that the additive noise is not sufficient - ideally, the algorithm should be robust to model imperfections and noise at the same time.

To avoid the inverse crime, we now consider an acoustic 2D setting where the simulated data is first generated on a fine grid of size $121 \times 121 \times 6121$, before solving the inverse problem on a coarse grid of size $25 \times 25 \times 1225$ (both grids simulate a virtual 2D space of size $5 \times 5\text{m}^2$ with recording and emission times set to 5s). Microphone positions correspond to the nodes of the crude grid, while white noise sources are arbitrarily distributed at the nodes of the fine grid. Before downsampling, the fine model data is temporally low-pass filtered to reduce the aliasing effects. The product[16] $\mathbf{A}\tilde{\mathbf{x}}$ is now only *approximately* sparse, and to account for

[16] $\tilde{\mathbf{x}}$ is the crude version of the "fine" data vector $\mathbf{x}$.



(a) Localization RMSE *vs* k (in 2D).

(b) Localization error for k = 1 (in 3D).

Figure 4.16 – Experiments with model error included.

this model error we increased the data fidelity parameter ε to 0.1. The results shown in the figure 4.16a imply that we are able to localize up to two sources with an error on the order of the crude grid's spatial step size.

It is noteworthy that *automatic* localization, based on the ℓ_2 norm of a spatial group $\hat{\mathbf{z}}_j$, is now a non-trivial task. This is due to the spatial aliasing effect, which makes actual source positions contained within "blurry zones", as opposed to distinct points in space in the inverse crime setting. Thus, it is possible that the naive clustering procedure we use to determine source locations has a significant impact on localization performance and presented results.

In the last series we consider the case with only one sound source and m = 10 microphones, in a 3D spatial domain. The number of sources is restricted in order to compare localization performance against a state-of-the-art algorithm [84], based on SRP-PHAT, that supports only the single source scenario. The "physical" settings are the same as for the previously conducted 3D experiments in the inverse crime setting (cubic room of size 2.5m × 2.5m × 2.5m, acquisition time τ = 5s, $f_s \approx$ 2.4kHz), except that now different grids are used for generating data and solving the inverse problem (fine grid consisting of 0.125m × 0.125m × 0.125m voxels, and crude grid with voxels of size 0.25m × 0.25m × 0.25m). Boundary conditions are modeling a *hard wall* structure, therefore reverberation is significant. As in the 2D case, the measurement data is low-pass filtered before processing by the algorithms.

The results in figure 4.16b reveal that cosparse acoustic localization, in this setting, performs significantly better than the SRP-PHAT algorithm. The median localization error of our method is only slightly higher than the tolerance threshold, suggesting that the estimated locations are close to the ground truth positions. On the other hand, the localization error of SRP-PHAT is on the order of half the room size, which means that it does not perform better than the uniformly random location selection. We remark that this is a highly unfavorable setting for

SRP-PHAT, due to very high reverberation times. Additionally, we feel that the simulations may be biased towards the cosparse approach, which uses the analysis operator based on the crude grid, but otherwise accurate (whereas SRP-PHAT is completely unaware of the environment). Therefore, more exhaustive experiments are needed, using the data generated by a fundamentally different model, such as the image source model [8].

4.5 Summary

This chapter was concerned with sound source localization in the physics-driven context.

After discussing physics of sound propagation, along with state-of-the-art in source localization, we proceeded to development of the localization method based on physics-driven framework. This was done by the "recipe": the physical model of sound propagation was used to formulate an optimization problem regularized by the (structured) sparse analysis or sparse synthesis data model.

The equivalence of the two models in the physics-driven setting, as well as the predicted numerical advantage of the analysis approach, have been experimentally verified. Additionally, we observed that the analysis-based optimization benefits from the increased amount of observation data, whereas the synthesis-based one exhibits an increase in computational complexity, resulting in orders of magnitude gain in processing time for the analysis versus synthesis approach. Favorable scaling capability of the analysis approach allowed us to formulate and solve the regularized inverse problem in three spatial dimensions and realistic recording time. This required solving a huge scale convex optimization problem, which was addressed by the Weighted SDMM algorithm.

We investigated the effects of various model parameters on the source localization performance. It has been observed that the increased boundary absorption has a negative impact on the accuracy of source localization, confirming our claim that the major benefit of this physics-driven approach comes from the acoustic multipath diversity. Finally, the robustness of the approach was exercised using different discretization for the generative and inverse models. In this simplified setting, the cosparse approach outperformed state-of-the-art SRP-PHAT source localization algorithm.

The following chapter extends the cosparse sound source localization concept to more challenging scenarios, where some physical parameters of the environment are unknown beforehand, and have to estimated simultaneously with localization. In addition, we discuss application of the physics-driven approach to an interesting *hearing behind walls* localization problem.

5 Extended scenarios in cosparse acoustic source localization

In the previous chapter we applied the physics-driven framework to develop a sound source localization approach. By exploiting the acoustic multipath and scaling capabilities of the cosparse regularization, we proposed a method that offers competitive and robust performance in indoor source localization. However, it is based on rather strong assumptions, regarding the accurate knowledge of domain geometry, propagation medium and boundary conditions. Such information is rarely available in practice, which may seriously limit the usefulness of the approach in real applications. Therefore, in this chapter, we aim at weakening some of these assumptions, by learning physical parameters directly from measurement data.

The methods developed for this purpose can be seen as (semi) *blind analysis operator learning* algorithms. In fact, the last decade has seen major breakthroughs in addressing the related, so-called *dictionary learning* problems [239, 92, 169, 4, 150] for the sparse synthesis data model. These are even more difficult than standard sparse recovery inverse problems, as in this case even the dictionary $\mathbf{D}$ is not known in advance. On the other hand, as demonstrated on several occasions [168, 216, 260, 93, 4], even if general-purpose dictionary is available for a certain class of signals, learning (or improving) it adaptively may offer substantial performance gains. As for the cosparse analysis data model in general, learning the analysis operators (or the "analysis dictionaries") is a recent and fruitful research axis [224, 258, 215, 261]. In our case, we would not learn the operator from scratch, but rather adapt it to the observed data, by learning some of its parameters.

More particularly, we address the problem of blind sound speed estimation in the first section of the chapter. In the second section, we challenge the problem of blind specific acoustic impedance learning, *i.e.* learning the absorption properties of the boundaries. The third section, is different: here we do not discuss the analysis operator learning problem, but instead provide a glimpse of potential applications enabled by the physics-driven framework. It is dedicated to an application we label *hearing behind walls*. In the final, fourth section we provide the summary and note the contributions. The material in this chapter is mostly based on publications [26, 25, 142].

5.1 Blind estimation of sound speed

It was already mentioned, in subsection 4.1.1, that the speed of sound is considered only *approximately* constant, *i.e.* the propagating medium is only approximately isotropic. In reality, however, the speed of sound is a slowly varying function of position and time, *e.g.* due to differences in temperature in different regions of the spatial domain Γ. Even if we assume that the sound speed is indeed constant, its exact value might be unknown (it could be estimated from (4.6) if the temperature information is available). If the speed of sound estimate is very inaccurate, the physical model embedded in the analysis operator will be wrong. The effects of such model inaccuracies have been exhaustively investigated [126, 50], and are known to significantly alter regularization performance.

Therefore, our goal here is to simultaneously recover the pressure signal (more precisely, to localize the sound sources) and estimate the *slowly varying* sound speed $c(\mathbf{r}, t)$. We are particularly interested in this scenario for practical reasons. Imagine a room equipped with an air-conditioner or a radiator that are turned on during the recording process. The induced temperature gradient slowly changes in space and time due to the diffusion process.

5.1.1 The BLESS algorithm

Let us first formally define the inverse problem. Consider the FDTD-SLF discretization scheme in appendix B.1 (we use 2D for clarity, but this easily extends to 3D case). Instead of a fixed sound speed c, we now have varying $c(\mathbf{r}, t) = c^t_{i,j}$. By denoting $\mathbf{q} = \mathbf{c}^{-2} = \left[c^1_{1,1}\, c^1_{2,1} \ldots c^t_{i,j} \ldots\right]^{T^{-2}} \in \mathbb{R}^n$, we can represent the analysis operator $\mathbf{A}$ as follows:

$$\mathbf{A} = \mathbf{A}_1 + \mathtt{diag}\left(\mathbf{q}\right)\mathbf{A}_2, \tag{5.1}$$

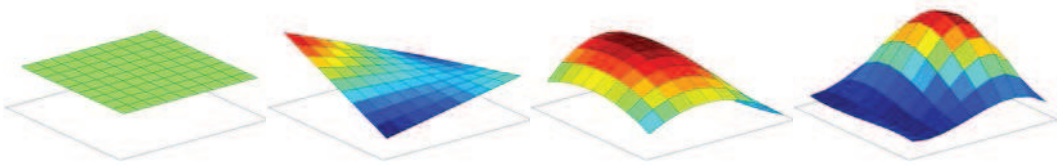
where the singular matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ are obtained after factorizing $\mathbf{q}$ in (B.1). This leads to the optimization problem involving both the pressure $\mathbf{p}$ and the unknown inverse of squared sound speed $\mathbf{q}$ (here we use $f_d = \lambda f_q$ from (1.5) for the data fidelity term):

$$\underset{\mathbf{p},\mathbf{q}}{\text{minimize}}\; f_r(\mathbf{A}\mathbf{p}) + \lambda\|\mathbf{M}\mathbf{p} - \mathbf{y}\|_2^2 + \chi_{\ell_2=0}\left(\begin{bmatrix}\mathbf{A}_0\\ \mathbf{A}_{\partial\Gamma}\end{bmatrix}\mathbf{p}\right) \quad \text{s.t.} \quad \mathbf{A} = \mathbf{A}_1 + \mathtt{diag}\left(\mathbf{q}\right)\mathbf{A}_2. \tag{5.2}$$

Recall that $\mathbf{A}_0$ and $\mathbf{A}_{\partial\Gamma}$ are row-reduced matrices of $\mathbf{A}$, corresponding to initial and boundary conditions, and that the terminal condition ($t = \tau$) has been abandoned as of subsection 4.4.3. For clarity, we will denote $\mathbf{B} = \begin{bmatrix}\mathbf{A}_0\\ \mathbf{A}_{\partial\Gamma}\end{bmatrix}$, and, since this is only a submatrix of $\mathbf{A}$, it can be also represented as a sum

$$\mathbf{B} = \mathbf{B}_1 + \mathtt{diag}\left(\mathbf{q}_{0,\partial\Gamma}\right)\mathbf{B}_2 \tag{5.3}$$

where $\mathtt{diag}\left(\mathbf{q}_{0,\partial\Gamma}\right)$ denotes the corresponding submatrix of $\mathtt{diag}\left(\mathbf{q}\right)$. The regularizer f_r is again the joint $\ell_{2,1}$ norm (4.26).



(a) r = 1 (b) r = 2 (c) r = 3 (d) r = 4

Figure 5.1 – Shape of $\mathbf{q} = \mathbf{F}^{[r]}_{\texttt{null}}\mathbf{a}$ for different values of r.

Due to the presence of the bilinear form $\texttt{diag}\left(\mathbf{q}\right)\mathbf{A}_2\mathbf{p}$, the problem (5.2) is non-convex. However, it is *biconvex, i.e.* it is convex in each variable (but not jointly!). This allows for simple alternating minimization with respect to $\mathbf{p}$ and $\mathbf{q}$, which is locally convergent[1], by definition:

$$\begin{aligned}
\mathbf{p}^{(\mathrm{i}+1)} &= \underset{\mathbf{p}}{\arg\min}\; f_r(\mathbf{A}^{(\mathrm{i})}\mathbf{p}) + \lambda\|\mathbf{M}\mathbf{p}-\mathbf{y}\|_2^2 + \chi_{\ell_2=0}\left(\mathbf{B}^{(\mathrm{i})}\mathbf{p}\right) \\
\mathbf{q}^{(\mathrm{i}+1)} &= \underset{\mathbf{q}}{\arg\min}\; f_r\left(\mathbf{A}_1\mathbf{p}^{(\mathrm{i}+1)} + \texttt{diag}\left(\mathbf{A}_2\mathbf{p}^{(\mathrm{i}+1)}\right)\mathbf{q}\right) + \chi_{\ell_2=0}\left(\mathbf{B}_1\mathbf{p}^{(\mathrm{i}+1)} + \texttt{diag}\left(\mathbf{B}_2\mathbf{p}^{(\mathrm{i}+1)}\right)\mathbf{q}_{0,\partial\Gamma}\right) \\
\mathbf{A}^{(\mathrm{i}+1)} &= \mathbf{A}_1 + \texttt{diag}\left(\mathbf{q}^{(\mathrm{i}+1)}\right)\mathbf{A}_2, \quad \mathbf{B}^{(\mathrm{i}+1)} = \begin{bmatrix}\mathbf{A}_0^{(\mathrm{i}+1)} \\ \mathbf{A}_{\partial\Gamma}^{(\mathrm{i}+1)}\end{bmatrix}.
\end{aligned} \tag{5.4}$$

Unfortunately, iterating this scheme reveals that there are points $\hat{\mathbf{p}}$ and $\hat{\mathbf{q}}$ which yield very low objective value f_r, but do not recover the original signals. If we inspect the iterates, it becomes clear why this is the case. Since the matrix $\mathbf{A}_1$ is singular, for the squared inverse speed of sound estimate $\hat{\mathbf{q}} = \mathbf{0}$, the corresponding pressure field estimate is $\hat{\mathbf{p}} \in \texttt{null}\left(\left[\begin{smallmatrix}\mathbf{A}_1 \\ \mathbf{M}\end{smallmatrix}\right]\right)$ (there is always a non-trivial intersection between the null spaces of the two matrices[2]). Obviously, these are not the estimates we are looking for, and to avoid this behavior, we need to bound the estimate $\hat{\mathbf{q}}$. Fortunately, one can easily devise lower and upper bounds for the speed of sound in practice: for instance, the upper bound may be given by the maximal allowed propagation speed prescribed by the CFL condition (appendix B.1), while the lower bound can be based on the lowest considered temperature of the environment (using the formula (4.6)).

Moreover, the intrinsic degrees of freedom in the structure of a real world medium are often much smaller. Recalling the assumption of a slowly-varying speed of sound (analogously, a slowly varying $\mathbf{q}$), we may decide to promote some level of smoothness in the estimate $\hat{\mathbf{q}}$. One way to do it, is to assume that the r^{th} derivative approximation of the field $\mathbf{q}$ along each spatiotemporal dimension is close to zero, *i.e.*

$$\mathbf{F}^{[r]}\mathbf{q} = \mathbf{0} \Leftrightarrow \mathbf{q} = \mathbf{F}^{[r]}_{\texttt{null}}\mathbf{a} \tag{5.5}$$

where $\mathbf{F}^{[r]}_{\texttt{null}}$ is a null space basis of the r^{th} order finite difference matrix $\mathbf{F}^{[r]}$. This is actually the space of sampled polynomials of order $r-1$, as illustrated in figure 5.1. Given the assumption

[1] In terms of the objective value [114].

[2] The last s columns of the matrix $\mathbf{A}_1$ are all-zero, as well as most of the corresponding columns of $\mathbf{M}$.

(5.5), the number of degrees of freedom is dramatically reduced (since the dimension of $\mathtt{null}\left(\mathbf{F}^{[r]}\right)$ is only r^d, where d is the number of dimensions).

By usual variable splitting, the optimization problem is now modified to

$$\begin{aligned} \underset{\mathbf{p},\mathbf{z},\mathbf{a}}{\text{minimize}}\ & f_r(\mathbf{z}) + \lambda\|\mathbf{M}\mathbf{p}-\mathbf{y}\|_2^2 + \chi_{\ell_2=0}\left(\mathbf{z}_{0,\partial\Gamma}\right) + \chi_{\alpha_{\min}\preceq\cdot\preceq\alpha_{\max}}\left(\mathbf{F}^{[r]}_{\mathtt{null}}\mathbf{a}\right) \\ \text{s.t.}\quad & \mathbf{z} = \left[\mathbf{A}_1 + \mathtt{diag}\left(\mathbf{F}^{[r]}_{\mathtt{null}}\mathbf{a}\right)\mathbf{A}_2\right]\mathbf{p}. \end{aligned} \tag{5.6}$$

We could apply alternating minimization in order to find a local optimum of the optimization problem (5.6). However, we will instead use a biconvex version of ADMM, which is known to exhibit better empirical performance [39, 1, 253, 232]. Following the formulation of augmented Lagrangian (A.2), the development is straightforward. We term the algorithm *Blind Localization and Estimation of Sound Speed (BLESS)* (the pseudocode is given in Algorithm 3).

Algorithm 3 *BLESS*

Require: $\mathbf{y}$, $\mathbf{M}$, $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{F}^{[r]}$, λ, $\alpha_{\min}$, $\alpha_{\max}$, μ_1, μ_2

$\mathbf{p}^{(0)} = \mathbf{M}^{\mathsf{T}}\mathbf{y}$, $\mathbf{a}^{(0)} = \mathbf{1}\cdot 0.5(\alpha_{\max}-\alpha_{\min})$, $\mathbf{w}^{(0)} = \mathbf{0}$, $\mathbf{u}_1^{(0)} = \mathbf{0}$, $\mathbf{u}_2^{(0)} = \mathbf{0}$

$\mathbf{z}^{(i+1)} = P_\Phi\left(\mathrm{prox}_{1/\lambda f}\left(\mathbf{A}^{(i)}\mathbf{p}^{(i)} + \mathbf{u}_1^{(i)}\right)\right)$

$\mathbf{p}^{(i+1)} = \arg\min_{\mathbf{p}} \frac{\mu_1}{2}\|\mathbf{A}^{(i)}\mathbf{p} - \mathbf{z}^{(i+1)} + \mathbf{u}_1^{(i)}\|_2^2 + \frac{\mu_2}{2}\|\mathbf{M}\mathbf{p} - \mathbf{y} + \mathbf{u}_2^{(i)}\|_2^2$

$\mathbf{a}^{(i+1)} = \arg\min_{\mathbf{a}} \|\mathtt{diag}\left(\mathbf{A}_2\mathbf{p}^{(i+1)}\right)\mathbf{F}^{[r]}\mathbf{a} - \mathbf{z}^{(i+1)} + \mathbf{A}_1\mathbf{p}^{(i+1)} + \mathbf{u}_1^{(i)}\|_2^2$
subject to $\alpha_{\min} \preceq \mathbf{F}^{[r]}\mathbf{a} \preceq \alpha_{\max}$

$\mathbf{A}^{(i+1)} = \mathbf{A}_1 + \mathtt{diag}\left(\mathbf{F}^{[r]}\mathbf{a}^{(i+1)}\right)\mathbf{A}_2$

$\mathbf{u}_1^{(i+1)} = \mathbf{u}_1^{(i)} + \mathbf{A}^{(i+1)}\mathbf{p}^{(i+1)} - \mathbf{z}^{(i+1)}$

$\mathbf{u}_2^{(i+1)} = \mathbf{u}_2^{(i)} + \mathbf{M}\mathbf{p}^{(i+1)} - \mathbf{y}$

if convergence **then**
    terminate
else
    $i \leftarrow i+1$
    go to 2
end if

return $\hat{\mathbf{p}} = \mathbf{p}^{(i+1)}$, $\hat{\mathbf{z}} = \mathbf{z}^{(i+1)}$, $\hat{\mathbf{c}} = \left[\mathbf{F}^{[r]}\mathbf{a}^{(i+1)}\right]^{-1/2}$

$P_\Phi(\cdot)$ is the projection operator on the constraint set $\Phi = \left\{\mathbf{z} \mid \mathbf{z}_{0,\partial\Gamma} = \mathbf{0}\right\}$, while f_r is the joint $\ell_{2,1}$ norm, as mentioned before. The $\mathbf{z}^{(i+1)}$ update is closed form evaluation of $\mathrm{prox}_{1/\lambda\ell_{2,1}+\chi_\Phi}(\cdot)$.

5.1.2 Simulations

In order to demonstrate the joint estimation performance of the proposed approach, a rectangular 15×15 grid is simulated by placing $1 \le k \le 18$ white noise sources uniformly at random in space. Sources emit random frequencies for a duration of 100 time samples, which is equal to the experiment duration (acquisition time τ). The boundaries are modeled by Neumann boundary condition, and the field $\mathbf{q}$ is generated by randomly selecting the vector $\tilde{\mathbf{a}}$ and then

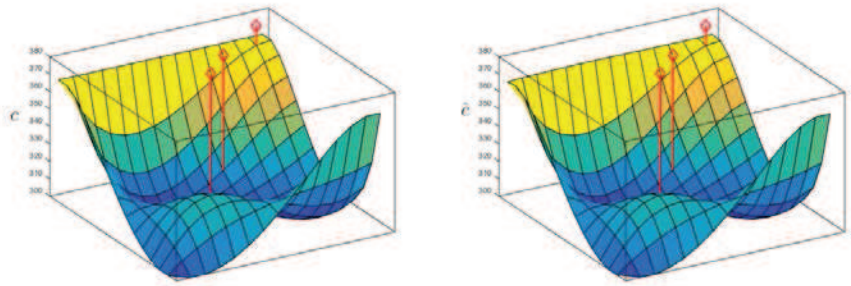


Figure 5.2 – The original sound speed (left) $\mathbf{c}$ and the estimate (right) $\hat{\mathbf{c}}$ (the diamond markers indicate the source spatial positions).

solving the optimization problem

$$\mathbf{q} = \mathbf{F}^{[\mathsf{r}]}_{\mathtt{null}} \cdot \arg\min_{\mathbf{a}} \|\mathbf{a} - \tilde{\mathbf{a}}\|_2^2 \quad \text{s.t.} \quad \alpha_{\mathtt{min}} \preceq \mathbf{F}^{[\mathsf{r}]}_{\mathtt{null}} \mathbf{a} \preceq \alpha_{\mathtt{max}}$$

with parameter r set to 1, 2, 3 and 4. The pressure field is measured by $3 \leq \mathsf{m} \leq 90$ randomly located *noiseless* microphones and the field parameters $\mathbf{a}$ and $\mathbf{p}$ are estimated by means of the BLESS algorithm. In current implementation, we enforce only *spatial* smoothness, meaning that the finite difference matrix $\mathbf{F}^{[\mathsf{r}]}$ applies differentiation exclusively along spatial dimensions (the speed of sound is constant over time). In each setting we conduct 50 experiments, and after every simulation, k positions with highest energy are chosen as estimates of the source locations. When localization is successful, the parameter field is often perfectly recovered, as demonstrated in figure 5.2. However, the degrees of freedom within $\mathbf{q}$ increase with r and, therefore, the performance is expected to decrease, which is confirmed experimentally in figure 5.3. Overall, the results are promising, but the experiments with spatiotemporally varying speed of sound have not been conducted, and will be the subject of future work.

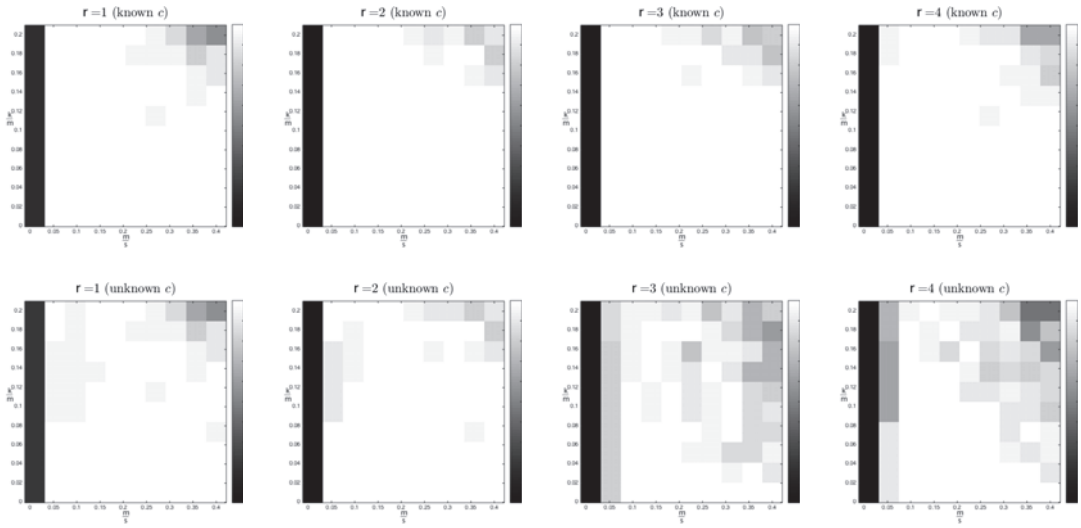


Figure 5.3 – Empirical localization probability with known (top) and estimated (bottom) sound speed.

5.2 Blind estimation of specific acoustic impedance

Fortunately, while estimating the sound speed in air is important, in many cases the initial guess can be accurate enough (for instance, if the temperature of the room is in approximately steady-state, calculating c from (4.6) is sufficient). On the other hand, guessing the specific acoustic impedance ξ in (4.10) is much more difficult. In practice, one may consult standardized tables of absorption properties of different structures (*e.g.* from construction engineering literature), but this is very inflexible and may be inaccurate (in practice, we need to be certain about the type of material the boundaries consist of). Another way is to physically measure the acoustic impedance (*e.g.* [190, 70, 71]) of the concerned room and use this information to build the analysis operator. Unfortunately, this requires specific hardware setup and calibration, which makes it also inflexible when considering different acoustic environments. A less demanding approach is to use only a microphone array and a known sound source, as presented in [11]. This way one would parametrize the analysis operator for later use with unknown sources.

Our goal is to go a step beyond aforementioned approaches: we want a method that can simultaneously estimate the specific impedance ξ *and* the acoustic pressure/sound source component (thus also perform source localization). Hence, we are not given any new source field information, except spatial sparsity assumed earlier. Again, this approach is practically motivated - the goal is to develop a flexible method, which may be used later on in the real physical environment where the wall (ceiling, floor etc) structure is not given in advance. We will see that the problem shares some similarities with the speed of sound estimation problem. However, there are some notable differences, the most distinct one coming from the natural observation that boundaries do not change through time (if we exclude opening doors and similar actions). This assumption will help us reduce the dimensionality of the problem, which is, otherwise, very ill-posed.

5.2.1 The CALAIS algorithm

The development of the approach begins in a similar fashion as in the speed of sound estimation case (same as before, we will discuss the 2D case, but the main principles can be straightforwardly extended to 3D setting). Again, the FDTD-SLF discretization in appendix B.1 is considered, but we are now interested in discretization of the boundary terms (B.3) and (B.4). By grouping coefficients involving $\xi_{\cdot,\cdot}^{-1}$ in these expressions, we can represent the boundary part $\mathbf{A}_{\partial\Gamma}$ of the analysis operator as the sum of the following two component matrices:

$$\mathbf{A}_{\partial\Gamma} = \underbrace{\begin{bmatrix} \tilde{\mathbf{A}}_{\partial\Gamma_1} & & \\ & \tilde{\mathbf{A}}_{\partial\Gamma_1} & \\ & & \ddots \\ & & & \tilde{\mathbf{A}}_{\partial\Gamma_1} \end{bmatrix}}_{\mathbf{A}_{\partial\Gamma_1}\in\mathbb{R}^{\mathrm{bt}\times\mathrm{n}}} + \mathtt{diag}(\mathbf{Sb}) \underbrace{\begin{bmatrix} \tilde{\mathbf{A}}_{\partial\Gamma_2} & & \\ & \tilde{\mathbf{A}}_{\partial\Gamma_2} & \\ & & \ddots \\ & & & \tilde{\mathbf{A}}_{\partial\Gamma_2} \end{bmatrix}}_{\mathbf{A}_{\partial\Gamma_2}\in\mathbb{R}^{\mathrm{bt}\times\mathrm{n}}}. \tag{5.7}$$

The vector $\mathbf{b} = \left[\xi_{1,1}\, \xi_{2,1}\, \ldots\, \xi_{\mathsf{i},\mathsf{j}}\, \ldots\right]^{-\mathsf{T}} \in \mathbb{R}^{\tilde{\mathsf{b}}}$ contains inverse acoustic impedances, *i.e. specific acoustic admittances*. The number of elements in $\mathbf{b}$ is equal to $\tilde{\mathsf{b}} = \mathsf{b} + (\mathsf{d}-2)\mathsf{e} + (\mathsf{d}-1)\mathsf{c}$, where b is the number of spatial elements corresponding to discretized boundary, e is the number of edges, c the number of corner nodes and d is the number of spatial dimensions (*i.e.* $\mathsf{d} = 2$ for 2D)[3]. The blocks $\tilde{\mathbf{A}}_{\partial\Gamma_1}$ and $\tilde{\mathbf{A}}_{\partial\Gamma_2}$ contain fixed coefficients of $\mathbf{A}_{\partial\Gamma}$, while the matrix $\mathbf{S} \in \mathbb{R}^{\mathsf{bt}\times\tilde{\mathsf{b}}}$ is a row-wise concatenation of blocks $\tilde{\mathbf{S}} \in \mathbb{R}^{\mathsf{b}\times\tilde{\mathsf{b}}}$. Each individual block $\tilde{\mathbf{S}}$ is almost an identity matrix, except for the edge and corner nodes, such as in (B.4). We also define the matrix $\mathbf{A}_{\Gamma\setminus\partial\Gamma}$, which is the analysis operator $\mathbf{A}$ *without* the boundary rows.

Before proceeding to formulation of an optimization problem, we will discuss the nature of the admittance vector $\mathbf{b}$. First, its elements are always positive, since $\xi > 0$ (in practice, we observed that the positivity constraint $\chi_{\mathbb{R}^{\mathsf{b}}_+}(\mathbf{b})$ does not have an influential role). Second, they are parametrizing enclosure boundaries, which, for rooms, may be composed of walls, floor, ceiling, windows etc. One immediately realizes that, at least on macroscopic scale, these structures are approximately homogeneous. For instance, if a side of the room is occupied by a concrete wall, all admittances corresponding to this part of the boundary should have very similar values. Hence, $\mathbf{b}$ admits even stronger *piecewise constant* model (of course, one would need to take care of the ordering of elements within $\mathbf{b}$). This is a weak assumption, and it usually holds in practice unless the discretization is very crude.

Now, having introduced the involved matrices and assumptions, we proceed to formulating the optimization problem:

$$\begin{aligned}\underset{\mathbf{p},\mathbf{b}}{\text{minimize}}\; & f_r(\mathbf{A}_{\Gamma\setminus\partial\Gamma}\mathbf{p}) + \chi_{\ell_2=0}\left(\mathbf{A}_0\mathbf{p}\right) + \chi_{\ell_2\leq\varepsilon}\left(\mathbf{M}\mathbf{p}-\mathbf{y}\right) + \|\mathbf{b}\|_{\mathrm{TV}} + \chi_{\mathbb{R}^{\mathsf{b}}_+}(\mathbf{b}) + \lambda\|\mathbf{A}_{\partial\Gamma}\mathbf{p}\|_2^2 \\ & \text{s.t. } \mathbf{A}_{\partial\Gamma} = \mathbf{A}_{\partial\Gamma_1} + \mathtt{diag}(\mathbf{S}\mathbf{b})\,\mathbf{A}_{\partial\Gamma_2}, \end{aligned} \tag{5.8}$$

where $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation norm (1.10), mentioned in chapter 1. By approximating the gradient with, *e.g.* the previously defined $\mathbf{F}^{[1]}$ matrix operator, this penalty simply writes as $\|\mathbf{b}\|_{\mathrm{TV}} = \|\mathbf{F}^{[1]}\mathbf{b}\|_1$. It is known to promote piecewise constant solutions. We remark that, in principle, the cost (5.8) could be straightforwardly extended to account for the unknown sound speed, in an attempt to jointly estimate the two parameters. In this work, however, we assume simplified scenario where the sound speed is known beforehand.

Again, there is a bilinear form embedded into product $\mathbf{A}_{\partial\Gamma}\mathbf{p}$, and thus, (5.8) is another biconvex problem. Same as before, we will apply the biconvex ADMM heuristics. The conceptual pseudocode is given in Algorithm 4, and we term it *Cosparse Acoustic Localization, Acoustic Impedance estimation and Signal recovery (CALAIS)*. Note that evaluating the minimizers of the intermediate convex problems (steps 2 and 3) is left to *black-box* approaches, *e.g.* we used the Weighted SDMM algorithm. To accelerate computation, it is warm-started by the estimate obtained in the preceding CALAIS iteration.

[3] This corresponds to the number of impedance coefficients needed to model different parts of the boundary, as explained in appendix B.1.

Algorithm 4 *CALAIS*

Require: $\mathbf{y}$, $\mathbf{M}$, $\mathbf{A}_{\partial\Gamma_1}$, $\mathbf{A}_{\partial\Gamma_2}$, $\mathbf{A}_0$, $\mathbf{A}_{\Gamma\setminus\partial\Gamma}$, $\mathbf{b}^{(0)}$, $\mathbf{S}$, ε, λ

1: $\mathbf{u}^{(0)} = \mathbf{0}$, $\mathbf{A}_{\partial\Gamma}^{(0)} = \mathbf{A}_{\partial\Gamma_1} + \mathtt{diag}\left(\mathbf{S}\mathbf{b}^{(0)}\right)\mathbf{A}_{\partial\Gamma_2}$

2: $\mathbf{p}^{(i+1)} = \arg\min_{\mathbf{p}} f_r(\mathbf{A}_{\Gamma\setminus\partial\Gamma}\mathbf{p}) + \lambda\|\mathbf{A}_{\partial\Gamma}^{(i)}\mathbf{p} + \mathbf{u}^{(i)}\|_2^2 \quad \text{s.t. } \mathbf{A}_0\mathbf{p} = \mathbf{0}, \quad \|\mathbf{M}\mathbf{p} - \mathbf{y}\|_2 \leq \varepsilon$

3: $\mathbf{b}^{(i+1)} = \arg\min_{\mathbf{b}} \|\mathbf{b}\|_{\mathrm{TV}} + \lambda\|\mathtt{diag}\left(\mathbf{A}_{\partial\Gamma_2}\mathbf{p}^{(i+1)}\right)\mathbf{S}\mathbf{b} + \mathbf{A}_{\partial\Gamma_1}\mathbf{p}^{(i+1)} + \mathbf{u}^{(i)}\|_2^2 \quad \text{s.t. } \mathbf{b} \geq \mathbf{0}$

4: $\mathbf{A}_{\partial\Gamma}^{(i+1)} = \mathbf{A}_{\partial\Gamma_1} + \mathtt{diag}\left(\mathbf{S}\mathbf{b}^{(i+1)}\right)\mathbf{A}_{\partial\Gamma_2}$

5: $\mathbf{u}^{(i+1)} = \mathbf{u}^{(i)} + \mathtt{diag}\left(\mathbf{A}_{\partial\Gamma_2}\mathbf{p}^{(i+1)}\right)\mathbf{S}\mathbf{b}^{(i+1)} + \mathbf{A}_{\partial\Gamma_1}\mathbf{p}^{(i+1)}$

6: **if** convergence **then**

7: terminate

8: **else**

9: $i \leftarrow i + 1$

10: go to 2

11: **end if**

12: **return** $\hat{\mathbf{p}} = \mathbf{p}^{(i+1)}$, $\hat{\mathbf{b}} = \mathbf{b}^{(i+1)}$, $\hat{\mathbf{A}}_{\partial\Gamma} = \mathbf{A}_{\partial\Gamma}^{(i+1)}$

5.2.2 Simulations

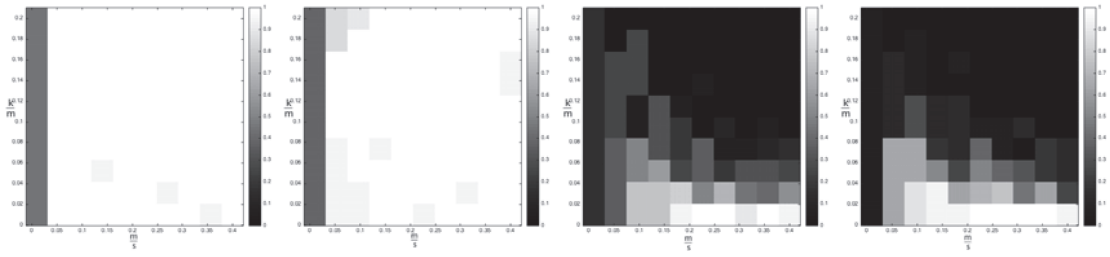
Experimental setup All experiments are conducted in the simulated two dimensional (rectangular) room of size 15×15m, with acquisition time set to t = 1s and speed of sound c = 343m/s. With spatial step size of $1\text{m} \times 1\text{m}$, this corresponds to ~ 110000 discrete points in space and time. For the experiments where we avoid the inverse crime, different grid is used for generating data: in this case, the spatial step size is $0.5\text{m} \times 0.5\text{m}$, which yields a discrete model of size ~ 450000. The specific acoustic impedance parameters are set to ξ_1 and ξ_2, for each pair of opposite "walls". The number of microphones m and number of sources k are varied, and the final results are averages over 20 realizations. White noise sound sources emit during the whole acquisition period $(0, \tau]$, except for t = 0, due to the homogeneous initial conditions. Four series of experiments are performed:

1. In the first series, we assume idealized (inverse crime) conditions: noiseless measurements and an accurate geometric model. The impedance values are $\xi_1 = 100$ (hard wall) and $\xi_2 = 0.3$ (soft wall).

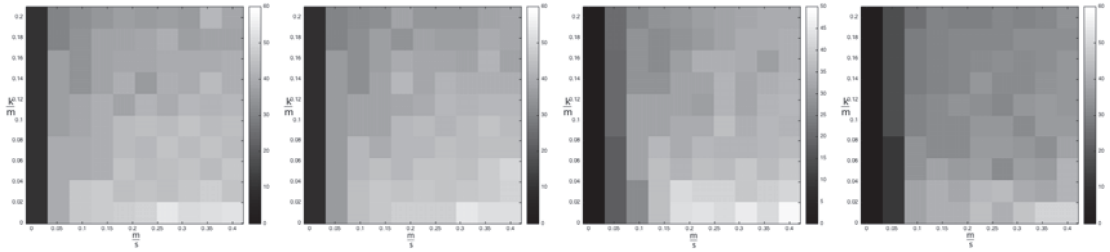
2. In the second series, we still consider inverse crime conditions, but with $\xi_1 = 100$ (hard wall) and $\xi_2 = 0.9$ (absorbing) values. We recall the results of subsection 4.4.5, stating that the absorbing conditions decrease localization performance. However, this is a somewhat different setting, since *not all* boundaries are absorbent.

3. In the third series, we preserve the previous setup, but use different generative and inversion discretization grids, to avoid the inverse crime. We remark that the sources are allowed to take positions anywhere in the generative (finer) grid.

4. In the fourth series, the setup is further adversed: we add white Gaussian measurement noise, such that *per-sample* SNR is around 20dB, and the piecewise constant model is compromised by corrupting the vector $\boldsymbol{\xi}$ with AWGN, distributed as $\mathcal{N}(0, 0.01)$.



(a) Empirical localization probability.



(b) Signal-to-(estimation) noise ratio corresponding to the acoustic pressure estimation ($\mathrm{SNR}_{\hat{\mathbf{p}}}$).

Figure 5.4 – From left to right: i) inverse crime, $\xi_2 = 0.3$; ii) inverse crime, $\xi_2 = 0.9$; iii) grid model error, $\xi_2 = 0.9$, noiseless; iv) grid and impedance model errors, $\xi_2 = 0.9$, SNR = 20dB.

In each experiment, we first compute the minimal localization RMSE. In the inverse crime setting, localization is considered successful when this error is zero. In the non-inverse crime setting, an error on the order of crude grid's stepsize is tolerated. Having defined the successfulness criterion, we can calculate the empirical probability of accurate source localization (RMSE lower than the tolerance). Additionally, we compute the signal estimation error, in terms of signal-to-noise ratio $\mathrm{SNR}_{\hat{\mathbf{p}}} = 20\log\frac{\|\mathbf{p}\|_2}{\|\mathbf{p}-\hat{\mathbf{p}}\|_2}$.

In all experiments, we use $\lambda = 0.1$ and the initial estimate $\mathbf{b}^{(0)} = \mathbf{1}$ (the vector of all ones). The stopping criterion is based on the relative distance between the objective function (5.8) values (denoted here by $\mathrm{f}^{(\mathrm{i})}$) at two successive iterations:

$$\frac{|\mathrm{f}^{(\mathrm{i})} - \mathrm{f}^{(\mathrm{i}-1)}|}{\min\left(\mathrm{f}^{(\mathrm{i})}, \mathrm{f}^{(\mathrm{i}-1)}\right)} \leq 10^{-4}.$$

The data fidelity parameter ε is set to 0, in all experiments. Thus, the noise variance is assumed unknown beforehand.

Results According to phase transition graphs in figure 5.4, in the inverse crime setting, the algorithm achieves high localization accuracy and signal recovery, and is almost unaffected by the change of impedance value ξ_2 (admittance estimate accuracy is demonstrated in figure 5.5 - left). This suggest that, as long as there is *any* multipath diversity in the system, the physics-driven approach will be able to exploit it. Moreover, the CALAIS algorithm achieves almost identical results as our standard physics-driven localization (comparable to results in figure 4.7), despite the fact that the impedance parameters are unknown.

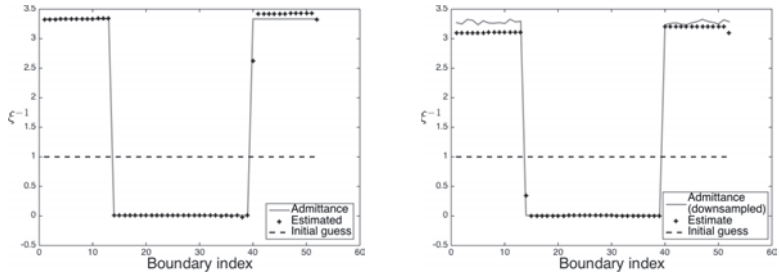


Figure 5.5 – Original and estimated acoustic admittances in the inverse crime (left) and non-inverse crime (right, the ground truth admittance has been downsampled for better visualization) settings.

When the models are inaccurate, the localization performance deteriorates, as shown in the two top right diagrams in figure 5.4. Model error affects signal recovery performance to a lesser extent (bottom row of figure 5.4), which again[4] suggests that location estimation probably needs to be more sophisticated in realistic conditions. The CALAIS algorithm seems robust to moderate additive noise and reduced accuracy of the piecewise constant model (fig. 5.4, far right column). Moreover, a high $\mathrm{SNR}_{\hat{\mathbf{p}}}$ implies accurate admittance estimation, as presented in fig. 5.5. Figure 5.6 shows the RMSE statistics when the number of sources is fixed to $k \in \{1, 2, 3, 4\}$ and the number of microphones is varied ($3 \leq m \leq 90$), for the most adverse experimental setting (4). These figures (based on the same series of experiments) give a different picture than the phase transition diagrams in fig. 5.4. Observing the median results in fig. 5.6, we conclude that localization is successful provided that $k \leq 2$ and $m > 12$. Moreover, even though the RMSE is above the tolerance for $k > 2$, it is lower than twice the crude grid's stepsize, suggesting that the sources are localized in their immediate neighborhoods.

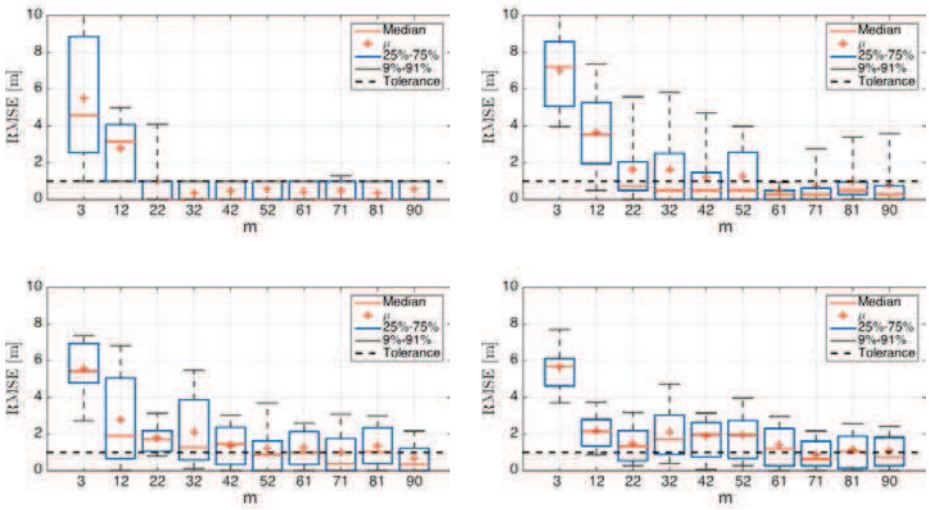


Figure 5.6 – Non-inverse crime setting (top: $k = 1$ and $k = 2$; bottom: $k = 3$ (left), $k = 4$).

[4] We advocated this in subsection 4.4.6 of the previous chapter.

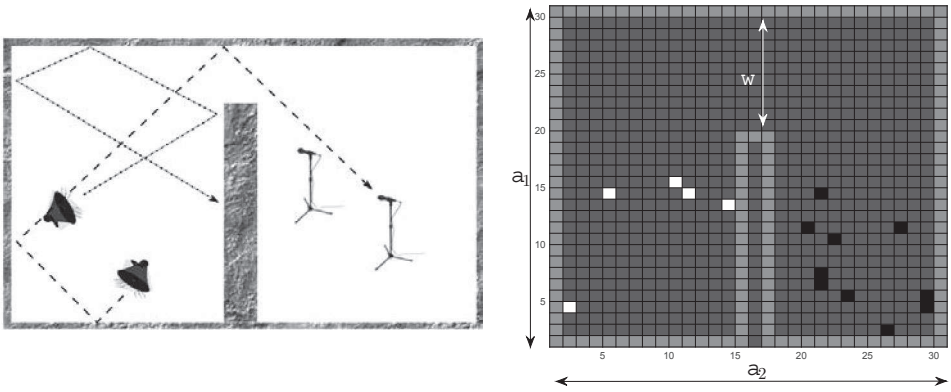


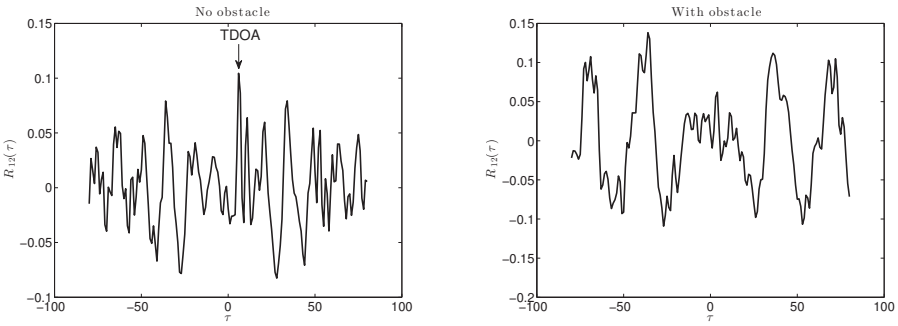
Figure 5.7 – Prototype "split room" in 2D (left) and the discretized example (right). *White pixels*: sources, *black pixels*: sensors, *light gray*: "walls", *dark gray*: propagation medium.

5.3 Hearing Behind Walls

Equipped with algorithmic tools that can help us learn some physical parameters characterizing the environment, we turn our attention to potential applications. It was already demonstrated, in chapter 4, that the physics-driven approach is a viable candidate for sound source localization, particularly in reverberant spaces. On the other hand, its high computational cost compared to some traditional localization methods may still seem unjust. In this section, we will demonstrate an application in which the traditional methods necessarily fail, but the physics-driven localization is possible. Namely, we are interested in the localization problem where the sources and microphones are separated by a soundproof obstacle, as illustrated in figure 5.7 (notice that "the wall" does not completely divide the room). We term the localization problems of this type "*hearing behind walls*" problems[5]. These problems are reminiscent of *through-the-wall radar imaging* problems (see [9] and the references therein). The difference is that the latter are based on active techniques (*i.e.* sending the probe signal through a semi-penetrable wall to the target), whereas we operate in passive mode and instead aim at localizing uncontrolled active targets (sound sources).

If the spatial domain Γ includes an obstacle between the microphones and sources, the problem is insolvable by traditional goniometric methods based on TDOA estimation (subsection 4.2.1 in the previous chapter). The problem is that, for TDOA estimation, *the direct propagation path is assumed.* The estimation usually involves computing the cross-correlations between the recorded signals, and then using this information for computing the positions of the sources. For the spatial domain proposed here, however, cross-correlation between microphones is not informative, which can be seen on Figure 5.8 (the highest peaks on the right graph correspond to the reflections). The same can be said for the beamforming methods discussed in subsection 4.2.2, which, we recall, localize sources by "browsing" the spatial

[5] *"Hearing around walls"* may be more accurate.



(a) Without a wall.

(b) With a wall (split room).

Figure 5.8 – The cross-correlations of the impulse responses in a reverberant 2D "room".

domain for the locations of highest radiated energy. Unfortunately, from the point of observer (microphone positions), most energy actually comes from the small "gap" between the two separated areas in the figure 5.7.

However, the physics-driven approach does not impose any explicit restrictions on the shape of the spatial domain, as long as we can parametrize it by the analysis operator or the synthesis dictionary. In fact, the problem was already tackled in [87] and approached by the synthesis regularization, but we abandon this approach due to scalability and computational issues, as discussed in previous chapters. Moreover, in the previous two sections we developed methods, with the analysis regularization in mind, that can be used to relax the assumptions on the knowledge of the physical domain. Therefore, we again discretize Γ to generate the analysis operator $\mathbf{A}$, which is now assumed known beforehand. Moreover, for demonstration purpose, we consider the noiseless case. On the other hand, we do not explicitly enforce constraints on the initial and boundary conditions. Instead, we leave everything to the objective, hoping that the regularizer is strong enough to identify these homogeneous conditions. Thus, the optimization problem is now formulated as a simple linearly constrained minimization ($\|\mathbf{Mp}-\mathbf{y}\|_2 = \varepsilon = 0$):

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} f_r(\mathbf{Ap}) \quad \text{s.t.} \quad \mathbf{y} = \mathbf{Mp}. \tag{5.9}$$

The problem is solved by our standard tool, the Weighted SDMM algorithm.

5.3.1 Simulations

Experimental setup in 2D Experiments in two dimensions have been conducted in a simulated "split room" environment ($\mathsf{a}_1 = 30$) × ($\mathsf{a}_2 = 30$) presented on Figure 5.7. The number of sensors is set constant to $\mathsf{m} = 10$ and they have been randomly distributed in the right bottom quarter of the room. The acquisition time is set to $\mathsf{t} = 400$ and the source emitting duration is set to $\mathsf{t}_\mathsf{e} = 10$. Increasing the emission time is not an issue, as demonstrated in the experiments in subsection 4.4.1, as long as the appropriate regularizer is used. In the following

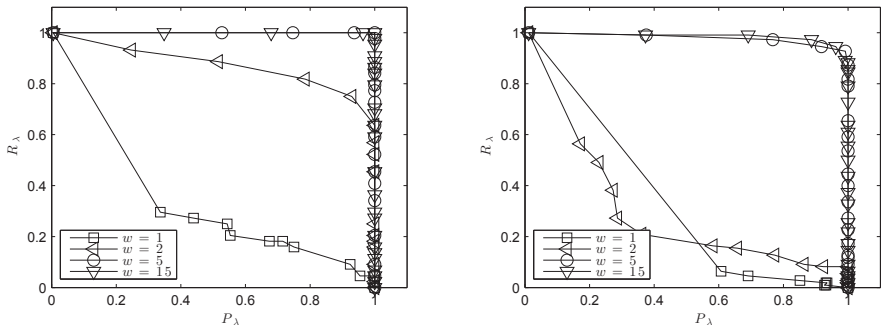


Figure 5.9 – Precision/recall diagrams for k = 4 (left) and k = 10 (right) sources.

experiments, we actually use the ℓ_1 norm instead of more appropriate structured norms, but we expect good performance since the acquisition time is considerably long.

For each experiment, we place $1 \le k \le m$ wideband sources (modeled as white Gaussian noise emitters with the amplitude distribution $\mathcal{N}(0,1)$) randomly in the left bottom quarter of the room. Then, for a given free space distance w between the obstacle and the opposite wall (from w = 1 - only one pixel wide, to w = 28 - no obstacle), we compute the ground truth signal $\mathbf{p}$ using the standard FDTD-SLF explicit scheme. Finally, the numerical solution $\hat{\mathbf{p}}$ of (5.9) is computed using the Weighted SDMM algorithm and the experiment is repeated 20 times.

As discussed in subsection 4.3.3, the most likely locations (i,j) are the ones having the highest temporal ℓ_2 norm $\hat{z}_{i,j} = \sqrt{\sum_{\tilde{t}=1}^{t} \left(\hat{\mathbf{z}}_{i,j}^{\tilde{t}}\right)^2}$. If the number of sources in not known in advance, the detection of source locations is done by applying some threshold λ. In this case, standard *precision* P_λ and *recall* R_λ measures are used to evaluate the localization performance. If we term the number of correctly identified sources by $\bar{k}(\lambda)$ and the total number of identified sources by $\hat{k}(\lambda)$, these values are equal to $P_\lambda = \bar{k}(\lambda)/\hat{k}(\lambda)$ and $R_\lambda = \bar{k}(\lambda)/k$. In addition, we compute the empirical probability of accurate source localization given the total number of sources: $P_k = \bar{k}/k$. Here $\bar{k}$ represents the number of correctly identified sources from the set of locations (i,j) obtained by keeping k highest in magnitude sums $\hat{\mathbf{Z}}_{i,j}$. For measuring the localization performance, we still maintain the *total accuracy* principle: to compute $\bar{k}(\lambda)$ and $\bar{k}$ we classify as correctly identified only those locations (i,j) which **exactly** correspond to the ground truth position of the sources. In addition, we evaluate the wavefield signal-to-noise ratio $\mathrm{SNR}_{\mathbf{p}} = 20\log_{10} \|\mathbf{p}\|_2 / \|\mathbf{p} - \hat{\mathbf{p}}\|_2$.

Results in 2D Figure 5.9 shows precision and recall graphs for the cases of k = 4 and k = 10 sources in space, and different widths w. The presented results indicate that already a small width (w = 5) is sufficient to localize the sources with high accuracy, even when their number is high. Figure 5.10 (left) presents the empirical probability P_k for varying k and w parameters. We can see that the localization probability is high, even in those cases where the door width is considerably small. As expected, the performance is lower for higher number of sources

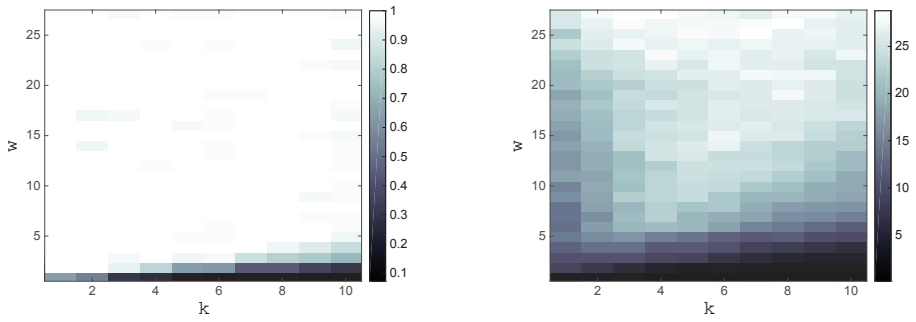


Figure 5.10 – Probability of accurate source localization given k (left) and wavefield SNR (right).

(*i.e.* lower cosparsity) and smaller door width. Figure 5.10 (right) depicts the estimated $\mathrm{SNR}_{\mathbf{p}}$, for the same range of k and w. It seems that these results are correlated with the source localization probability, although there are some surprises, namely the fact that $\mathrm{SNR}_{\mathbf{p}}$ is not the highest for the signals having the highest cosparsity (left side of the SNR graph).

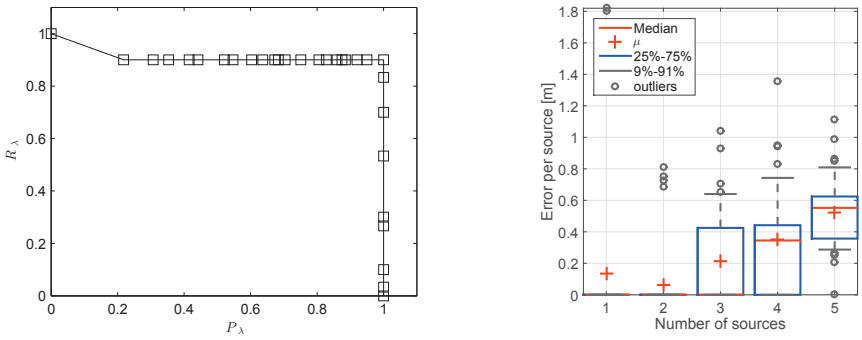
The obtained results are in accordance with physics of propagation. The well-known Huygens-Fresnel principle[6] suggests that there is a minimal door width $\tilde{w}$ beyond which it will be impossible to detect the sources in the other half of the room: it will always appear as if they are located at the gap position. This is exactly what happens for very small values of w in our experiments.

Experiments in 3D As an additional outlook to the scaling capabilities we also conduct two illustrative experiments in three dimensions.

In the first experiment, we concentrate on a single setup (fixing $k = 3$ and $w = 10$) in a simulated space of size $(a_1 = 20) \times (a_2 = 20) \times (a_3 = 20)$ with duration $t = 400$, whose results were obtained by averaging the outcome of 10 consecutive experiments. Figure 5.11 (left) is the precision/recall graph for this three-dimensional setup. For conveniently chosen range of thresholds, it was possible to accurately localize the sources in 9 out of 10 experiments. The computational time per experiment was approximately 2 to 3 times higher than needed for the 2D experiments presented before.

In the second, more physically relevant experiment, sound propagation is simulated in a virtual 3D space of size $2.5 \times 2.5 \times 2.5\mathrm{m}^3$, with a separating wall of length 1.5m. The recording time is set to $\tau = 5$s and the white noise sources emit during the entire acquisition period. The spatial domain step size is $0.25 \times 0.25 \times 0.25\mathrm{m}^3$ and the sampling frequency is $f_s \approx 2.38$kHz. In this case we assumed that the number of sources is known and the experiment is repeated 50 times. The median results presented in figure 5.11 (right), indicate that we are able to perfectly localize up to three sources.

[6] Fresnel extended Huygens's ideas on wave propagation (section 3.2.4) to include interference principles and diffraction effects.



(a) Precision/recall for the "unitless" setting.

(b) Localization error in the "physical" setting.

Figure 5.11 – Localization performance for the hearing behind walls 3D problems.

5.4 Summary and contributions

In this chapter we addressed some major obstacles on our path to apply physics-driven sound localization in practice. Additionally, we discussed a localization scenario, generally unsolvable by traditional methods, that can be addressed by our approach.

By assuming spatiotemporal smoothness of the speed of sound function, we argued that, in some cases, one may simultaneously perform source localization, signal recovery and speed of sound estimation. Concerning the blind estimation of specific acoustic impedance, we have shown that, as long as the piecewise constant model approximately holds, this parameter can be also estimated in parallel to localization and recovery. To the best of the author's knowledge, there have been no attempts so far to perform blind estimation of these physical parameters, by exploiting spatial sparsity of sound sources. Furthermore, we feel that it is possible to combine these data models in a joint optimization problem, which would lead to simultaneous estimation of all parameters and source locations.

Concerning the "hearing behind walls" scenario, we have shown that the absence of direct propagation path does not render source localization impossible. Indeed, the physics-driven approach does not depend on convexity of the spatial domain, nor on the existence of direct propagation path. In an idealized case, the method is able to accurately localize sources even if the gap connecting the two enclosures is relatively small. With respect to other works based on the same idea, we are the first to actually demonstrate the effectiveness of spatial sparsity regularization in three spatial dimensions (thanks to sparse analysis regularization) and without strong prior assumption on the source frequency band.

Now that we are convinced in the potential of our approach for various real-world acoustic scenarios, we are also interested in its generalization capability to other physics-driven inverse problems. Thus, in the forthcoming chapter we shift away from sound and consider a different source localization problem: localization of cortical sources responsible for epileptic seizures.

6 (Co)sparse brain source localization

ElectroEncephaloGraphy (EEG) represents a group of methods used for diagnosing physiological abnormalities of the brain. It is based on recordings of brain electrical activity, either in invasive (*i.e. intracranial EEG* [154]) or non-invasive manner (*i.e. scalp EEG*, or only EEG). It is assumed that the brain electric activity of the human fetus starts between the 17th and 23rd week of pregnancy and continues throughout entire life. These signals can be related to electrical potentials measured by a sensor array placed at the surface of the head (in the case of non-invasive EEG), or surface of the brain (in the case of intracranial EEG). The EEG measurements have been used for variety of clinical applications: monitoring alertness, coma and brain death, locating damaged brain areas, controlling anesthesia depth, testing drugs for convulsive effects, diagnosing sleep disorders and investigation of epilepsy and seizure origin localization (the reader is referred to [235] and the references therein).

Nowadays, EEG is most often exploited for diagnosing and examination of epilepsy. The "roots" of this irregular brain activity lie in sudden abnormal electric bursts, called paroxysmal discharges [19]. If an epileptic patient is not responding to drug treatment, a surgical removal of the epileptic region is an option. Naturally, a prerequisite is that the sources of epileptic activity are localized by means of EEG, which is the problem formulated as follows:

Given an array of m *electrodes, placed on the surface of the head (with a known geometry), determine locations of* k *epileptic sources in the brain cortex (the outer layer of the brain).*

This challenging inverse problem has been investigated by many researches and approached from different angles. As the reader might expect, in this chapter we will address the problem using the physics-driven (co)sparse framework.

The first section briefly introduces the physical model of brain electrical activity in the EEG context. In the second section, we discuss some methods in brain source localization. The third section is dedicated to proposed physics-driven localization, whose performance is exercised through simulations, in the fourth section. In the fifth section we summarize the chapter and note our contributions. For the greater part, the material in this part of the thesis is based on publications [7, 139].

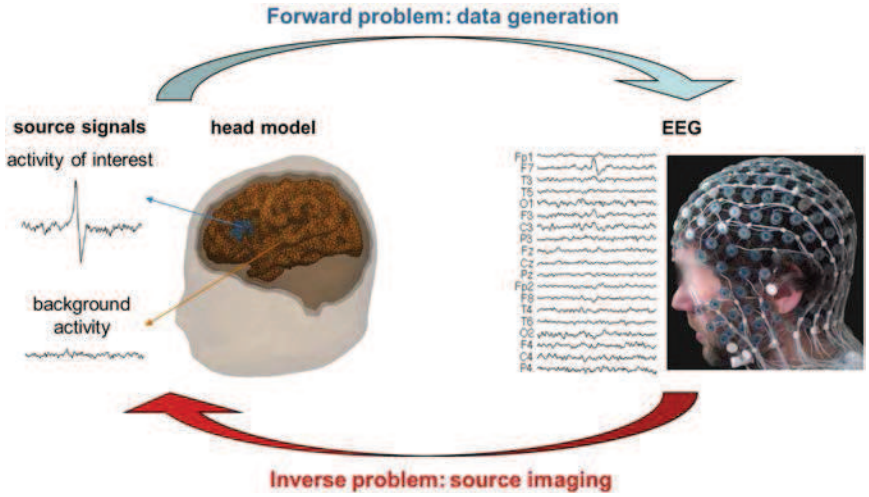


Figure 6.1 – The forward and inverse EEG problem [1].

6.1 EEG basics

The forward and inverse EEG problems are illustrated in figure 6.1. The forward problem consists of determining the measured electric potentials on the head surface given the distribution of "brain sources", whose origin we elaborate in the following text.

The brain tissue is composed of nerve cells or *neurons*, depicted in figure 6.2 (left). The simplified structure comprises the base (or *soma*), equipped with short, branched projections called *dendrites*, and the long projection called *axon* (or nerve fibre) ending with *synapses*. Neurons "communicate" by exchanging specific chemicals, known as *neurotransmitters*, which are released by synapses and absorbed by dendrites [19, 217]. For a neuron to release the neurotransmitter substance, it is necessary that so-called action potential exists, which causes an electric current to flow through axon. When a sufficiently large group of nerve cells "fires" simultaneously, the cumulative activity can be detected by measuring electric potentials on the scalp. It is presumed that this type of activity mostly originates from so-called pyramidal cells, which are oriented perpendicular to the cortical surface [122]. Therefore, the "brain sources" represent synchronous activity of a group of pyramidal cells.

Since the cells within each group are approximately parallel, their somas and synapses form groups of current sources and sinks often modeled as *current dipoles*. The dipole model, consisting of the pair of monopoles with the same magnitudes and opposite signs, is often applied in EEG signal processing. However, it is not the only one: the blurred dipole [45], the quadrupole [122], irrotational current density source model [209] etc are also used.

In order to be able to interpret the EEG measurements, we need to establish a mathematical model relating sources to scalp potentials, which is the subject of the following subsection.

[1] Image by courtesy of [19].

6.1.1 Physical model

The head is considered to be a volume conductor with a nonuniform anisotropic conductivity functional $\sigma(\mathbf{r})$ and a magnetic permeability[2] of free space $\mu = \mu_0$. Electromagnetic fields are governed by the well-known Maxwell's equations (to be consistent with usual physics notation, here we use the vector symbol $\vec{\cdot}$ to denote a physical vector field):

$$\nabla \cdot \vec{\mathbf{E}} = \frac{\rho}{\epsilon_0} \tag{6.1}$$

$$\nabla \times \vec{\mathbf{E}} = -\frac{\partial \vec{\mathbf{B}}}{\partial \mathrm{t}} \tag{6.2}$$

$$\nabla \cdot \vec{\mathbf{B}} = 0 \tag{6.3}$$

$$\nabla \times \vec{\mathbf{B}} = \mu_0 \left(\vec{\mathbf{J}} + \epsilon_0 \frac{\partial \vec{\mathbf{E}}}{\partial \mathrm{t}} \right), \tag{6.4}$$

where "×" denotes the *curl* operator[3]. The symbols $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ denote *electric* and *magnetic fields*, respectively, ρ is *electric charge density* and $\vec{\mathbf{J}}$ is *current density*. Informally, electric and magnetic fields are vector fields defined with regards to electric, respectively magnetic, force that would be induced on a test particle of unit charge. Electric charge density measures the amount of electric charge per unit volume, while current density measures the electric current per cross section of the volume. Precise definition of all these concepts can be found in any relevant textbook on electromagnetism (*e.g.* [193], in the context of EEG).

In addition, *charge conservation*, or continuity equation of electromagnetism, states that electric charge can neither be created nor destroyed, *i.e.* the net quantity of positive and negative charges is the same. Mathematically:

$$\frac{\partial \rho}{\partial \mathrm{t}} + \nabla \cdot \vec{\mathbf{J}} = 0. \tag{6.5}$$

Fortunately, when brain sources are considered, one may adopt a simplified *quasi-static approximation* of the previous laws. Indeed, usual frequencies in neuromagnetism, *i.e.* electromagnetic phenomena related to brain activity, are well below 100Hz [265, 136]. This means that, if we assume a relatively high sampling rate, the temporal derivatives in the previous expressions almost vanish: $\nabla \times \vec{\mathbf{E}} \approx 0$, $\nabla \times \vec{\mathbf{B}} \approx \mu_0 \vec{\mathbf{J}}$ and $\nabla \cdot \vec{\mathbf{J}} \approx 0$.

Another common approximation is *Ohm's law*, stating that current density is directly proportional to electric field. In terms of EEG, a convenient modification of Ohm's law [193, 248, 265] is to represent total current density as a sum of so-called *impressed* or *primary* current density

[2] Conductivity and permeability are intrinsic electromagnetic properties of the medium.

[3] Curl is the infinitesimal rotation of a 3D vector field $\vec{\mathbf{F}}$. It is implicitly defined by means of a certain line integral, however the most common definition is given with regards to Cartesian coordinate system: $\nabla \times \vec{\mathbf{F}} = \left(\frac{\partial F_z}{\partial \mathrm{y}} - \frac{\partial F_y}{\partial z} \right) \mathbf{e}_\mathrm{x} + \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial \mathrm{x}} \right) \mathbf{e}_\mathrm{y} + \left(\frac{\partial F_y}{\partial \mathrm{y}} - \frac{\partial F_y}{\partial \mathrm{x}} \right) \mathbf{e}_\mathrm{z}$, where $\mathbf{e}_\mathrm{x}$, $\mathbf{e}_\mathrm{y}$ and $\mathbf{e}_\mathrm{z}$ are the unit coordinate vectors.

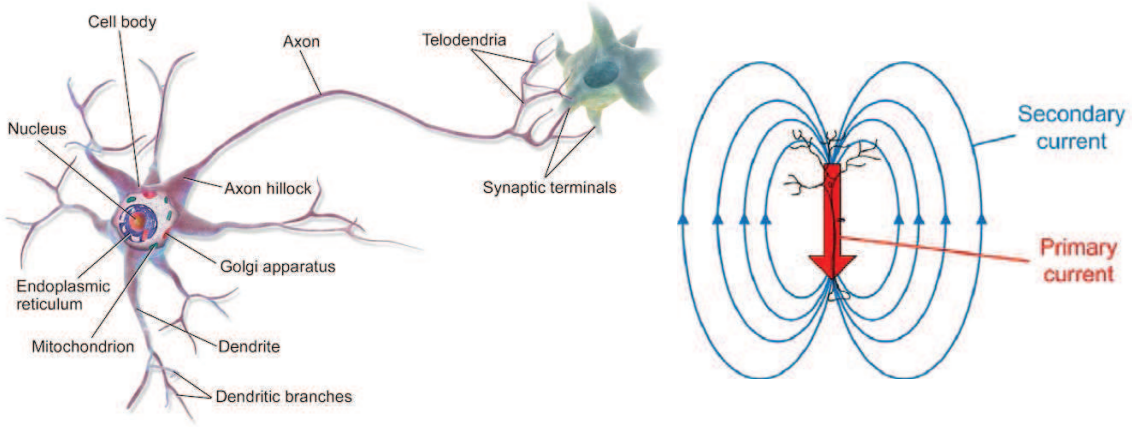


Figure 6.2 – Structure of the multipolar neuron (left) and its electric currents (right) [4].

$\vec{\mathbf{J}}_p = \sigma\vec{\mathbf{E}}$, and *return* or *secondary* current density $\vec{\mathbf{J}}_s$, shown in figure 6.2 (right):

$$\vec{\mathbf{J}} = \sigma\vec{\mathbf{E}} + \vec{\mathbf{J}}_s, \tag{6.6}$$

where σ, as noted before, represents conductivity.

Substituting (6.6) into the quasi-static charge conservation law (6.5), we have $\nabla\cdot\left(\sigma\vec{\mathbf{E}}+\vec{\mathbf{J}}_s\right)=0$. Finally, due to vanishing curl (6.2), the electric field $\vec{\mathbf{E}}$ can be represented by the gradient of some scalar potential $\vec{\mathbf{E}} = -\nabla u$. This provides us with the final relation between current density and the induced electric potentials:

$$\nabla\cdot(\sigma\nabla u) = \nabla\cdot\vec{\mathbf{J}}_s. \tag{6.7}$$

Recalling the dipole model of brain sources, this equation can be translated into [122]:

$$\nabla\cdot(\sigma\nabla u) = \sum_{j=1}^{k} z(\mathbf{r}_j)(\delta(\mathbf{r}-\mathbf{r}_j^-) - \delta(\mathbf{r}-\mathbf{r}_j^+))/\mathrm{d} := z, \tag{6.8}$$

where z_j is the magnitude of the j^{th} brain source, d is the distance between two composing monopoles (*i.e.* somas and synapses) and $\mathbf{r}_j^+$ and $\mathbf{r}_j^-$ are their respective locations.

We have actually recovered the familiar Poisson equation (3.11), this time applied to electric potentials and dipole sources. From subsection 3.2.3, we know that when appropriate boundary conditions are met, the problem is well-posed, and can be represented compactly as $Au = z$ (with A being the system composed of (6.8) and the boundary conditions). Then, the unique integral form (3.6) exists, and expresses surface potentials as a linear function of current densities: $u = Dz$. The linear operator D is the Green's functions basis. The measurements $\mathbf{y} = Mu$ are obtained by sampling the potential field u at a finite number of points on the head surface, by appropriate arrangement of measuring electrodes (*e.g.* [196]). Therefore,

[4] Images by courtesy of https://en.wikipedia.org/wiki/Neuron and [19].

the measurement operator $MD\colon z \mapsto \mathbf{y}$ (in EEG terminology, often called *lead field*) is also linear.

6.1.2 Boundary conditions

To define a well-posed forward model dictated by a linear PDE, we need to impose certain boundary conditions. In EEG, two types of boundary conditions are defined, depending on the considered spatial position [122]:

- Boundary between two compartments in the head is modeled by (non-homogeneous) Neumann or Dirichlet boundary condition, *e.g.*:

  $$\vec{\mathbf{J}}_{s1} \cdot \vec{\mathbf{n}} = \vec{\mathbf{J}}_{s2} \cdot \vec{\mathbf{n}},$$

  where $\vec{\mathbf{J}}_{s1}$ and $\vec{\mathbf{J}}_{s2}$ represent current densities within compartments 1 and 2.

- Surface boundary (head-to-air boundary) is modeled by homogeneous Neumann condition:

  $$\vec{\mathbf{J}}_{s} \cdot \vec{\mathbf{n}} = 0,$$

  which is more relevant for us, since it (straightforwardly) fulfills the self-adjointness criterion discussed in subsection 3.2.3.

When all boundaries are modeled by Neumann boundary condition, one needs to impose additional condition to ensure uniqueness (recall that all-Neumann condition leads to an additive constant ambiguity). To remedy this non-uniqueness, a reference electrode is chosen (equivalent to imposing Dirichlet boundary condition at one point of the model), or the potentials u are assumed to have zero mean [248].

6.1.3 Epilepsy

About one percent of the entire population is affected by epilepsy, making it one of the most common neuronal diseases, second only to stroke [219]. As mentioned, it arises due to paroxysmal discharges, that can occur either in the whole cortex (generalized epilepsy), or they can be localized in the limited brain regions called epileptogenic zones. In the latter case, use of EEG and MEG (*MagnetoEncephaloGraphy*) source localization can help identify epileptogenic zones, which may be then surgically removed. The electroencephalogram during seizures contains prominent *interictal spikes*, known to be related to epileptic sources [193, 219]. Epileptic source localization is especially challenging due to *background activity*, which can be seen as "biological noise" - these are the non-epileptic brain sources (*e.g.* due to muscle activation) and non-cortical activity (*e.g.* cardiac). In addition, the measurement (instrumental) noise is always present.

6.2 Brain source localization methods

Given the well-posed forward model, we turn our attention to the EEG localization, which is an inverse source problem. Localization methods can be applied to a single "snapshot" of observation data (measurements collected at one time instant), or to a temporal block of measurements (collected during several consecutive time instances). For snapshot methods, one usually considers time instants corresponding to interictal spikes, as this data usually has highest SNR with respect to the background activity [19].

However, even in the unrealistic setting where all surface potentials are known, it has been proven that the EEG inverse source problem (without additional constraints) is ill-posed [73] (namely, non-unique). Therefore, some form of regularization is always required. Based on applied hypothesis and techniques, localization methods in EEG can be roughly divided in two groups [118], briefly discussed in the rest of this section.

6.2.1 Parametric methods

Parametric models generally assume that only few dipoles (brain sources), with unknown locations and orientations, are responsible for the observed surface potentials. This form of spatial sparsity is known as *equivalent dipole model*. Sources are defined by six-dimensional vectors $\mathbf{p}_j$ consisting of three spatial coordinates $\mathbf{r}_j$ (center of the j^{th} dipole), two angles indicating dipole orientation $\boldsymbol{\theta}_j$ and the magnitude z_j. These parameters appear non-linearly in (6.8), hence the optimization problems in this category are nonlinear and non-convex.

Non-linear least squares There are several methods that directly attempt to minimize $\|\mathbf{y} - MD\hat{z}(\mathbf{p}_1, \mathbf{p}_2 \dots \mathbf{p}_k)\|_2^2$, where $\hat{z}$ is the (parametrized) estimate of the right hand side of (6.8). Since the problem is non-linear, methods based on Gauss-Newton [221], Nelder-Mead [69], simulated annealing [175] and genetic algorithms [194] have been proposed. When used in a "snapshot" measurement regime, these methods are known as "moving dipole models", since the dipole positions are unconstrained, whereas in the block measurement case the dipole positions can be fixed over the entire interval [15].

Besides uncertainty in the quality of an estimate, due to non-convexity, an obvious downside of these methods is that the number of dipoles k is required a priori. This is difficult to estimate automatically, and, in practice, analysts parametrize the problem with different k until the result is physiologically plausible. However, increasing k makes the problem more ill-posed, ultimately leading to non-uniqueness. Additionally, it increases the search space and therefore, the computational cost (which can be significant for simulated annealing and similar approaches).

Beamforming In the context of EEG, the output of standard beamformer can be seen as a linear spatial filter applied to the electrode measurements $\mathbf{Y} \in \mathbb{R}^{m \times t}$ at time $\tilde{t}$:

$$\mathbf{w}^T \mathbf{y}_{\tilde{t}} = \mathbf{w}^T \left(MDz_{\tilde{t}} + \mathbf{e} \right), \tag{6.9}$$

where $\mathbf{y}_{\tilde{t}}$ is the $\tilde{t}^{\text{th}}$ column of the matrix $\mathbf{Y}$, $z_{\tilde{t}}$ are the brain current densities at time $\tilde{t}$ and $\mathbf{e}$ is the additive noise.

Same as in the acoustic case, the goal is to browse the search space for the high-energy locations. Therefore, the filter weights $\mathbf{w}^\top \in \mathbb{R}^{1\times m}$, parametrized by the target vector $\mathbf{r}$, should suppress all brain sources but the one at the position $\mathbf{r}$. In practice, only a finite number of locations $\mathbf{r}$ is searched, which is equivalent to a discrete source grid $\mathbf{z}_{\tilde{t}} \in \mathbb{R}^s$. Moreover, to compute filter weights, the lead field operator MD is required - usually, it is numerically computed and stored as in a matrix form $\mathbf{MD} \in \mathbb{R}^{m\times s}$. The discrete version $\mathbf{M} \in \mathbb{R}^{m\times s}$ of the subsampling operator M, is simply a row-reduced identity matrix, while the Green's functions basis D is discretized into a dictionary $\mathbf{D} \in \mathbb{R}^{s\times s}$.

Usually, some minimization criterion is chosen for the filter weights, *e.g.* the ℓ_2 norm. Beamformers designed this way are data-independent. An alternative is to exploit the statistics of the recorded data, to build an adaptive beamformer. One well-known example is the *Linearly Constrained Minimum Variance (LCMV)* beamformer, where the idea is to use the minimum variance estimate, constrained by the unit response at location $\mathbf{r}$ [118, 244]:

$$\underset{\mathbf{w}}{\text{minimize}}\ \mathbf{w}^\top \mathbb{E}(\mathbf{y}_{\tilde{t}}\mathbf{y}_{\tilde{t}}^\top)\mathbf{w} \quad \text{s. t.} \quad \mathbf{w}^\top \mathbf{M}\mathbf{d}_{\mathbf{r}} = 1, \tag{6.10}$$

where $\mathbf{d}_{\mathbf{r}}$ is the column of $\mathbf{D}$ coinciding with the unit dipole source at $\mathbf{r}$ (*i.e.* the corresponding Green's function). The covariance matrix $\mathbb{E}(\mathbf{y}_{\tilde{t}}\mathbf{y}_{\tilde{t}}^\top)$ is usually estimated from the data, and possibly badly conditioned. Solution of (6.10) requires inverting this matrix, which is done by applying Tikhonov regularization and/or using iterative methods. The major drawbacks of LCMV beamformer are sensitivity to noise and correlation between targeted and interfering sources, which may cause mutual cancellation [72].

MUSIC We already mentioned *MUltiple SIgnal Classification (MUSIC)* [223] algorithm in chapter 4, subsection 4.2.1, in the acoustic context. In contrast to its ineffectiveness in indoor sound source localization, it is widely used in EEG source localization. In a way, MUSIC can also be seen as a beamforming approach, but with the preprocessing step that is used to identify signal and noise subspaces. First, the measurement matrix $\mathbf{Y}$ is factorized by the Singular Value Decomposition (SVD), *i.e.* $\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ and p largest singular values are identified (ideally, $p = k$, but the number of sources is unknown beforehand). The corresponding p left singular vectors $\mathbf{U}_s = \mathbf{U}_{1:p}$ represent the signal subspace, whereas the remaining ones are considered to form the noise subspace.

In the second step, scanning is performed, by comparing the signal subspace with the columns of the lead field matrix $\mathbf{MD}$ (more precisely, the Green's functions for all dipole positions $\mathbf{r}$ and orientations $\boldsymbol{\theta}$). The goal is to find global maximums of *subspace correlation*, *i.e.*:

$$\texttt{subcorr}\,(\mathbf{d}(\mathbf{r},\boldsymbol{\theta}), \mathbf{U}_s) = \frac{\|\mathbf{U}_s^\top \mathbf{d}(\mathbf{r},\boldsymbol{\theta})\|_2}{\|\mathbf{d}(\mathbf{r},\boldsymbol{\theta})\|_2}, \tag{6.11}$$

which requires exhaustive search over (discrete) parameter space $(\mathbf{r},\boldsymbol{\theta})$. To reduce computa-

tional cost, *quasi-linear* approximation is commonly undertaken - the dipole's Greens function is assumed to be separable: $\mathbf{d}(\mathbf{r},\boldsymbol{\theta}) = \mathbf{d}(\mathbf{r})\boldsymbol{\theta}$. Then, search can be done sequentially, by maximizing subspace correlation for the position $\mathbf{r}$ only, keeping p locations that exhibited highest subspace correation, and then improving these values by tuning their angle vectors $\boldsymbol{\theta}$.

Problems arise if the signal subspace is inaccurately estimated: then, only the first location can be easily identified during the search. In that case local maxima may compromise the search for the remaining $\mathsf{p}-1$ global estimators, and some sort of peak-picking routine may be used. An alternative is the *Recursively Applied and Projected (RAP) MUSIC* algorithm [181], where the sources are iteratively localized, by removing their contribution before the new source is estimated (this way local maxima around the old sources is also removed).

The number of resolvable sources, for MUSIC-like algorithms, is upper-bounded by the number of microphones: $\mathsf{k} < \mathsf{m}$. Since these algorithms are based on correlations, they are inherently sensitive to synchronized sources. However, for high SNRs, they are more robust compared to the LCMV beamformer, and can tolerate partially correlated sources [15]. This means that partially correlated sources whose second moments are statistically independent can still be resolved. An improvement in this direction was proposed in [6], termed *4^{th}-order Deflation MUSIC (4-D MUSIC)*, which further relaxes the partial correlation assumption by requiring statistical independence of the fourth-order cumulants.

6.2.2 Non-parametric methods

In addition to equivalent dipole model, non-parametric methods also consider so-called distributed source models. The assumption is that dipole sources with fixed locations and possibly fixed orientations are distributed in the whole brain volume or cortical surface. Thus, only the amplitudes and directions need to be estimated. Furthermore, it is usually not necessary to consider the entire head volume (grid), since it is widely believed that sources of primary interest are restricted to the cortex. This constraint is usually inferred from Magnetic Resonance Imaging (MRI) data, where image segmentation techniques are used to determine the cortical surface of interest.

Bayesian framework If the (discrete) source signal is represented in spatiotemporal matrix form $\mathbf{Z} \in \mathbb{R}^{\mathsf{s}\times\mathsf{t}}$, where each column is the dipole distribution at time instant $\tilde{\mathsf{t}}$, the measurement data can be expressed as:

$$\mathbf{Y} = \mathbf{MDZ} + \mathbf{E}, \tag{6.12}$$

where $\mathbf{E}$ is the additive noise matrix.

In the Bayesian framework, the variables $\mathbf{Y}$, $\mathbf{Z}$ and $\mathbf{E}$ are considered to be random variables, denoted here by Y, Z and E. The goal is to yield a posterior point estimate $\hat{\mathbf{Z}}$, usually in a form of *Maximum A Posteriori (MAP)* or *conditional mean*. The latter, even though argued as more accurate [16, 135], is usually difficult to compute, due to necessity of evaluating high-

dimensional posterior integrals. Indeed, for general distributions, evaluating expectations cannot be done by analytically. Therefore, use of techniques that approximate expectations, such as Markov Chain Monte Carlo method, is often unavoidable.

MAP is defined as the source signal that maximizes the mode of a posterior distribution:

$$\hat{\mathbf{Z}} = \arg\max_{\mathbf{Z}} p(\mathrm{Y} = \mathbf{Y} \mid \mathrm{Z} = \mathbf{Z})\, p(\mathrm{Z} = \mathbf{Z}) = \arg\min_{\mathbf{Z}} -\ln p(\mathrm{Y} = \mathbf{Y} \mid \mathrm{Z} = \mathbf{Z}) - \ln p(\mathrm{Z} = \mathbf{Z}), \quad (6.13)$$

where $p(\mathrm{Y} = \mathbf{Y} \mid \mathrm{Z} = \mathbf{Z})$ is the conditional *probability density function (pdf)* related to probability of observing the measurements $\mathbf{Y}$ given the dipole arrangement $\mathbf{Z}$, while $p(\mathrm{Z} = \mathbf{Z})$ is the evaluated *prior* pdf of the source random variable Z.

A common choice for the log likelihood term $\ln p(\mathrm{Y} = \mathbf{Y} \mid \mathrm{Z} = \mathbf{Z})$ is given by assuming Gaussian distribution of the noise E:

$$p(\mathrm{Y} = \mathbf{Y} \mid \mathrm{Z} = \mathbf{Z}) \propto \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{Y} - \mathbf{MDZ}\|_2^2\right), \quad (6.14)$$

which leads to the MAP estimator:

$$\hat{\mathbf{Z}} = \arg\min_{\mathbf{Z}} \frac{1}{2\sigma^2}\|\mathbf{Y} - \mathbf{MDZ}\|_2^2 - \ln p(\mathrm{Z} = \mathbf{Z}). \quad (6.15)$$

When the prior pdf is also Gaussian, the MAP estimate is equivalent to the minimum mean square error estimate, or Wiener solution [135]. In addition to Gaussian prior pdf, many other prior distributions have been proposed, such as the ones generated from hierarchical Bayesian framework / Markov random field models [113, 209].

Penalized least squares This group of methods has been introduced, in broader context, in subsection 1.2.1 on *Variational regularization*. The general problem is to find an estimate of the source term $\hat{\mathbf{Z}}$ by solving the following problem:

$$\underset{\mathbf{Z}}{\text{minimize}}\ \|\mathbf{Y} - \mathbf{MDZ}\|_2^2 + \lambda f_r(\mathbf{Z}), \quad (6.16)$$

where f_r is, as usual, a variational prior functional, and λ is a user-defined scalar parameter.

As mentioned earlier, in some cases the problems generated by the Bayesian and variational frameworks are identical: *e.g.* MAP estimation (6.15) with Gaussian prior coincides with Tikhonov regularization ($f_r = \|\mathbf{LZ}\|_2^2$), for $\mathbf{L} = \mathbf{I}$ and an appropriate choice of λ. However, interpretation of one framework through another may be misleading, as pointed out in [119].

In EEG, specific Tikhonov regularization methods are known as *Minimum Norm Estimates (MNE)* [19]. Particularly, chooosing $\mathbf{L} = \mathbf{F}^{[2]}\mathtt{diag}(\mathbf{w})$, where $\mathbf{F}^{[2]}$ is the discrete Laplacian and $\mathbf{w} > \mathbf{0}$, leads to popular *LOw Resolution Electrical Tomography Algorithm (LORETA)* [203] and its variants [202, 249]. As expected, MNE regularization is known to produce only low-resolution solutions [15], which is why different priors have been considered. *Minimum Current Estimate*

(MCE) [241], on the other hand, is based on $f_r = \|\mathtt{diag}(\mathbf{w})\,\mathbf{Z}\|_1$, *i.e.* it encourages sparsity of the weighted source term. Note that, in the approaches where the dipole model is assumed, a particular structure is imposed on the optimization variable $\mathbf{Z}$ (otherwise, $\mathbf{Z}$ corresponds to monopole sources). Some approaches assume additional regularity, *e.g.* a spatiotemporal structured sparsity, promoted by the joint $\ell_{2,1}$ norm [81].

The aforementioned sparsity-promoting penalized least squares are based on the sparse synthesis data model, due to the explicit use of the lead field matrix. The reader may presume that all these approaches fit into our physics-driven regularization framework. We recall that this is true only when an analogous sparse analysis approach can be derived, which is not always the case. In the following section we will apply the physics-driven (co)sparse framework of chapter 3, to develop a brain source localization method, and compare it to the synthesis-based one (*i.e.* an MCE variant) that does not have an exact cosparse analog.

Perhaps unsurprisingly, the literature on the sparse analysis regularization in EEG is extremely scarce. The author is aware of only few references, *e.g.* [178, 170, 164], which are, however, very recent and not directly related to our work.

6.3 Physics-driven (co)sparse brain source localization

One of our goals is to demonstrate that the physics-driven framework, along with the sparse analysis regularization, can be employed in the EEG context. As a proof of concept, we will apply our prescribed procedure to construct a simple "snapshot" signal estimator.

First, let us state the main assumptions:

(A1) Potentials u are recorded by electrodes placed at locations $\mathbf{r}_\mathsf{i}$, with $\mathsf{i} = [1, \mathsf{m}]$.

(A2) Out of the total number of point sources q, only $\mathsf{k} < \mathsf{m}$ are active.

(A3) Sources are located in the cortex Σ and their orientation is known.

(A4) Each current dipole is represented by two monopoles with opposite amplitudes.

The relation between potentials and dipole sources is given by the equation (6.8), with boundary conditions noted in subsection 6.1.2. The reference electrode's potential is set to 0V. Thus, we have satisfied the three requirements for the physics-driven approach: existence of the well-posed physical model, measurement system and sparse singularities. Now, the continuous domain model needs to be discretized to yield the usual system $\mathbf{Au} = \mathbf{z}$.

6.3.1 Discretization and the dipole model embedding

For the left hand side of (6.8) we use a finite difference method proposed by Witwer et al. [257], described in appendix B.2. This method implicitly assumes all-Neumann homogeneous

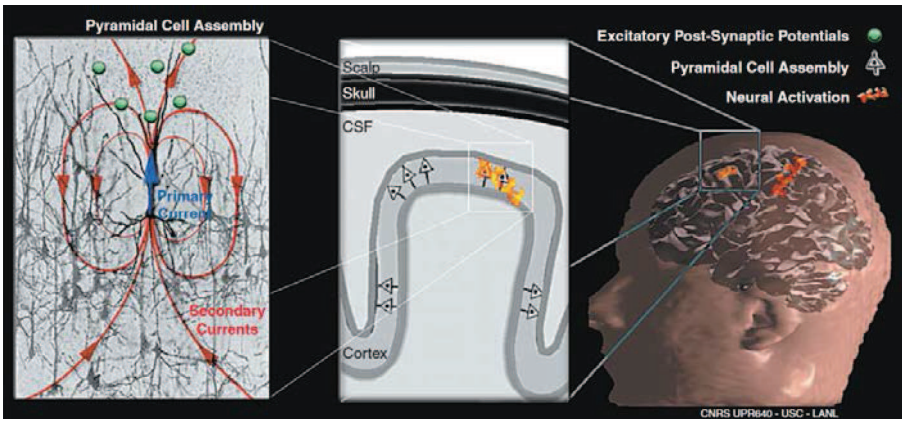


Figure 6.3 – The dipole model: physiological origin (left), orientation (center) and location (right).[5]

boundary conditions, thus the difference system it generates is ill-posed. Fortunately, given the zero reference electrode, the system is easy to stabilize: one simply has to remove the row and column corresponding to the reference electrode's position in discretized space Γ. This is legal because the current density at this location is also equal to zero (by the assumption (A3)). Since the problem setting is time-independent, we have $n := |\Gamma| - 1 = s - 1$ (due to variable removal), and the generated analysis operator is $\mathbf{A} \in \mathbb{R}^{n \times n}$. Furthermore, $\mathbf{Z} := \mathbf{z} \in \mathbb{R}^n$, $\mathbf{u} \in \mathbb{R}^n$, $\mathbf{M} \in \mathbb{R}^{n \times m}$ and $\mathbf{Y} := \mathbf{y} \in \mathbb{R}^m$.

To model the right hand side of (6.8), we impose certain structure onto the vector $\mathbf{z} = \mathbf{Au}$. If the dipole model is assumed, we can think of $\mathbf{z}$ as composed of q pairs[6] of components with equal magnitude and opposite sign, and the remaining zeros. Moreover, the non-zeros of $\mathbf{z}$ are restricted by the support set Σ, and their orientations, *i.e.* positions of the positive and negative monopoles, are known (illustrated in figure 6.3). This knowledge can be encoded through a sparse matrix $\mathbf{B}$. Let u^- and u^+ index the elements of $\mathbf{z}$ giving the amplitude of the monopoles located at positions $\mathbf{r}^-$ and $\mathbf{r}^+$, respectively. Knowing the geometry of the cortex, it is possible to build a set of q different couples (u_j^-, u_j^+) indexing the q dipoles of Σ. Indeed, by covering the surface of the gray matter with q monopoles indexed by the integers u_j^-, we can deduce the q corresponding integers u_j^+ such that each dipole is oriented orthogonally to the surface of Σ, as the pyramidal cells it represents. We can thus express the vector $\mathbf{z}$ as $\mathbf{z} = \mathbf{B}\tilde{\mathbf{z}}$, where $\mathbf{B} = (b_{u_1,u_2})$ is an $(n \times n)$ sparse matrix with entries

$$b_{u_1,u_2} = \begin{cases} 1 & \text{if } u_1 = u_2 \\ -1 & \text{if } u_1 = u_j^+ \text{ and } u_2 = u_j^- \\ 0 & \text{otherwise,} \end{cases} . \tag{6.17}$$

and $\tilde{\mathbf{z}}$ is an n-dimensional k-sparse vector with $n - q$ known zero elements (the u^{th} element of $\tilde{\mathbf{z}}$ is zero if $u \neq u_j^-$ for $j \in \{1, \ldots, q\}$). Non-zero elements of $\tilde{\mathbf{z}}$ represent the amplitude of monopoles

[5] Image by courtesy of [5].
[6] Due to finite resolution of the discrete grid, some pairs might share their elements.

restricted to the cortical surface.

To summarize, the model

$$\mathbf{Au} = \mathbf{z} = \mathbf{B}\tilde{\mathbf{z}} \tag{6.18}$$

can be rewritten as

$$\tilde{\mathbf{A}}\mathbf{u} = \tilde{\mathbf{z}}, \tag{6.19}$$

with $\tilde{\mathbf{A}} := \mathbf{B}^{-1}\mathbf{A}$.

By construction, $\mathbf{B}$ is invertible, and the cost of computing some matrix-vector product $\mathbf{B}^{-1}\mathbf{v}$ is $O(n)$ if one employs the forward substitution algorithm [111]. Indeed, it can be shown that there is a permutation of rows and columns of $\mathbf{B}$ that yields a lower triangular matrix with ones on the main diagonal. Since the sparse structure of a matrix is preserved by permutations, the forward substitution has only linear complexity. In practice, we observed that $\mathbf{B}^{-1}$ is also sparse, as well as the matrix $\tilde{\mathbf{A}}$ which is the product of the two sparse matrices.

6.3.2 The analysis *vs* synthesis regularization

As in the acoustic context, the Green's functions dictionary $\mathbf{D}$ can be computed as the inverse of the analysis operator $\mathbf{A}$. It is clear from previous discussions, that the equivalent representation of (6.20) is given by the synthesis model:

$$\mathbf{u} = \mathbf{DB}\tilde{\mathbf{z}} = \tilde{\mathbf{D}}\tilde{\mathbf{z}}, \tag{6.20}$$

where $\mathbf{D} = \mathbf{A}^{-1}$ and $\tilde{\mathbf{D}} = \tilde{\mathbf{A}}^{-1}$. The two models are equivalent, but, unlike the matrix $\mathbf{A}$, the dictionary $\mathbf{D}$ is *not* sparse, and neither is the matrix $\tilde{\mathbf{D}}$. The physical interpretation is that Poisson's equation models a non-zero steady-state electrical field on the domain given a source distribution. This can be mathematically verified for the spherical head model where the analytical solution of the forward problem is available [6].

Support constraints Let Γ_1 be the set of possible locations of monopoles at the surface of the gray matter Σ. The rows of the analysis operator $\tilde{\mathbf{A}}$ can then be accordingly split into $\tilde{\mathbf{A}}_{\Gamma_1}$ and $\tilde{\mathbf{A}}_{\Gamma_2}$, where $\tilde{\mathbf{A}}_{\Gamma_1}$ is the $(q \times n)$ submatrix of $\tilde{\mathbf{A}}$ obtained by extracting the rows corresponding to the support set Γ_1, and $\tilde{\mathbf{A}}_{\Gamma_2}$ corresponds to the rows indexed by the complementary set Γ_2.

Convex relaxation Signal estimation is performed by solving either the analysis sparse

$$\underset{\mathbf{u}}{\text{minimize}}\ f_r(\tilde{\mathbf{A}}\mathbf{u}) + f_d(\mathbf{Mu} - \mathbf{y}) + f_c(\tilde{\mathbf{A}}_{\Gamma_2}\mathbf{u}), \tag{6.21}$$

or, the synthesis sparse problem

$$\underset{\tilde{\mathbf{z}}}{\text{minimize}}\ f_r(\tilde{\mathbf{z}}) + f_d(\mathbf{M}\tilde{\mathbf{D}}\tilde{\mathbf{z}} - \mathbf{y}) + f_c(\tilde{\mathbf{z}}_{\Gamma_2}), \tag{6.22}$$

from which the dipole estimate $\hat{\mathbf{z}} = \mathbf{B}\tilde{\mathbf{z}} = \mathbf{A}\mathbf{u}$ is easily deduced. The reader may observe that the boundary term is again excluded, but the homogeneous boundary conditions are promoted through the objective functional f_r. Particularly, we used standard convex relaxation by the ℓ_1 norm, since the sources are assumed sparse and the dipole structure is already embedded into the modified analysis operator/synthesis dictionary. As a measure of data-fidelity, we again use the characteristic function $\chi_{\ell_2 \leq \varepsilon}(\cdot)$, while for the f_c penalty, we used either the characteristic function $\chi_{\ell_2 \leq \sigma}(\cdot)$, with $\sigma = \varepsilon$, or the weighted sum-of-squares term $\lambda \|\cdot\|_2^2$.

Localization The brain sources can then be detected either by thresholding, or by keeping k highest in magnitude dipoles $\hat{\mathbf{z}} = \mathbf{B}\tilde{\mathbf{z}}$, given the estimate of $\tilde{\mathbf{z}}$.

Minimum Current Estimate Another way of encoding knowledge about the source support set Γ_1 is to factorize $\tilde{\mathbf{z}} = \mathbf{R}\mathbf{s}$, where $\mathbf{R} = (r_{u_1,u_2})$ is an $(n \times q)$ expanding sparse matrix defined by:

$$r_{u_1,u_2} = \begin{cases} 1 & \text{if } u_1 = u_j^- \text{ and } u_2 = j, \\ 0 & \text{otherwise.} \end{cases} \tag{6.23}$$

for $j \in \{1, \ldots, q\}$.

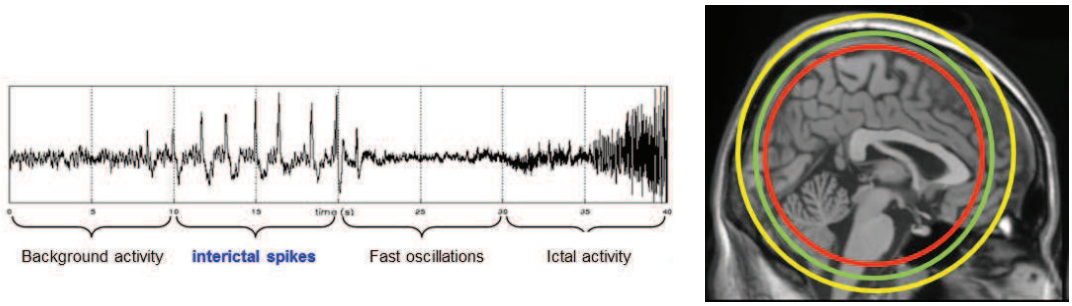
Using the expanding matrix $\mathbf{R}$ (6.23), a second synthesis sparse problem can be solved in order to localize brain sources. It consists in solving (3.22) with the dictionary $\bar{\mathbf{D}} = \mathbf{A}^{-1}\mathbf{R}$. In section 6.4, the corresponding optimization algorithm will be named Minimum Current Estimate (MCE) to refer to a similar approach proposed in [241]. Note that, in this case, an equivalent cosparse optimization problem cannot be readily designed, since the matrix $\bar{\mathbf{D}}$ is not invertible.

6.4 Simulations

The experiments conducted here are not as exhaustive as the ones performed in the sound source localization chapter. Instead, we targeted computational complexity issue of the sparse *vs* cosparse approaches, recognized also in the acoustic case. Nevertheless, we included the MCE algorithm in all simulations. Moreover, in the experiments investigating the robustness, we compare the physics-driven (co)sparse localization against several other state-of-the-art methods.

6.4.1 Settings

In all experiments, $k = 3$ distant epileptic dipoles are placed in the gray matter of the superior cortex. A physiologically-relevant stochastic model [130] is used to generate the spike-like interictal epileptic activity. It is noteworthy that this activity is the same for the three epileptic dipoles, leading to synchronous epileptic sources. On the other hand, the background activity, *i.e.*, the activity of non-epileptic dipoles of the gray matter, is generated as Gaussian and as



(a) The example time series of a single EEG sensor electrode.

(b) The spherical head model.

Figure 6.4 – An example measurements and the head model.[7]

temporally and spatially white. Its power is controlled by a multiplicative coefficient in order to get different Signal to Noise Ratio (SNR) values. Recall that we decided to work in "snapshot" regime, as explained in section 6.3, where the selected electrode measurement matches the time sample where the highest signal amplitude is observed, corresponding to the top of the interictal spike (illustrated in figure 6.4a). The electrodes were placed on the scalp sphere using the $10-5$ system [196]. For the head model, we used three nested concentric spheres (figure 6.4b) with radius (cm) equal to 7, 8, 9.2, and piecewise constant conductivities (siemens/cm) equal to 1, 0.0667, 1, 10^{-10}.

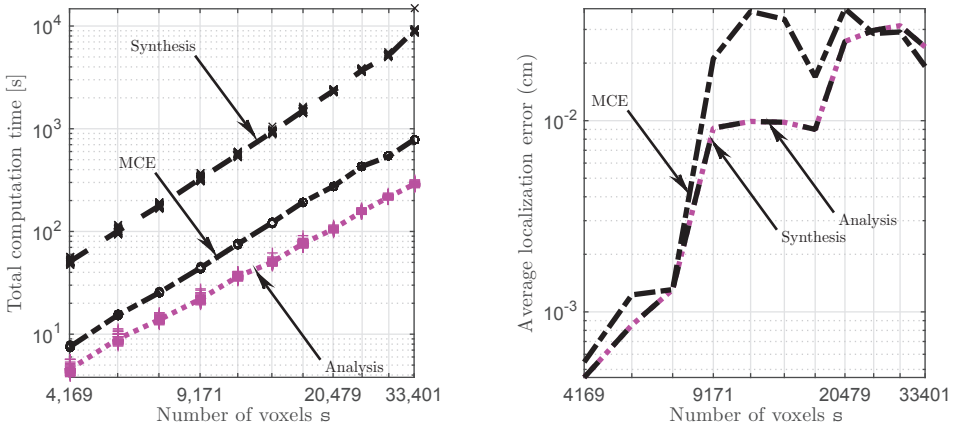
Concerning the sparsity- and cosparsity-based approaches, we used our standard convex optimization tool, the Weighted SDMM algorithm. The stopping criterion is parametrized by the relative accuracy ($\mu = 0.01$) and the iteration count (10^4). The data-fidelity parameter ε set to 0, *i.e.* enforcing the equality constraint (thus. we make no assumptions on the noise variance). Due to the moderate scale of the EEG problem, we are able to use a direct solver (Cholesky factorization) for the evaluation of the least squares minimizer in (3.28).

The quality of localization is presented in the form of RMSE $\|\hat{\mathbf{r}}-\bar{\mathbf{r}}\|_2$, per source (with $\hat{\mathbf{r}}$ and $\bar{\mathbf{r}}$ the estimated and true source location, respectively). We did 50 realizations of each experiment, all run on Intel® Xeon® 4-Core 2.8GHz, equipped with 32GB RAM.

6.4.2 Scalability

Analogously to the acoustic case in subsection 4.4.4, we first investigate how the analysis and synthesis regularizations compare in terms of scalability. The number of measurement electrodes is kept fixed at $\mathsf{m} = 128$. We vary the discretization at 11 different scales, yielding uniform grids with a number of nodes ranging between $\mathsf{s} = 4169$ and $\mathsf{s} = 33401$. We recall that the number of optimization variables is equal to $\mathsf{s}-1$ for the synthesis/analysis approaches, and equal to $\mathsf{q} \ll \mathsf{s}$ for the MCE-like approach. Therefore, the results are presented with respect to the number s of voxels in the head, which is the same for all methods. In these experiments,

[7] Images by courtesy of [5].



(a) Total computational time *vs* size s.

(b) RMSE *vs* size s.

Figure 6.5 – Computational and performance effects with regards to the problem size.

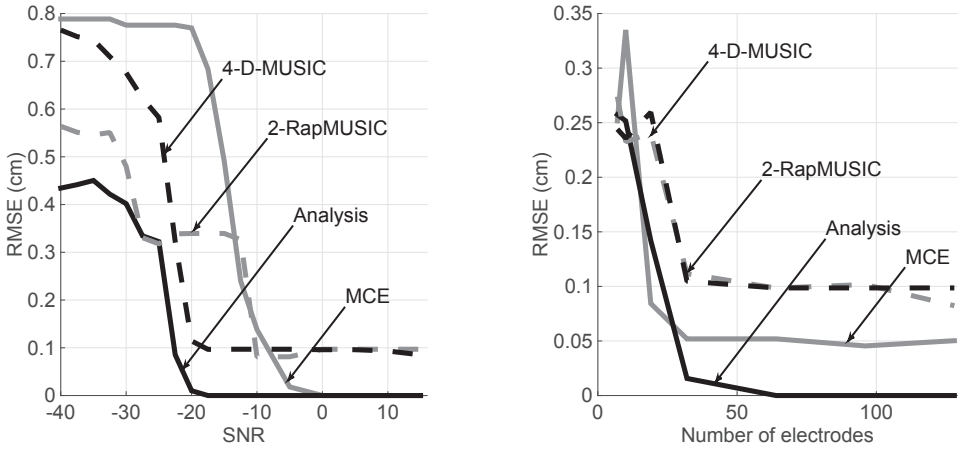
the source locations are chosen randomly for each realization, and the non-cortex support was suppressed by the linear constraints f_c.

First, as for the acoustic case, the computational cost of the synthesis approach is higher and grows faster than that of if its analysis counterpart (figure 6.5a). Interestingly, the latter and the MCE-like approach show a similar behavior. In fact, the direct computation of the proximal operators and the SDMM involved in the MCE-like approach require a slightly lower cost, but the additional cost due to the initial computation of the dictionary slightly increases the MCE-like technique's total cost ; the impact of this additional cost is not negligible, since the dictionary has to be recomputed for each patient in clinical practice.

Second, the localization error increases with spatial resolution (figure 6.5b), which is expected due to the fixed number of electrodes. Furthermore, it can be confirmed that the analysis and synthesis models are also equivalent in the EEG context. Since the analysis approach has better scaling capabilities, we no longer investigate the sparse synthesis regularization given in (6.22). However, the MCE approach does not have a sparse analysis analog, thus we still keep it the forthcoming series.

6.4.3 Robustness

A common type of model error encountered in the biomedical context is due to the presence of background activity in non-epileptic regions of the gray matter. Non-epileptic dipoles of the gray matter also have a non-zero amplitude, even if it should be ideally lower than that of epileptic dipoles. Consequently, the n-dimensional vector **Au** is not really k-sparse, but it should have k dominant components. We conduct two experiments to evaluate the influence of such model errors, in which we included state-of-the-art approaches RapMUSIC [181] and 4-D-MUSIC [6] (discussed in subsection 6.2.1). Three synchronous epileptic dipoles were



(a) RMSE *vs* SNR.

(b) RMSE *vs* m, for SNR=−6dB.

Figure 6.6 – Robustness to (background activity) noise.

placed at fixed positions, but the background noise is regenerated for each realization. For these experiments, the weighted square ℓ_2 norm has been used in place of the f_c penalty in (6.21).

Figure 6.6a presents the simulation results as a function of the SNR value, with the same number of sensors (m = 128), as in the previous series. This is representative of localization problems with synchronous and focal epileptic sources, which are challenging for existing techniques. The analysis approach exhibits a remarkable robustness with respect to background activity and manages to perfectly localize epileptic sources even for a very low SNR value. This is, however, not the case for other approaches, which either do not succeed in achieving the same performance (4-D-MUSIC, RapMUSIC), or require significantly higher SNR (MCE).

Figure 6.6b displays the RMSE criterion at the output of the four algorithms as a function of the number m of electrodes for an SNR value of −6dB. The physics-driven approach requires at least 64 electrodes to achieve a perfect result for such an SNR value while the other algorithms do not manage to perfectly localize the three synchronous epileptic dipoles regardless of the available number of electrodes. Given all these results, along with the results from the previous subsection, among the considered algorithms, the analysis approach is a clear winner in terms of both localization performance and computational cost.

6.5 Summary and contributions

In this chapter, we applied the physics-driven framework to the ill-posed inverse problem of epileptic source localization in electroencephalography.

After discussing the physical model governed by Poisson's equation and dipole sources, we briefly presented some state-of-the-art localization methods in EEG. We argued that some of these approaches, based on the sparse synthesis regularization, can fit into our physics-driven (co)sparse context. However, some of them (*e.g.* MCE), cannot be interpreted through this framework, even though they use the lead field matrix, *i.e.* the Green's functions.

We proceeded by developing a physics-driven (co)sparse source localization method, that exploits the dipole structure and the cortical spatial constraints and can be naturally formulated in the sparse analysis and sparse synthesis context. The experiments on simulated, but physically-relevant data, have demonstrated that the two versions perform equally in terms of localization accuracy. However, as expected, the analysis version is computationally more efficient, due the inherited sparsity of the analysis operator matrix. Moreover, the localization method is more robust with respect to background (non-epileptic) brain source activity than several state-of-the-art algorithms in the field, including the MCE approach.

Experiments on the real data with realistic head models are required to confirm the usefulness of the approach for clinical applications. Envisioned improvements are going beyond the "snapshot" regime and incorporating structured (co)sparsity priors, in order to exploit spatiotemporal correlation of the epileptic sources. Additionally, one could reduce the physical prior information and learn some parameters (*e.g.* conductivities) from the data, in similar fashion as it was done in the acoustic case.

7 Conclusions and perspectives

The central theme of this work was the interplay between cosparse and sparse regularizations with an emphasis on physics-driven linear inverse problems. We discussed three applications: audio declipping (desaturation), acoustic source localization and mapping of epileptic sources in EEG imaging. In the remaining text we will note the principal contributions and propose several directions for future research on the subject.

7.1 Conclusions

Chapter 2 was the prologue of a main subject. It concerned the audio declipping inverse problem, addressed by sparse analysis and sparse synthesis regularizations. To avoid biasing the results using different algorithms, but at the same time insisting on competitive declipping performance, we developed a versatile non-convex approach that can straightforwardly accommodate either the analysis or the synthesis prior. Moreover, when the analysis operator forms a tight frame, the analysis-based algorithm has a very low computational complexity per iteration, equivalent to the cost of two matrix-vector products. The experimental results indicate that, while the synthesis version is slightly advantageous in terms of audio recovery, the analysis one is significantly more efficient in terms of processing time. Moreover, both methods perform highly competitive against state-of-the-art declipping algorithms, especially when the saturation is severe (confirmed numerically and by MUSHRA perceptual evaluation). This work has lead to industrial collaboration with a leading professional audio restoration company.

In Chapter 3 we introduced the physical problems of our interest and proposed a general regularization framework termed the *physics-driven (co)sparse regularization*. This class of problems is governed by physical laws expressed by linear partial differential equations or, equivalently, in integral form through the Green's functions basis. Analogously, the sparse analysis regularization is based on discretization of the coupled system, formed by the partial differential equation and the appropriate initial/boundary conditions. The sparse synthesis regularization is based on discretization of the Green's functions, or the system's impulse

response, leading to a nominally equivalent representation. Nevertheless, we argued that the inherited sparsity of the analysis operator yields computationally much more scalable optimization problems, provided that the employed discretization is locally supported. Additionally, we developed a version of Alternating Direction Method of Multiplier algorithm, termed Weighted SDMM, tailored for solving large scale convex optimization problems generated by the two regularization approaches, and used throughout our work.

Chapter 4 was devoted to a particular physics-driven inverse problem: sound source localization in reverberant rooms. After discussing current issues with state-of-the-art methods in this context, we showed how the physics-driven (co)sparse framework can be applied to the localization problem. We assumed (approximate) knowledge of the room geometry and structure, and suggested *full wavefield interpolation* technique, by exploiting the inhomogeneous wave equation in the time domain. This way, making strong assumptions on the source frequency band is avoided, and a signal estimate for every spatiotemporal coordinate of the discretized domain is obtained (as a "byproduct"). The computational complexity of the analysis- and synthesis-based regularization was discussed, and later on empirically validated. It was argued that, for problems of this scale, only the analysis-based regularization is a viable choice. Indeed, to the best of our knowledge, this is the first time that full (inverse) wavefield interpolation was performed on a regular laptop, in a physically relevant three-dimensional simulation setting. Additionally, we investigated how the performance is affected by various model parameterizations, which confirmed our intuition that, for this technique, the reverberation is a welcome phenomenon that actually improves localization performance. Finally, physics-driven cosparse localization was confronted to a stochastic version of state-of-the-art SRP-PHAT algorithm, which it outperformed in a given experimental setting.

In chapter 5 we continued the discussion on physics-driven cosparse acoustic localization, and investigated the possibility of relaxing the physical prior knowledge. First, we explore the scenario with unknown, but smoothly varying speed of sound in a room. The smoothness assumption enables significant reduction of degrees of freedom, which was exploited to design a biconvex optimization algorithm that blindly estimates the speed, and the pressure signal at the same time (thereby, performing source localization). Obtained results are promising, although the current implementation exploits only the spatial smoothness of sound speed. We then moved to the problem of blind estimation of specific acoustic impedance, *i.e.* the absorption parameters of the boundaries. The physically justified piecewise constant model was assumed, and an ADMM-based, biconvex algorithm was developed. The empirical results show that the algorithm performs almost identically as physics-driven localization in the accurate setting (where the boundary parameters are perfectly known). Moreover, it exhibits robustness to moderate model error and additive noise. This part is concluded by demonstrating the ability of physics-driven (co)sparse localization to detect sources in a scenario where a direct propagation path does not exist, and where TDOA-based and classical beamforming methods necessarily fail. By exploiting only echoes, the method shows remarkable performance even when the obstructing wall is quite large relative to the room size (both in two- and three-dimensional setting).

A last physics-driven inverse problem was addressed in chapter 6 - epileptic brain source localization in electroencephalography. We argued that some state-of-the-art methods in this field fit into our framework, and proceeded by developing a new one, by exploiting Poisson's equation and the dipole source model. The numerical efficiency of the sparse analysis compared to the sparse synthesis approach was again verified. Numerical simulations, with physiologically-relevant source signal generator, were used for comparison of our method against several state-of-the-art algorithms outside the physics-driven regularization framework. It was demonstrated that the proposed approach requires the least number of electrodes for accurate localization, and is the most robust with respect to biological noise (non-epileptic brain activity).

7.2 Perspectives

We envision several possibilities for future research, both theoretical and practical, for each of the problems previously discussed.

7.2.1 Declipping

Theoretical aspects

There are no results to date on the theory behind a declipping inverse problem. Although declipping algorithms seem to work in practice, theoretical understanding of the problem may help us estimate performance bounds achievable by these methods. Recent studies on generalized RIP conditions [37] could be a good direction.

Deriving a convergence proof for the SPADE algorithms is equally important. Recent theoretical advances on non-convex proximal algorithms and ADMM [34, 127] may be a good starting point. We also predict existing connections with Dykstra's projection algorithm [41].

Practical aspects

From the experience with social sparsity algorithm, we expect that applying structured (co)sparsity priors could substantially improve perceptual quality of a declipped estimate. Moreover, one can imagine incorporating psychoacoustic information, to further improve audible performance.

In our work, we considered only the single channel setting. Parallel declipping of a multichannel recording is a possibility for extending the approach.

Finally, more difficult problems could be investigated, such as simultaneous declipping and source separation, or declipping a signal that was convolved by an acoustic path filter.

7.2.2 Physics-driven cosparse acoustic localization

Theoretical aspects

We already know that the inverse source problem is generally ill-posed, even with spatial (co)sparsity regularization. Difficulty of the problem is illustrated by the existence of the non-radiating source phenomenon [79, 78]. Nevertheless, our problem is even more complex, due to the fact that i) the acoustic pressure field is heavily subsampled and, ii) the discretization alters the underlying physical model.

Related to the latter issue, we predict connections between our work and the so-called *sparse spikes deconvolution* problem [89]. In that work, the authors argue that there are fundamental performance limits for discrete regularization of a genuinely continuous domain problem (whatever is the "resolution" of a grid). A challenging research axis is constructing a method that avoids discretization all together, in the spirit of *super-resolution* [48].

Concerning BLESS (section 5.1) and CALAIS (section 5.2), we suspect that adapting the theory of non-convex ADMM, presented in [127], could lead to convergence proofs for these two algorithms. An alternative to biconvex optimization may be a *lifting scheme*, which has been successfully exploited in, *e.g.* phase retrieval problems [17, 47, 162, 28].

Practical aspects

Even though the optimization problems generated by physics-driven cosparse regularization scale gracefully, non-smooth convex optimization involving tens of millions (and easily, much more) variables is a challenging task, certainly not well-suited for practical implementation. We envision several possibilities:

Multilevel approach Instead of performing, in a way "brute force" convex optimization, one can envision a hierarchy of models at different scales, *e.g.* controlled by the finesse of the discretization grid. Apart from dimensionality reduction, benefit would also come from the improved conditioning of the linear problems at coarse grids. One way to exploit multilevel idea is to accelerate the Weighted SDMM algorithm, by employing algebraic multigrid [101, 218] as a solver or preconditioner for the normal equations (3.28). Another possibility is to directly apply some of the recently proposed multilevel convex optimization algorithms, *e.g.* [240, 200]. One can also think of multilevel source localization, where the cosupport is iteratively refined: first, estimation at a lower scale (coarser grid) would be performed, and then used as a cosupport constraint for estimation at a higher scale (finer grid). The cosupport constraints could be interpreted as additional observations, hence there is a possibility of acceleration noted in subsection 4.4.4 and illustrated in Figure 4.13. With regards to this time-data tradeoff, a technique called *aggressive smoothing* [42] could be used as well.

Helmholtz domain Moving into the frequency domain is tempting, as it would immediately lead to significant dimension reduction. However, there is a compromise: as mentioned, assumptions need to be made on the frequency range where sources emit. Perhaps more worrying is the result from [58], mentioned in subsection 4.3.1, demonstrating the sensitivity of the sparse synthesis method [87] with regards to modal frequencies. On the other hand, the approach taken in [87] is based on the OMP algorithm, thus convex regularization might be more stable. Lastly, due to oscillatory behavior, the discretized Helmholtz equation becomes increasingly difficult to solve at large wavenumbers [98], which may influence the stability of the inverse problem.

Implicit discretizations The CFL condition is upper bounding the temporal stepsize in explicit discretization schemes (appendix B.1). Implicit discretizations, however, are unconditionally stable, and thus allow for arbitrary large stepsizes. The caveat is that the stepsize is inversely proportional to the discretization accuracy, hence larger steps lead to larger errors. A way to partially remedy this problem is to use higher order schemes, which may, in turn, increase the computational complexity (fortunately, only by a constant).

Robustness to inevitable modeling error should be further investigated. We observed a decrease in localization performance for the non-inverse crime scenario, attributed to spatial aliasing - thus, localization criteria probably needs to be adjusted. There is a possibility that the robustness may be improved by exploiting temporal source models, such as time-frequency (co)sparsity used for audio declipping in the second chapter.

Concerning the physical parameter estimation, we foresee using different parameter models, depending on the environment - or even user-defined (*e.g.* piecewise smooth instead of spatiotemporally smooth sound speed, to account for the interfaces between different propagation media). As a long term objective, we envision learning the geometry from data, possibly using more flexible discretizations, such as FEM. The ultimate goal is to develop a fully-blind algorithm, *i.e.* one that simultaneously estimates the geometry, parameters and pressure field (thus, source locations), using only the measurements and weak data models (such as the ones exploited for sound speed and acoustic impedance learning).

7.2.3 Physics-driven cosparse brain source localization

Theoretical aspects

The lack of recovery guarantees makes the approach, despite its empirical success, somewhat less convincing. To begin, one may consider analytic solutions of Poisson's equation available for the spherical head model and isotropic conductivities.

Practical aspects

Incorporating realistic head models (*i.e.* FEM templates inferred from MRI or *Computed Tomography (CT)* scans) and using non-simulated data is necessary. As an alternative to MRI/CT templates, a blind estimation method, similar to the ones used for parameter learning in the acoustic case, could be designed in order to estimate conductivities of different head compartments.

The "snapshot" regime which we used for simplicity does not exploit the temporal support, which may be necessary for successful regularization of high-dimensional problems generated by FEM discretization. In this case, too, refined time-frequency source models could be used.

Concerning the spatial source model, clustered dipoles are physiologically more relevant than the point dipole model we used in our work. To accomplish this task, a tailored group (co)sparse regularization should be used. Namely, we envision that applying the newly proposed *Ordered Weighted* ℓ_1 regularization [102] to this problem could exploit spatial and magnitude correlations of epileptic sources.

A Alternating Direction Method of Multipliers

Convex optimization is one of the main tools in sparse (and cosparse) signal recovery. Usually, convex programs generated by these problems are non-smooth, *i.e.* they do not admit a unique differential at every point of their domain (think of the ℓ_1 norm minimization, for instance). Moreover, the physics-driven (co)sparse regularized problems, introduced in chapter 3 and to which most of the thesis is devoted, are usually of a very large scale. These problems are difficult to solve using second order optimization algorithms, such as the interior-point method [40]. As opposed to first order algorithms, which require the functional and gradient oracle, second order algorithms additionally need the Hessian matrix (the nonsmoothness here is dealt by reformulating the original problem into a smooth, but usually higher-dimensional variant). Instead, first order algorithms, particularly the ones from *proximal splitting* framework [67, 199] are more applicable.

The aim of this appendix is not to be an exhaustive review of proximal splitting methods, for what the reader may consult the literature cited above, and the references therein. Instead, we will briefly introduce a specific proximal splitting algorithm known as *Alternating Direction Method of Multipliers (ADMM)* [90, 39], whose adapted version was introduced in section 3.4, and used throughout our work to address physics-driven (co)sparse problems. Moreover, we also used ADMM (like many other practitioners) as a heuristic non-convex optimization algorithm. Some theoretical guarantees for ADMM applied in the non-convex setting have been recently provided by different authors.

The emphasis is on the intuition and practical issues, without an in-depth discussion on the convex analysis principles behind the algorithm. In the first section, we introduce ADMM through augmented Lagrangian framework. In the second section, we discuss an ADMM variant tailored for minimization of separable objectives, which is followed by noting the proximal operator frequently used in our work, in the third section. In the last, fourth section, we discuss the ADMM algorithm as a non-convex heuristics.

A.1 Origins of ADMM

In its canonical form (adopted from [39]), ADMM can be used to solve problems in the following form:

$$\underset{\mathbf{x},\mathbf{z}}{\text{minimize}}\ f_1(\mathbf{x}) + f_2(\mathbf{z}) \quad \text{subject to} \quad \mathbf{A}\mathbf{x} - \mathbf{B}\mathbf{z} = \mathbf{h}. \tag{A.1}$$

where $\mathbf{A} \in \mathbb{R}^{p\times n}$, $\mathbf{B} \in \mathbb{R}^{p\times m}$ and $\mathbf{h} \in \mathbb{R}^{p}$ (these are generic matrices and vectors).

To address this problem, consider the augmented Lagrangian [192] relaxation:

$$L_\rho(\mathbf{x}, \mathbf{z}, \mathbf{w}) = f_1(\mathbf{x}) + f_2(\mathbf{z}) + \mathbf{w}^\mathsf{T}(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{h}) + \frac{\rho}{2}\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{h}\|_2^2, \tag{A.2}$$

where ρ is a positive constant. By variable substitution $\mathbf{u} = \frac{1}{\rho}\mathbf{w}$, an equivalent expression is called *scaled Lagrangian* form:

$$L_\rho(\mathbf{x}, \mathbf{z}, \mathbf{u}) = f_1(\mathbf{x}) + f_2(\mathbf{z}) + \frac{\rho}{2}\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{h} + \mathbf{u}\|_2^2 - \frac{\rho}{2}\|\mathbf{u}\|_2^2. \tag{A.3}$$

Then, the dual problem [40, 192] writes as

$$\underset{\mathbf{u}}{\text{maximize}} \left(\underset{\mathbf{x},\mathbf{z}}{\text{minimize}}\ L_\rho(\mathbf{x}, \mathbf{z}, \mathbf{u}) \right), \tag{A.4}$$

for which the standard *dual ascent* or *augmented Lagrangian* method [39, 91] may be applied:

$$\begin{aligned}(\mathbf{x}^{(i+1)}, \mathbf{z}^{(i+1)}) &= \underset{\mathbf{x},\mathbf{z}}{\arg\min}\ L_\rho(\mathbf{x}, \mathbf{z}, \mathbf{u}^{(i)}) \\ \mathbf{u}^{(i+1)} &= \mathbf{u}^{(i)} + \mathbf{A}\mathbf{x}^{(i+1)} + \mathbf{B}\mathbf{z}^{(i+1)} - \mathbf{h}.\end{aligned}$$

The issue with the scheme above is that the minimization in the first step has to be performed jointly over $(\mathbf{x}, \mathbf{z})$. The advantage of ADMM is that this joint optimization is decoupled into two sequential steps (independent minimization over $\mathbf{x}$ and $\mathbf{z}$):

$$\begin{aligned}\mathbf{x}^{(i+1)} &= \underset{\mathbf{x}}{\arg\min}\ L_\rho(\mathbf{x}, \mathbf{z}^{(i)}, \mathbf{u}^{(i)}) = \underset{\mathbf{x}}{\arg\min}\ f_1(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}^{(i)} - \mathbf{h} + \mathbf{u}^{(i)}\|_2^2 \\ \mathbf{z}^{(i+1)} &= \underset{\mathbf{x}}{\arg\min}\ L_\rho(\mathbf{x}^{(i+1)}, \mathbf{z}, \mathbf{u}^{(i)}) = \underset{\mathbf{z}}{\arg\min}\ f_2(\mathbf{z}) + \frac{\rho}{2}\|\mathbf{A}\mathbf{x}^{(i+1)} + \mathbf{B}\mathbf{z} - \mathbf{h} + \mathbf{u}^{(i)}\|_2^2 \\ \mathbf{u}^{(i+1)} &= \mathbf{u}^{(i)} + \mathbf{A}\mathbf{x}^{(i+1)} + \mathbf{B}\mathbf{z}^{(i+1)} - \mathbf{h}.\end{aligned} \tag{A.5}$$

Proving convergence of ADMM is not trivial, and has been erroneously interpreted as performing one pass of nonlinear Gauss-Seidel method to the joint optimization step of augmented Lagrangian [91]. However, the method *does* converge asymptotically to a stationary point of the original convex optimization problem, which can be proven through operator splitting theory and composition of so-called nonexpansive mappings [91]. More precisely, ADMM can be interpreted as *Douglas-Rachford splitting* [163] applied to the standard (non-augmented) Lagrangian dual problem of (A.1) [90].

A.2 Simultaneous Direction Method of Multipliers

When the optimization problem is the sum of more than two functionals, it can be solved by *Simultaneous Direction Method of Multipliers (SDMM)* [67]. Consider the following problem

$$\underset{\mathbf{x},\mathbf{z}_i}{\text{minimize}} \sum_{i=1}^{f} f_i(\mathbf{z}_i) \quad \text{subject to} \quad \mathbf{H}_i\mathbf{x}-\mathbf{h}_i=\mathbf{z}_i. \tag{A.6}$$

Actually, the above is easily reformulated as an ADMM problem (hence, its convergence is guaranteed through ADMM framework). Define:

$$g_1(\mathbf{z}) = \sum_{i=1}^{f} f_i(\mathbf{z}_i), \text{ where } \mathbf{z} = \begin{bmatrix} \mathbf{z}_1^T & \mathbf{z}_2^T & \dots & \mathbf{z}_f^T \end{bmatrix}^T. \tag{A.7}$$

By choosing $g_2(\mathbf{x}) = 0$, we can recover the express ADMM problem as follows:

$$\underset{\mathbf{x},\mathbf{z}}{\text{minimize}}\, g_1(\mathbf{z}) + g_2(\mathbf{x}) \quad \text{subject to} \quad \mathbf{H}\mathbf{x}-\mathbf{h}=\mathbf{z}, \tag{A.8}$$

where $\mathbf{H} = \begin{bmatrix}\mathbf{H}_1^T \mathbf{H}_2^T \dots \mathbf{H}_f^T\end{bmatrix}^T$ and $\mathbf{h} = \begin{bmatrix}\mathbf{h}_1^T \mathbf{h}_2^T \dots \mathbf{h}_f^T\end{bmatrix}^T$.

The iterates, given in (A.5), are now expressed as follows:

$$\begin{aligned}
\mathbf{z}^{(j+1)} &= \text{prox}_{\frac{1}{\rho}g_1}\left(\mathbf{H}\mathbf{x}^{(j)}-\mathbf{h}+\mathbf{u}^{(j)}\right),\\
\mathbf{x}^{(j+1)} &= \arg\min_{\mathbf{x}} \frac{\rho}{2}\|\mathbf{H}\mathbf{x}-\mathbf{h}+\mathbf{u}^{(j)}-\mathbf{z}^{(j+1)}\|_2^2,\\
\mathbf{u}^{(j+1)} &= \mathbf{u}^{(j)}+\mathbf{H}\mathbf{x}^{(j+1)}-\mathbf{h}-\mathbf{z}^{(j+1)},
\end{aligned} \tag{A.9}$$

where $\text{prox}_{\frac{1}{\rho}g_1}(\mathbf{v})$ denotes the famous *proximal operator* [179, 199, 67] of the function $\frac{1}{\rho}g_1$ applied to some vector $\mathbf{v}$:

$$\text{prox}_{\frac{1}{\rho}g_1}(\mathbf{v}) = \arg\min_{\mathbf{z}} g_1(\mathbf{z}) + \frac{\rho}{2}\|\mathbf{z}-\mathbf{v}\|_2^2. \tag{A.10}$$

Since $g_1(\mathbf{z})$ in (A.7) is block-separable, so is the proximal operator $\text{prox}_{\frac{1}{\rho}g_1}(\cdot)$ [67] (the least squares step and the $\mathbf{u}$-updates are trivially separable as well). Finally, we can recover the SDMM iterates:

$$\begin{aligned}
\mathbf{z}_i^{(j+1)} &= \text{prox}_{\frac{1}{\rho}f_i}\left(\mathbf{H}_i\mathbf{x}^{(j)}-\mathbf{h}_i+\mathbf{u}_i^{(j)}\right),\\
\mathbf{x}^{(j+1)} &= \arg\min_{\mathbf{x}} \sum_{i=1}^{f} \frac{\rho}{2}\|\mathbf{H}_i\mathbf{x}-\mathbf{h}_i+\mathbf{u}_i^{(j)}-\mathbf{z}_i^{(j+1)}\|_2^2,\\
\mathbf{u}_i^{(j+1)} &= \mathbf{u}_i^{(j)}+\mathbf{H}_i\mathbf{x}^{(j+1)}-\mathbf{h}_i-\mathbf{z}_i^{(j+1)}.
\end{aligned} \tag{A.11}$$

Note that the SDMM algorithm is straightforwardly applicable to distributed computing, due to the decoupled minimization in the $\mathbf{z}_i$ update step.

Although asymptotically convergent, in practice, the SDMM algorithm often terminates to infeasible point (*e.g.* due to an iteration threshold). Therefore, in section 3.4 we propose a modified *Weighted SDMM*, which reaches feasibility in far fewer iterations (empirically).

Stopping criterion Following [39], the SDMM stopping criterion is based on primal and dual residuals $\mathbf{r}^{(j)}_{\text{prim}}$ and $\mathbf{r}^{(j)}_{\text{dual}}$, which are defined as:

$$\begin{aligned}\mathbf{r}^{(j)}_{\text{prim}} &= \mathbf{H}\mathbf{x}^{(j)} - \mathbf{h} - \mathbf{z}^{(j)} \\ \mathbf{r}^{(j)}_{\text{dual}} &= \mathbf{H}\left(\mathbf{x}^{(j)} - \mathbf{x}^{(j-1)}\right).\end{aligned} \tag{A.12}$$

We stop iterating once their norms fall below the thresholds:

$$\begin{aligned}\mathrm{q}^{(j)}_{\text{prim}} &= \mu \min\left\{\|\mathbf{H}\mathbf{x}^{(j)} - \mathbf{h}\|_2, \|\mathbf{z}^{(j)}\|_2\right\} \\ \mathrm{q}^{(j)}_{\text{dual}} &= \mu \left\| \begin{bmatrix} \sqrt{\rho_1}\mathbf{u}^{(j)}_1 \\ \vdots \\ \sqrt{\rho_f}\mathbf{u}^{(j)}_f \end{bmatrix} \right\|_2,\end{aligned} \tag{A.13}$$

where μ denotes the relative accuracy.

A.3 Proximal operators

Efficient evaluation of proximal operators is of crucial importance for the computational efficiency of proximal algorithms. Fortunately, many interesting functionals admit computationally cheap, even explicit solutions (these are well-known results by now, available in the references). Here we state some of them frequently used throughout the thesis.

The ℓ_1 norm: $(\text{prox}_{\lambda\ell_1}(\mathbf{v}))_i = v_i\left(1 - \frac{\lambda}{|(v_i|}\right)_+$.

The joint $\ell_{2,1}$ norm: $(\text{prox}_{\lambda\ell_{2,1}}(\mathbf{v}))_i = v_i\left(1 - \frac{\lambda}{\|\mathbf{v}_\Upsilon\|_2}\right)_+$.

The hierarchical $\ell_{2,1}$ norm[1] $\text{prox}_{\frac{1}{\rho}(\ell_{2,1}+\ell_1)}(\mathbf{v}) = \text{prox}_{\frac{1}{\rho}\ell_{2,1}}\left(\text{prox}_{\frac{1}{\rho}\ell_1}(\mathbf{v})\right)$.

The operator $(\cdot)_+$ denotes component-wise positive thresholding: $(\mathbf{v})_+ := \{\forall i \mid \max(v_i, 0)\}$, and $\mathbf{v}_\Upsilon$ denotes a vector composed by the elements of a vector $\mathbf{v}$ indexed by the indice-set Υ.

Instead of constraining, one may choose to add a penalty term $\lambda\|\mathbf{v}\|_2^2$. Then, the proximal operator associated with this function is easily obtained, since it is the minimizer of a sum of quadratic forms:

$$\text{prox}_{\lambda\ell_2^2}(\mathbf{v}) = \frac{\mathbf{v}}{2\lambda + 1}. \tag{A.14}$$

[1] Denotes the hierarchical $\ell_{2,1}$ norm for joint groups of variables (similar to the standard joint $\ell_{2,1}$ norm) and singletons, together.

Proximal operator of the **characteristic function** $\chi_\Xi(\mathbf{v})$

$$\chi_\Xi(\mathbf{v}) = \begin{cases} 0, & \mathbf{v} \in \Xi, \\ +\infty & \text{otherwise}, \end{cases} \tag{A.15}$$

corresponds to the orthogonal projection $P_\Xi(\mathbf{v})$ to a set Ξ. A common case is the characteristic function $\chi_{\ell_2 \le \mu}(\cdot)$, which bounds the ℓ_2 norm of $\mathbf{v}$ by some noise level ($\|\mathbf{v}\|_2 \le \mu$). The proximal operator is the projection of the vector to the ℓ_2-ball of radius μ:

$$\text{prox}_{\ell_2 \le \mu}(\mathbf{v}) = \begin{cases} \mathbf{v}, & \text{if } \|\mathbf{v}\|_2 \le \mu \\ \mu \frac{\mathbf{v}}{\|\mathbf{v}\|_2}, & \text{otherwise}. \end{cases} \tag{A.16}$$

Projection to the ℓ_∞ ball amounts to constraining the absolute magnitude of a signal:

$$(\text{prox}_{\ell_\infty \le \mu}(\mathbf{v}))_i = \begin{cases} v_i, & |v_i| \le \mu \\ \text{sgn}(v_i)\mu, & \text{otherwise}. \end{cases} \tag{A.17}$$

A generalization of the previous functional are the component-wise magnitude constraints. The projection is similar to (A.17), except that the bound μ and the constraint sign ("$\le$" or "$=$") are specified per element of an input vector.

Another simple, but useful penalty is an **inequality constraint** $\chi_{\le \mathbf{c}}(\mathbf{v})$. The projection is straightforward:

$$(\text{prox}_{\le \mathbf{c}}(\mathbf{v}))_i = \begin{cases} v_i, & v_i \le c_i \\ c_i, & \text{otherwise}. \end{cases} \tag{A.18}$$

For the end of this short review, we left probably the most obvious case: the **affine equality constraints**. Standard way of accounting for these constraints in an SDMM program, is by introducing a new auxiliary variable $\mathbf{z}_i$ and a characteristic function of an affine set. Note, however, that the $\mathbf{x}^{(i+1)}$ update step in (A.11) is a simple unconstrained linear least squares minimization, *i.e.* its minimizer is obtained by (approximately) solving the normal equations [40]. Now, one can straightforwardly incorporate linear constraints and ensure that *every* iterate of the algorithm is feasible. This can be done by transforming the normal equations into another linear system, for instance by means of the *null-space method*, or by exploiting *Karush-Kuhn-Tucker (KKT)* conditions [192]. The former method leads to a lower-dimensional positive definite system, but requires a null space basis. In the section 3.3, we saw that the physics-driven *cosparse* regularization often yields convex problems with very simple linear equality constraints, whose null space basis is sparse and easily constructed.

Finally, we remark that there are many other useful proximal operators which may be efficiently computed, even if they do not admit closed-form solutions [67].

A.4 ADMM for non-convex problems

Many engineering applications give rise to non-convex optimization problems. In this case, global optimum is rarely theoretically guaranteed, but practitioners often obtain good results even though the computed local minimizers may be suboptimal. The ADMM algorithm has been exhaustively used as a non-convex optimization heuristics [39, 1, 253, 232, 60, 59], despite the lack of general theoretical justification in this setting.

This popularity is partially motivated by the fact that proximal operators of some useful non-convex functionals are very easy to compute. A well-known case is the ℓ_0 constraint, *i.e.* the characteristic function of the k-sparse set. The associated projection is done by applying the *hard thresholding* operator to the input vector, which preserves k highest in magnitude coefficients and sets the rest to zero. A related case is rank-k matrix constraint: the orthogonal projection corresponds to keeping k largest singular values of the associated *Singular Value Decomposition (SVD)*. Another class are *biconvex* problems [114], *i.e.* the problems which are genuinely (jointly) nonconvex, but become convex when considering only one variable (the other is kept fixed). Therefore, iterative steps in (A.5) are usually tractable in this case.

A very recent theoretical work [127] is one of the first that provided some theoretical evidence of local convergence for ADMM applied to non-convex problems. This work, however, makes two strong assumptions: first, the non-convex functionals need to be smooth (therefore, characteristic functions of non-convex sets are not included), and second, the ADMM multiplier ρ needs to be chosen sufficiently high such that the associated proximal operator is a resolvent of a strongly convex function. Although very restrictive, the theory still holds in some interesting cases, such as for certain biconvex problems involving bilinear forms.

General convergence proof for non-convex ADMMs is still an open problem, but recent advances [251] show that the research is going in good direction. Encouraged by these results, in this thesis, we developed and used several algorithms based on non-convex ADMM, although their convergence is not yet fully supported by theory.

B Discretization

B.1 Finite Difference Time Domain - Standard Leapfrog Method

For demonstration purpose, we consider Standard Leapfrog Method (SLF) discretization of the isotropic[1] acoustic wave-equation (4.7), in two dimensions (extension to 3D follows straightforwardly).

Pressure inside the *discrete* spatial domain $\Gamma \setminus \partial\Gamma$ (excluding the boundary), and corresponding to time samples $t > 2$, is discretized using the 7-point stencil in figure B.1, as follows:

$$\frac{\partial^2 p^t_{i,j}}{\partial x^2} + \frac{\partial^2 p^t_{i,j}}{\partial y^2} - \frac{1}{c^2}\frac{\partial^2 p^t_{i,j}}{\partial t^2} = \frac{p^t_{i-1,j} - 2p^t_{i,j} + p^t_{i+1,j}}{d_x^2} + \frac{p^t_{i,j-1} - 2p^t_{i,j} + p^t_{i,j+1}}{d_y^2} - \frac{1}{c^2}\frac{p^{t+1}_{i,j} - 2p^t_{i,j} + p^{t-1}_{i,j}}{d_t^2} + O(\max(d_x, d_y, d_t)^2) \quad \text{(B.1)}$$

where d_x, d_y and d_t denote the discretized spatial and temporal step sizes, respectively. Neglecting the $O(\cdot)$ term yields a convenient *explicit* scheme [231] to compute $p^{t+1}_{i,j}$ using pressure values at the previous two discrete time instances ($p^t_{(\cdot,\cdot)}$ and $p^{t-1}_{(\cdot,\cdot)}$).

For *all* nodes (including the boundary) at $t = 1$ and $t = 2$, the pressure values are prescribed by initial conditions. Alternatively, pressure values and the first derivative[2] could be assigned at $t = 1$, from which the values at $t = 2$ can be calculated.

Formulas for boundary nodes are obtained by substituting a non-existent spatial point in the scheme (B.1) by the expressions obtained from discretized boundary conditions. For the frequency-independent acoustic absorbing boundary condition

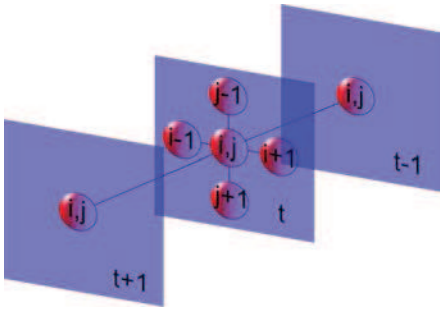


Figure B.1 – SLF stencil for the 2D wave equation.

[1] The speed of sound $c(\mathbf{r}, t) = c = 343\text{m/s}$ is uniform in all directions.

[2] Approximated, *e.g.* by the forward Euler scheme.

(4.10), proposed in [147], *e.g.* the missing point behind the right "wall" is evaluated as:

$$p^{t}_{i+1,j} = p^{t}_{i-1,j} + \frac{d_x}{c d_t \xi_{i,j}} \left(p^{t-1}_{i,j} - p^{t+1}_{i,j} \right), \tag{B.2}$$

which leads to a modified explicit expression:

$$p^{t+1}_{i,j} = \left[2(1-2\lambda^2) p^{t}_{i,j} + \lambda^2 (p^{t}_{i,j+1} + p^{t}_{i,j-1}) + 2\lambda^2 p^{t}_{i-1,j} - \left(1 - \frac{\lambda}{\xi_{i,j}} \right) p^{t-1}_{i,j} \right] \Big/ \left(1 + \frac{\lambda}{\xi_{i,j}} \right), \tag{B.3}$$

where $\lambda = c d_t / d_x = c d_t / d_y = c d_t / d_z$, assuming uniform spatial discretization, for simplicity.

When corners (and edges in 3D) are considered, the condition (4.10) is applied to all directions where the stencil points are missing. Generally, each part of the boundary (*i.e.* the missing node) is characterized by its own specific acoustic impedance coefficient, therefore their total number is higher than the number of boundary elements. For instance, bottom right corner in 2D requires two impedances, $\xi^{x}_{i,j}$ and $\xi^{y}_{i,j}$, representing the two intersecting walls:

$$p^{t+1}_{i,j} = \left[2(1-2\lambda^2) p^{t}_{i,j} + 2\lambda^2 (p^{t}_{i-1,j} + p^{t}_{i,j-1}) - \left(1 - \frac{\lambda}{\xi^{x}_{i,j}} - \frac{\lambda}{\xi^{y}_{i,j}} \right) p^{t-1}_{i,j} \right] \Big/ \left(1 + \frac{\lambda}{\xi^{x}_{i,j}} + \frac{\lambda}{\xi^{y}_{i,j}} \right). \tag{B.4}$$

Finally, to ensure stability of the scheme, spatial and temporal step sizes are bound to respect the *Courant-Friedrich-Lewy* (CFL) condition [231]: $c d_t / \min(d_x, d_y) \leq 1/\sqrt{2}$.

Concatenating these difference equations for the entire spatio-temporal dimension $s \times t$ yields a full rank / square-invertible matrix operator $\mathbf{A} \in \mathbb{R}^{st \times st}$.

Discretization of sound sources There are three ways of incorporating a sound source [226] in the FDTD-SLF scheme:

Hard source The pressure value $p^{t}_{i,j}$, at the source location (i, j, t), is replaced by a source signal sample at time t.

Soft source The expression (B.1) is enriched by adding a source signal sample at time t.

Transparent source Equivalent to soft source with additional negative term, introduced to compensate for the grid impulse response.

Hard sources are easy to implement, but suffer from serious scattering effects (hence, their name) and low-frequency artifacts. Soft sources do not scatter the waves, but change the source signal which is affected by the impulse response of the scheme. Transparent sources compensate for these effects, but require knowing the impulse response beforehand. We use soft sources, for simplicity, but ensure that the mean of a sound source signal is zero, to satisfy physical constraints noted in [226].

B.2 Witwer's Finite Difference Method for Poisson's equation

Witwer's FDM is derived by applying Kirchhoff's current law at each node [257]. Using this scheme, the total current flow value, $z_{i,j,k}$, at voxel[3] (i,j,k) is given by [122]:

$$\nabla\cdot\left(\sigma\nabla u_{i,j,k}\right)\approx z_{i,j,k}=g_{i,j,k}u_{i,j,k}-w_{i,j,k}^{(1,0,0)}u_{i+1,j,k}-w_{i,j,k}^{(-1,0,0)}u_{i-1,j,k}-w_{i,j,k}^{(0,1,0)}u_{i,j+1,k}- \\ w_{i,j,k}^{(0,-1,0)}u_{i,j-1,k}-w_{i,j,k}^{(0,0,1)}u_{i,j,k+1}-w_{i,j,k}^{(0,0,-1)}u_{i,j,k-1} \tag{B.5}$$

with:

$$w_{i,j,k}^{(m_1,m_2,m_3)}=\frac{2\alpha\sigma(\alpha[i,j,k]^T)\sigma(\alpha[i+m_1,j+m_2,k+m_3]^T)}{\sigma(\alpha[i,j,k]^T)+\sigma(\alpha[i+m_1,j+m_2,k+m_3]^T)}$$

$$g_{i,j,k}=w_{i,j,k}^{(1,0,0)}+w_{i,j,k}^{(-1,0,0)}+w_{i,j,k}^{(0,1,0)}+w_{i,j,k}^{(0,-1,0)}+w_{i,j,k}^{(0,0,1)}+w_{i,j,k}^{(0,0,-1)}$$

$$z_{i,j,k}=\begin{cases}z(\mathbf{r})/d & \text{if } \mathbf{r}^-=\alpha\,[i,j,k]^T\\ -z(\mathbf{r})/d & \text{if } \mathbf{r}^+=\alpha\,[i,j,k]^T\\ 0 & \text{otherwise,}\end{cases}.$$

where α denotes the spatial sampling stepsize and d is the distance between two monopoles.

Note that considering $(u_{i,j,k})$, $(w_{i,j,k}^{(m_1,m_2,m_3)})$ and $(g_{i,j,k})$ as third order arrays, and the Hadamard product between multi-way arrays, it is easy to implement the right hand side of equation (B.5) using matrix programming languages such as Matlab®.

[3] Here, the integer k represents the coordinate of the third grid axis and m. is an integer offset.

C Social sparsity declipper

Social sparsity declipper [227] is a sparse synthesis-based algorithm that shares some similarities with Consistent IHT (subsection 2.2). Both algorithms can be interpreted as sparsity-promoting, iterative shrinkages, applied to a *squared hinge* functional h^2 (recall the definitions of restriction operators $\mathbf{M}_r$ and $\mathbf{M}_c$):

$$h^2 = \| \mathbf{M}_r \mathbf{y} - \mathbf{M}_r \mathbf{x} \|_2^2 + \| \left(\mathbf{M}_c \mathbf{y} - \mathbf{M}_c \mathbf{x} \right)_+ \|_2^2, \tag{C.1}$$

where $(\cdot)_+$ is component-wise positive thresholding. The gradient of (C.1), along with a shrinkage operator, is used to simultaneously promote (structured) sparse solutions and maintain clipping consistency constraints. However, while Consistent IHT uses block-based processing, social sparsity declipper requires all data at once, in order to operate directly on time-frequency coefficients $x_{t,f}$. Consistent IHT features (progressive) hard thresholding, while social sparsity declipper uses so-called *Persistent Empirical Wiener (PEW)* shrinkage:

$$\mathfrak{S}_\lambda^{\text{PEW}}(x_{t,f}) = x_{t,f} \left(1 - \lambda^2 / \sum_{\tilde{t} \in \Upsilon} x_{\tilde{t},f}^2 \right)_+, \tag{C.2}$$

where Υ indicates the temporal neighborhood of the point (t,f). The associated optimization problem is not clearly defined, and presumed non-convex [148]. The authors propose relaxed *Iterative Soft Thresholding Algorithm (ISTA)* [88] heuristics, presented in Algorithm 5.

Algorithm 5 Social sparsity declipper

Require: $\mathbf{y}, \mathbf{M}_r, \mathbf{M}_c, \mathbf{D}, \tilde{\mathbf{z}}^{(0)}, \lambda, \gamma, \delta = \|\mathbf{D}\|_2^2, i = 0$
  repeat
    $\mathbf{g}_1 = \mathbf{D}^H \mathbf{M}_r{}^T \left(\mathbf{M}_r \mathbf{D} \mathbf{z}^{(i)} - \mathbf{M}_r \mathbf{y} \right)$
    $\mathbf{g}_2 = \mathbf{D}^H \mathbf{M}_c^T \left(\mathbf{M}_c \mathbf{D} \mathbf{z}^{(i)} - \mathbf{M}_C \mathbf{y} \right)_+$
    $\tilde{\mathbf{z}}^{(i+1)} = \mathfrak{S}_{\lambda/\delta} \left(\mathbf{z}^{(i)} - \frac{1}{\delta} (\mathbf{g}_1 + \mathbf{g}_2) \right)$
    $\mathbf{z}^{(i+1)} = \tilde{\mathbf{z}}^{(i+1)} + \gamma \left(\tilde{\mathbf{z}}^{(i+1)} - \tilde{\mathbf{z}}^{(i)} \right)$
    $i \leftarrow i + 1$
  until convergence
  return $\hat{\mathbf{x}} = \mathbf{D} \mathbf{z}^{(i+1)}$

D GReedy Analysis Structured Pursuit

The *GReedy Analysis Structured Pursuit (GRASP)* algorithm [188] is an extension of *Greedy Analysis Pursuit (GAP)* [186], tailored to group cosparse signals[1]. In the context of acoustic source localization, group cosparsity is modeled as joint cosparsity[2]. The pseudocode of GRASP is given in Algorithm 6.

Algorithm 6 GRASP

Require: $\mathbf{y} \in \mathbb{R}^{m}$, $\mathbf{m} \in \mathbb{R}^{m \times n}$, $\mathbf{A} \in \mathbb{R}^{d \times n}$, $\{\Psi_g\}_{g \in \Phi}$, σ, ϵ, h
Ensure: $i = 0$, $\hat{\Lambda} = [1, d]$, $\Phi^{(0)} = \Phi$
  repeat
    $\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{A}_{\hat{\Lambda}}\mathbf{x}\|_2^2$ s.t. $\|\mathbf{M}\mathbf{x} - \mathbf{y}\|_2 \leq \sigma$
    $\tilde{g} = \arg\max_{g \in \Phi^{(i)}} \|\mathbf{A}_{\Psi_g}\hat{\mathbf{x}}\|_2$
    $\Phi^{(i+1)} = \Phi^{(i)} \setminus \tilde{g}$
    $\hat{\Lambda} \leftarrow \hat{\Lambda} \setminus \Psi_{\tilde{g}}$
    $i \leftarrow i + 1$
  until $\|\mathbf{A}_{\hat{\Lambda}}\hat{\mathbf{x}}\|_\infty \leq \epsilon$ or $i = h$
  return $\hat{\mathbf{x}}$

Here, the set Ψ_g contains indices of the g^{th} group, such that $\cup_{g \in \Phi} \Psi_g = [1, d]$, where d is the number of rows of the analysis operator $\mathbf{A}$.

Evaluating the minimizer of the constrained linear least squares problem in the first iterative step is computationally the most expensive operation of GRASP (essentially, a nested optimization problem - may be solved by Weighted SDMM, for example).

[1] By *group cosparse* (w.r.t. the analysis operator $\mathbf{A}$), we denote a class of signals whose cosupport is represented by a union of row groups of the matrix $\mathbf{A}$. Hence, "usual" cosparsity is a special case of group cosparsity where the groups are individual rows of the analysis operator.

[2] If we express a vector $\mathbf{x} \in \mathbb{R}^{gp}$ as concatenation of p vectors of size g, *i.e.* $\mathbf{x} = \left[\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots \mathbf{x}_p^T\right]^T$, then $\mathrm{supp}(\mathbf{A}\mathbf{x}_1) = \mathrm{supp}(\mathbf{A}\mathbf{x}_2) = \ldots = \mathrm{supp}(\mathbf{A}\mathbf{x}_p)$, where $\mathrm{supp}(\mathbf{v})$ denotes the support of the vector $\mathbf{v}$.

Bibliography

[1] A. Adler, M. Elad, Y. Hel-Or, and E. Rivlin. Sparse coding with anomaly detection. In *IEEE International Workshop on Machine Learning for Signal Processing (MLSP), 2013*, pages 1–6. IEEE, 2013.

[2] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley. Audio inpainting. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(3):922–932, 2012.

[3] S. Affes, S. Gazor, and Y. Grenier. Robust adaptive beamforming via lms-like target tracking. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994*, volume 4, pages IV–269. IEEE, 1994.

[4] M. Aharon, M. Elad, and A. Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, 2006.

[5] L. Albera. Brain source localization using a physics-driven structured cosparse representation of eeg signals. Presentation at MLSP conference, 2014.

[6] L. Albera, A. Ferréol, D. Cosandier-Rimélé, I. Merlet, and F. Wendling. Brain source localization using a fourth-order deflation scheme. *IEEE Transactions on Biomedical Engineering*, 55(2):490–501, 2008.

[7] L. Albera, S. Kitić, N. Bertin, G. Puy, and R. Gribonval. Brain source localization using a physics-driven structured cosparse representation of EEG signals. In *IEEE International Workshop on Machine Learning for Signal Processing (MLSP), 2014*, pages 1–6. IEEE, 2014.

[8] J. B. Allen and D. A. Berkley. Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*, 65(4):943–950, 1979.

[9] M. G. Amin. *Through-the-wall radar imaging*. CRC press, 2011.

[10] N. Antonello, T. van Waterschoot, M. Moonen, P. Naylor, et al. Source localization and signal reconstruction in a reverberant field using the fdtd method. In *Proceedings of*

the 22nd European Signal Processing Conference (EUSIPCO), 2014, pages 301–305. IEEE, 2014.

[11] N. Antonello, T. van Waterschoot, M. Moonen, and P. A. Naylor. Evaluation of a numerical method for identifying surface acoustic impedances in a reverberant room. In *Proc. of the10th European Congress and Exposition on Noise Control Engineering*, pages 1–6, 2015.

[12] A. Asaei, M. Golbabaee, H. Bourlard, and V. Cevher. Structured sparsity models for reverberant speech separation. *IEEE Transactions on Audio, Speech, and Language Processing*, 22(3):620–633, 2014.

[13] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H. Seidel. Dynamic range independent image quality assessment. In *ACM Transactions on Graphics (TOG)*, volume 27, page 69. ACM, 2008.

[14] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski. Optimization with sparsity-inducing penalties. *Foundations and Trends® in Machine Learning*, 4(1):1–106, 2012.

[15] S. Baillet, J. C. Mosher, and R. M. Leahy. Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, 18(6):14–30, 2001.

[16] G. Bal. Introduction to inverse problems. *lecture notes*, 9:54, 2004.

[17] R. Balan, P. Casazza, and D. Edidin. On signal reconstruction without phase. *Applied and Computational Harmonic Analysis*, 20(3):345–356, 2006.

[18] R. Baraniuk and P. Steeghs. Compressive radar imaging. In *IEEE Radar Conference, 2007*, pages 128–133. IEEE, 2007.

[19] H. Becker. *Denoising, separation and localization of EEG sources in the context of epilepsy*. PhD thesis, Universite de Nice-Sophia Antipolis, 2014.

[20] J. Benesty. Adaptive eigenvalue decomposition algorithm for passive acoustic source localization. *The Journal of the Acoustical Society of America*, 107(1):384–391, 2000.

[21] J. Benesty, J. Chen, and Y. Huang. Time-delay estimation via linear interpolation and cross correlation. *IEEE Transactions on Speech and Audio Processing*, 12(5):509–519, 2004.

[22] J. Benesty, J. Chen, and Y. Huang. *Microphone array signal processing*, volume 1. Springer Science & Business Media, 2008.

[23] J. Benesty, S. Makino, and J. Chen. *Speech enhancement*. Springer Science & Business Media, 2005.

[24] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 417–424. ACM Press/Addison-Wesley Publishing Co., 2000.

[25] N. Bertin, S. Kitić, and R. Gribonval. Joint estimation of sound source location and boundary impedance with physics-driven cosparse regularization. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016*. IEEE, 2016. To appear.

[26] C. Bilen, S. Kitić, N. Bertin, and R. Gribonval. Sparse acoustic source localization with blind calibration for unknown medium characteristics. In *iTwist-2nd international-Traveling Workshop on Interactions between Sparse models and Technology*, 2014.

[27] C. Bilen, A. Ozerov, and P. Pérez. Joint audio inpainting and source separation. In *The 12th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA 2015)*, 2015.

[28] C. Bilen, G. Puy, R. Gribonval, and L. Daudet. Convex optimization approaches for blind sensor calibration using sparsity. *IEEE Transactions on Signal Processing*, 62(18):4847–4856, 2014.

[29] Å. Björck. *Numerical methods in matrix computations*. Springer, 2015.

[30] C. Blandin, A. Ozerov, and E. Vincent. Multi-source tdoa estimation in reverberant audio using angular spectra and clustering. *Signal Processing*, 92(8):1950–1960, 2012.

[31] N. Bleistein and J. K. Cohen. Nonuniqueness in the inverse source problem in acoustics and electromagnetics. *Journal of Mathematical Physics*, 18(2):194–201, 1977.

[32] T. Blumensath and M. E. Davies. Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27(3):265–274, 2009.

[33] T. Blumensath and M. E. Davies. Sampling theorems for signals from the union of finite-dimensional linear subspaces. *IEEE Transactions on Information Theory*, 55(4):1872–1882, 2009.

[34] J. Bolte, S. Sabach, and M. Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146(1-2):459–494, 2014.

[35] J. M. Borwein and A. S. Lewis. *Convex analysis and nonlinear optimization: theory and examples*. Springer Science & Business Media, 2010.

[36] G. E. Bottomley, T. Ottosson, and Y.-P. E. Wang. A generalized rake receiver for interference suppression. *IEEE Journal on Selected Areas in Communications*, 18(8):1536–1545, 2000.

[37] A. Bourrier, M. E. Davies, T. Peleg, P. Perez, and R. Gribonval. Fundamental performance limits for ideal decoders in high-dimensional linear inverse problems. *IEEE Transactions on Information Theory*, 60(12):7928–7946, 2014.

[38] G. E. Box. Science and statistics. *Journal of the American Statistical Association*, 71(356):791–799, 1976.

[39] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

[40] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[41] J. P. Boyle and R. L. Dykstra. A method for finding projections onto the intersection of convex sets in hilbert spaces. In *Advances in order restricted statistical inference*, pages 28–47. Springer, 1986.

[42] J. J. Bruer, J. A. Tropp, V. Cevher, and S. Becker. Time–data tradeoffs by aggressive smoothing. In *Advances in Neural Information Processing Systems*, pages 1664–1672, 2014.

[43] M. Bruneau. *Fundamentals of acoustics*. John Wiley & Sons, 2013.

[44] H. Buchner, R. Aichner, J. Stenglein, H. Teutsch, and W. Kellennann. Simultaneous localization of multiple sound sources using blind adaptive mimo filtering. In *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, volume 3, pages iii–97. IEEE, 2005.

[45] H. Buchner, G. Knoll, M. Fuchs, A. Rienäcker, R. Beckmann, M. Wagner, J. Silny, and J. Pesch. Inverse localization of electric dipole current sources in finite element models of the human head. *Electroencephalography and clinical Neurophysiology*, 102(4):267–278, 1997.

[46] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 31(1):59–73, 2011.

[47] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski. Phase retrieval via matrix completion. *SIAM Review*, 57(2):225–251, 2015.

[48] E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.

[49] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.

[50] E. J. Candès, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223, 2006.

[51] E. J. Candès and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, 2008.

[52] G. Casella and E. I. George. Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174, 1992.

[53] V. Cevher, A. Sankaranarayanan, M. F. Duarte, D. Reddy, R. G. Baraniuk, and R. Chellappa. Compressive sensing for background subtraction. In *Computer Vision–ECCV 2008*, pages 155–168. Springer, 2008.

[54] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky. The convex geometry of linear inverse problems. *Foundations of Computational mathematics*, 12(6):805–849, 2012.

[55] G. Chardon, A. Cohen, L. Daudet, et al. Reconstruction of solutions to the helmholtz equation from punctual measurements. In *SampTA-Sampling Theory and Applications 2013*, 2013.

[56] G. Chardon and L. Daudet. Narrowband source localization in an unknown reverberant environment usingwavefield sparse decomposition. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012*, pages 9–12. IEEE, 2012.

[57] G. Chardon, A. Leblanc, and L. Daudet. Plate impulse response spatial interpolation with sub-nyquist sampling. *Journal of sound and vibration*, 330(23):5678–5689, 2011.

[58] G. Chardon, T. Nowakowski, J. De Rosny, and L. Daudet. A blind dereverberation method for narrowband source localization. 2015.

[59] R. Chartrand. Nonconvex splitting for regularized low-rank+ sparse decomposition. *IEEE Transactions on Signal Processing*, 60(11):5810–5819, 2012.

[60] R. Chartrand and B. Wohlberg. A nonconvex ADMM algorithm for group sparsity with sparse groups. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013*, pages 6009–6013. IEEE, 2013.

[61] J. Chen, J. Benesty, and Y. Huang. Time delay estimation in room acoustic environments: an overview. *EURASIP Journal on applied signal processing*, 2006:170–170, 2006.

[62] J. C. Chen, K. Yao, and R. E. Hudson. Source localization and beamforming. *IEEE Signal Processing Magazine*, 19(2):30–39, 2002.

[63] Y. Chen, T. A. Davis, W. W. Hager, and S. Rajamanickam. Algorithm 887: Cholmod, supernodal sparse cholesky factorization and update/downdate. *ACM Transactions on Mathematical Software (TOMS)*, 35(3):22, 2008.

[64] I. Chiba, T. Takahashi, and Y. Karasawa. Transmitting null beam forming with beam space adaptive array antennas. In *Vehicular Technology Conference, 1994 IEEE 44th*, pages 1498–1502. IEEE, 1994.

[65] P.-J. Chung, J. F. Böhme, and A. O. Hero. Tracking of multiple moving sources using recursive em algorithm. *EURASIP Journal on Applied Signal Processing*, 2005:50–60, 2005.

[66] M. Cobos, A. Marti, and J. J. Lopez. A modified srp-phat functional for robust real-time sound source localization with scalable spatial sampling. *IEEE Signal Processing Letters*, 18(1):71–74, 2011.

[67] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212. Springer, 2011.

[68] A. Criminisi, P. Pérez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on Image Processing*, 13(9):1200–1212, 2004.

[69] B. N. Cuffin. A method for localizing eeg sources in realistic head models. *IEEE Transactions on Biomedical Engineering*, 42(1):68–71, 1995.

[70] J.-P. Dalmont. Acoustic impedance measurement, part i: A review. *Journal of Sound and Vibration*, 243(3):427–439, 2001.

[71] J.-P. Dalmont. Acoustic impedance measurement, part ii: a new calibration method. *Journal of sound and vibration*, 243(3):441–459, 2001.

[72] F. Darvas, D. Pantazis, E. Kucukaltun-Yildirim, and R. Leahy. Mapping human brain function with meg and eeg: methods and validation. *NeuroImage*, 23:S289–S299, 2004.

[73] G. Dassios and A. Fokas. The definite non-uniqueness results for deterministic eeg and meg data. *Inverse Problems*, 29(6):065012, 2013.

[74] T. A. Davis. *Direct methods for sparse linear systems*, volume 2. Siam, 2006.

[75] B. Defraene, N. Mansour, S. De Hertogh, T. van Waterschoot, M. Diehl, and M. Moonen. Declipping of audio signals using perceptual compressed sensing. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(12):2627–2637, 2013.

[76] A. Deleforge and R. Horaud. The cocktail party robot: Sound source separation and localisation with an active binaural head. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 431–438. ACM, 2012.

[77] L. Demanet. *Curvelets, wave atoms, and wave equations*. PhD thesis, California Institute of Technology, 2006.

[78] A. Devaney and G. C. Sherman. Nonuniqueness in inverse source and scattering problems. *IEEE Transactions on Antennas and Propagation*, 30(5):1034–1037, 1982.

[79] A. J. Devaney. *Mathematical Foundations of Imaging, Tomography and Wavefield Inversion*. Cambridge University Press, 2012.

[80] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein. Robust localization in reverberant rooms. In *Microphone Arrays*, pages 157–180. Springer, 2001.

[81] L. Ding and B. He. Sparse source imaging in electroencephalography with accurate field modeling. *Human brain mapping*, 29(9):1053–1067, 2008.

[82] J. Dmochowski, J. Benesty, and S. Affes. Direction of arrival estimation using the parameterized spatial correlation matrix. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(4):1327–1339, 2007.

[83] J. P. Dmochowski, J. Benesty, and S. Affes. A generalized steered response power method for computationally viable source localization. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(8):2510–2526, 2007.

[84] H. Do, H. F. Silverman, and Y. Yu. A real-time srp-phat source location implementation using stochastic region contraction (src) on a large-aperture microphone array. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007.*, volume 1, pages I–121. IEEE, 2007.

[85] I. Dokmanć, R. Scheibler, and M. Vetterli. Raking the cocktail party. *IEEE Journal of Selected Topics In Signal Processing*, 9, 2015.

[86] I. Dokmanić. *Listening to Distances and Hearing Shapes*. PhD thesis, EPFL, 2015.

[87] I. Dokmanić and M. Vetterli. Room helps: Acoustic localization with finite elements. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012*, pages 2617–2620. Ieee, 2012.

[88] D. L. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, 1995.

[89] V. Duval and G. Peyré. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, pages 1–41, 2015.

[90] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1-3):293–318, 1992.

[91] J. Eckstein and W. Yao. Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives. *Pacific Journal on Optimization, to appear*, 2014.

[92] M. Elad. *The Quest for a Dictionary*. Springer, 2010.

[93] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, 2006.

[94] M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. *Inverse problems*, 23(3):947, 2007.

[95] M. Elad, J.-L. Starck, P. Querre, and D. L. Donoho. Simultaneous cartoon and texture image inpainting using morphological component analysis (mca). *Applied and Computational Harmonic Analysis*, 19(3):340–358, 2005.

[96] Y. C. Eldar and G. Kutyniok. *Compressed sensing: theory and applications*. Cambridge University Press, 2012.

[97] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

[98] O. G. Ernst and M. J. Gander. Why it is difficult to solve helmholtz problems with classical iterative methods. In *Numerical analysis of multiscale problems*, pages 325–363. Springer, 2012.

[99] L. Evans. Partial differential equations (graduate studies in mathematics, vol. 19). *Instructor*, page 67, 2009.

[100] F. J. Fahy. *Foundations of engineering acoustics*. Academic press, 2000.

[101] R. Falgout. An introduction to algebraic multigrid computing. *Computing in Science Engineering*, 8(6):24–33, 2006.

[102] M. A. Figueiredo and R. D. Nowak. Sparse estimation with strongly correlated variables using ordered weighted l1 regularization. *arXiv preprint arXiv:1409.4005*, 2014.

[103] M. Fink. Time-reversal acoustics in complex environments. *Geophysics*, 71(4):SI151–SI164, 2006.

[104] D. C.-L. Fong and M. Saunders. Lsmr: An iterative algorithm for sparse least-squares problems. *SIAM Journal on Scientific Computing*, 33(5):2950–2971, 2011.

[105] M. Fornasier and H. Rauhut. Compressive sensing. In *Handbook of mathematical methods in imaging*, pages 187–228. Springer, 2011.

[106] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*. Springer, 2013.

[107] S. Gannot, D. Burshtein, and E. Weinstein. Signal enhancement using beamforming and nonstationarity with applications to speech. *IEEE Transactions on Signal Processing*, 49(8):1614–1626, 2001.

[108] W. R. Gilks. *Markov chain monte carlo*. Wiley Online Library, 2005.

[109] R. Giryes, S. Nam, M. Elad, R. Gribonval, and M. E. Davies. Greedy-like algorithms for the cosparse analysis model. *Linear Algebra and its Applications*, 441:22–60, 2014.

[110] S. Godsill, P. Rayner, and O. Cappé. *Digital audio restoration*. Springer, 2002.

[111] G. H. Golub and C. F. Van Loan. *Matrix computations*, volume 3. JHU Press, 2012.

[112] R. C. Gonzalez, R. E. Woods, and S. L. Eddins. *Digital image processing using MATLAB*. Pearson Education India, 2004.

[113] I. F. Gorodnitsky, J. S. George, and B. D. Rao. Neuromagnetic source imaging with focuss: a recursive weighted minimum norm algorithm. *Electroencephalography and clinical Neurophysiology*, 95(4):231–251, 1995.

[114] J. Gorski, F. Pfeuffer, and K. Klamroth. Biconvex sets and optimization with biconvex functions: a survey and extensions. *Mathematical Methods of Operations Research*, 66(3):373–407, 2007.

[115] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka. RWC music database: Popular, classical and jazz music databases. In *ISMIR*, volume 2, pages 287–288, 2002.

[116] A. Gotthelf and J. G. Lennox. *Philosophical issues in Aristotle's biology*. Cambridge University Press, 1987.

[117] R. M. Gray and D. L. Neuhoff. Quantization. *IEEE Transactions on Information Theory*, 44(6):2325–2383, 1998.

[118] R. Grech, T. Cassar, J. Muscat, K. P. Camilleri, S. G. Fabri, M. Zervakis, P. Xanthopoulos, V. Sakkalis, and B. Vanrumste. Review on solving the inverse problem in eeg source analysis. *Journal of neuroengineering and rehabilitation*, 5(1):25, 2008.

[119] R. Gribonval. Should penalized least squares regression be interpreted as maximum a posteriori estimation? *IEEE Transactions on Signal Processing*, 59(5):2405–2410, 2011.

[120] S. M. Griebel and M. S. Brandstein. Microphone array source localization using realizable delay vectors. In *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics, 2001*, pages 71–74. IEEE, 2001.

[121] J. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton university bulletin*, 13(49-52):28, 1902.

[122] H. Hallez, B. Vanrumste, R. Grech, J. Muscat, W. De Clercq, A. Vergult, Y. D'Asseler, K. P. Camilleri, S. G. Fabri, S. Van Huffel, et al. Review on solving the forward problem in eeg source analysis. *Journal of neuroengineering and rehabilitation*, 4(1):46, 2007.

[123] A. C. Hansen and B. Adcock. Generalized sampling and infinite dimensional compressed sensing. *Magnetic Resonance Imaging*, page 1, 2011.

[124] M. J. Harvilla and R. M. Stern. Least squares signal declipping for robust speech recognition. In *Fifteenth Annual Conference of the International Speech Communication Association*, 2014.

[125] M. Herman, T. Strohmer, et al. High-resolution radar via compressed sensing. *IEEE Transactions on Signal Processing*, 57(6):2275–2284, 2009.

[126] M. Herman, T. Strohmer, et al. General deviants: An analysis of perturbations in compressed sensing. *IEEE Journal of Selected topics in signal processing*, 4(2):342–349, 2010.

[127] M. Hong, Z.-Q. Luo, and M. Razaviyayn. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. *arXiv preprint arXiv:1410.1390*, 2014.

[128] Y. A. Huang and J. Benesty. Adaptive multi-channel least mean square and newton algorithms for blind channel identification. *Signal Processing*, 82(8):1127–1138, 2002.

[129] V. Isakov. *Inverse problems for partial differential equations*, volume 127. Springer Science & Business Media, 2006.

[130] B. H. Jansen and V. G. Rit. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological cybernetics*, 73(4):357–366, 1995.

[131] A. Janssen, R. Veldhuis, and L. Vries. Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 34(2):317–330, 1986.

[132] R. Jenatton, J.-Y. Audibert, and F. Bach. Structured variable selection with sparsity-inducing norms. *The Journal of Machine Learning Research*, 12:2777–2824, 2011.

[133] M. Kabanava and H. Rauhut. Cosparsity in compressed sensing. In *Compressed Sensing and its Applications*. Springer, 2015.

[134] M. Kahrs and K. Brandenburg. *Applications of digital signal processing to audio and acoustics*, volume 437. Springer Science & Business Media, 1998.

[135] J. Kaipio and E. Somersalo. *Statistical and computational inverse problems*, volume 160. Springer Science & Business Media, 2006.

[136] L. Kaufman and Z.-L. Lu. Basics of neuromagnetism and magnetic source imaging. In *Magnetic Source Imaging of Human Brain*. Taylor & Francis, 2003.

[137] R. L. Kirlin, D. F. Moore, and R. F. Kubichek. Improvement of delay measurements from sonar arrays via sequential state estimation. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 29(3):514–519, 1981.

[138] A. Kirsch. *An introduction to the mathematical theory of inverse problems*, volume 120. Springer Science & Business Media, 2011.

[139] S. Kitić, L. Albera, N. Bertin, and R. Gribonval. Physics-driven inverse problems made tractable with cosparse regularization. *IEEE Transactions on Signal Processing*, 2015.

[140] S. Kitić, N. Bertin, and R. Gribonval. A review of cosparse signal recovery methods applied to sound source localization. In *Le XXIVe colloque Gretsi*, 2013.

[141] S. Kitić, N. Bertin, and R. Gribonval. Audio declipping by cosparse hard thresholding. In *iTwist-2nd international-Traveling Workshop on Interactions between Sparse models and Technology*, 2014.

[142] S. Kitić, N. Bertin, and R. Gribonval. Hearing behind walls: localizing sources in the room next door with cosparsity. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014*, pages 3087–3091. IEEE, 2014.

[143] S. Kitić, N. Bertin, and R. Gribonval. Sparsity and cosparsity for audio declipping: a flexible non-convex approach. In *Latent Variable Analysis and Signal Separation*, pages 243–250. Springer, 2015.

[144] S. Kitić, L. Jacques, N. Madhu, M. P. Hopwood, A. Spriet, and C. De Vleeschouwer. Consistent Iterative Hard Thresholding for signal declipping. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013*, pages 5939–5943. IEEE, 2013.

[145] C. H. Knapp and G. C. Carter. The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24(4):320–327, 1976.

[146] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, (8):30–37, 2009.

[147] K. Kowalczyk and M. v. Walstijn. Modeling frequency-dependent boundaries as digital impedance filters in fdtd and k-dwm room acoustics simulations. *Journal of the Audio Engineering Society*, 56(7/8):569–583, 2008.

[148] M. Kowalski, K. Siedenburg, and M. Dorfler. Social sparsity! neighborhood systems enrich structured shrinkage operators. *IEEE Transactions on Signal Processing*, 61(10):2498–2511, 2013.

[149] S. Kraft and U. Zölzer. Beaqlejs: Html5 and javascript based framework for the subjective evaluation of audio quality. In *Linux Audio Conference, Karlsruhe, DE*, 2014.

[150] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T.-W. Lee, and T. J. Sejnowski. Dictionary learning algorithms for sparse representation. *Neural computation*, 15(2):349–396, 2003.

[151] H. W. Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.

[152] H. Kuttruff. *Acoustics: An Introduction*. CRC Press, 2007.

[153] H. Kuttruff. *Room acoustics*. CRC Press, 2009.

[154] J. P. Lachaux, D. Rudrauf, and P. Kahane. Intracranial eeg and human brain mapping. *Journal of Physiology-Paris*, 97(4):613–628, 2003.

[155] E. Larsen and R. M. Aarts. *Audio bandwidth extension: application of psychoacoustics, signal processing and loudspeaker design*. John Wiley & Sons, 2005.

[156] J. Le Roux, P. T. Boufounos, K. Kang, and J. R. Hershey. Source localization in reverberant environments using sparse optimization. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013*, pages 4310–4314. IEEE, 2013.

[157] J. Le Roux, H. Kameoka, N. Ono, A. De Cheveigne, and S. Sagayama. Computational auditory induction as a missing-data model-fitting problem with bregman divergence. *Speech Communication*, 53(5):658–676, 2011.

[158] J. Lee. *Introduction to topological manifolds*, volume 940. Springer Science & Business Media, 2010.

[159] R. J. LeVeque. *Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems*, volume 98. Siam, 2007.

[160] H. Lewy. An example of a smooth linear partial differential equation without solution. *Annals of Mathematics*, pages 155–158, 1957.

[161] X. Li and L. J. Cimini. Effects of clipping and filtering on the performance of OFDM. In *47th IEEE Vehicular Technology Conference*, volume 3, pages 1634–1638. IEEE, 1997.

[162] S. Ling and T. Strohmer. Self-calibration and biconvex compressive sensing. *arXiv preprint arXiv:1501.06864*, 2015.

[163] P.-L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

[164] Y. Liu, M. De Vos, and S. V. Huffel. Compressed sensing of multi-channel eeg signals: The simultaneous cosparsity and low rank optimization. 2015.

[165] D. Lorenz and N. Worliczek. Necessary conditions for variational regularization schemes. *Inverse Problems*, 29(7):075016, 2013.

[166] Y. M. Lu and M. N. Do. Sampling signals from a union of subspaces. *IEEE Signal Processing Magazine*, 25(2):41–47, 2008.

[167] Y. Luo and G. T. Schuster. Wave-equation traveltime inversion. *Geophysics*, 56(5):645–653, 1991.

[168] J. Mairal, F. Bach, and J. Ponce. Task-driven dictionary learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4):791–804, 2012.

[169] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online dictionary learning for sparse coding. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 689–696. ACM, 2009.

[170] A. Majumdar and R. K. Ward. Non-convex row-sparse multiple measurement vector analysis prior formulation for eeg signal reconstruction. *Biomedical Signal Processing and Control*, 13:142–147, 2014.

[171] D. Malioutov. A sparse signal reconstruction perspective for source localization with sensor arrays, master's thesis, 2003.

[172] D. Malioutov, M. Çetin, and A. S. Willsky. A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Transactions on Signal Processing*, 53(8):3010–3022, 2005.

[173] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

[174] J. McCaskill. Automatic camera tracking using beamforming, Dec. 27 2005. US Patent 6,980,485.

[175] M. Miga, T. E. Kerner, T. M. Darcey, et al. Source localization using a current-density minimization approach. *IEEE Transactions on Biomedical Engineering*, 49(7):743–745, 2002.

[176] J. Miller. Why the world is on the back of a turtle, 1974.

[177] S. Miron, N. Le Bihan, and J. I. Mars. Vector-sensor music for polarized seismic sources localization. *EURASIP Journal on Applied Signal Processing*, 2005:74–84, 2005.

[178] M. Mohsina and A. Majumdar. Gabor based analysis prior formulation for eeg signal reconstruction. *Biomedical Signal Processing and Control*, 8(6):951–955, 2013.

[179] J.-J. Moreau. Proximité et dualité dans un espace hilbertien. *Bulletin de la Société mathématique de France*, 93:273–299, 1965.

[180] V. A. Morozov. *Methods for solving incorrectly posed problems*. Springer Science & Business Media, 2012.

[181] J. C. Mosher and R. M. Leahy. Source localization using recursively applied and projected (rap) music. *IEEE Transactions on Signal Processing*, 47(2):332–340, 1999.

[182] W. Munk, P. Worcester, and C. Wunsch. *Ocean acoustic tomography*. Cambridge University Press, 2009.

[183] S. K. Naik and C. A. Murthy. Hue-preserving color image enhancement without gamut problem. *IEEE Transactions on Image Processing*, 12(12):1591–1598, 2003.

[184] K. Nakadai, H. Nakajima, K. Yamada, Y. Hasegawa, T. Nakamura, and H. Tsujino. Sound source tracking with directivity pattern estimation using a 64 ch microphone array. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2005.(IROS 2005).*, pages 1690–1696. IEEE, 2005.

[185] K. Nakadai, H. G. Okuno, H. Kitano, et al. Real-time sound source localization and separation for robot audition. In *INTERSPEECH*, 2002.

[186] S. Nam, M. E. Davies, M. Elad, and R. Gribonval. Cosparse analysis modeling-uniqueness and algorithms. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011*, pages 5804–5807. IEEE, 2011.

[187] S. Nam, M. E. Davies, M. Elad, and R. Gribonval. The cosparse analysis model and algorithms. *Applied and Computational Harmonic Analysis*, 34(1):30–56, 2013.

[188] S. Nam and R. Gribonval. Physics-driven structured cosparse modeling for source localization. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012*, pages 5397–5400. IEEE, 2012.

[189] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM journal on computing*, 24(2):227–234, 1995.

[190] G. P. Nava, Y. Yasuda, Y. Sato, and S. Sakamoto. On the in situ estimation of surface acoustic impedance in interiors of arbitrary shape by acoustical inverse methods. *Acoustical science and technology*, 30(2):100–109, 2009.

[191] D. Needell and J. A. Tropp. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2009.

[192] J. Nocedal and S. Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[193] P. L. Nunez and R. Srinivasan. *Electric fields of the brain: the neurophysics of EEG*. Oxford university press, 2006.

[194] H. Ocak. Optimal classification of epileptic seizures in eeg using wavelet analysis and genetic algorithm. *Signal processing*, 88(7):1858–1867, 2008.

[195] M. Omologo and P. Svaizer. Use of the crosspower-spectrum phase in acoustic event location. *IEEE Transactions on Speech and Audio Processing*, 5(3):288–292, 1997.

[196] R. Oostenveld and P. Praamstra. The five percent electrode system for high-resolution eeg and erp measurements. *Clinical neurophysiology*, 112(4):713–719, 2001.

[197] Y. Oualil, F. Faubel, and D. Klakow. A fast cumulative steered response power for multiple speaker detection and localization. In *Proceedings of the 21st European Signal Processing Conference (EUSIPCO), 2013*, pages 1–5. IEEE, 2013.

[198] J. L. Paredes, G. R. Arce, and Z. Wang. Ultra-wideband compressed sensing: channel estimation. *IEEE Journal of Selected Topics in Signal Processing*, 1(3):383–395, 2007.

[199] N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in optimization*, 1(3):123–231, 2013.

[200] P. Parpas, D. V. Luong, D. Rueckert, and B. Rustem. A multilevel proximal algorithm for large scale composite convex optimization. 2014.

[201] F. Parvaresh, H. Vikalo, S. Misra, and B. Hassibi. Recovering sparse signals using sparse measurement matrices in compressed dna microarrays. *IEEE Journal of Selected Topics in Signal Processing*, 2(3):275–285, 2008.

[202] R. D. Pascual-Marqui et al. Standardized low-resolution brain electromagnetic tomography (sloreta): technical details. *Methods Find Exp Clin Pharmacol*, 24(Suppl D):5–12, 2002.

[203] R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *International Journal of psychophysiology*, 18(1):49–65, 1994.

[204] Y. C. Pati, R. Rezaiifar, and P. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers, 1993.*, pages 40–44. IEEE, 1993.

[205] C. Perkins, O. Hodson, and V. Hardman. A survey of packet loss recovery techniques for streaming audio. *IEEE Network*, 12(5):40–48, 1998.

[206] M. D. Plumbley, T. Blumensath, L. Daudet, R. Gribonval, and M. E. Davies. Sparse representations in audio and music: from coding to source separation. *Proceedings of the IEEE*, 98(6):995–1005, 2010.

[207] Y. L. Polo, Y. Wang, A. Pandharipande, and G. Leus. Compressive wide-band spectrum sensing. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009.*, pages 2337–2340. IEEE, 2009.

[208] G. Puy, M. E. Davies, and R. Gribonval. Linear embeddings of low-dimensional subsets of a Hilbert space to R^m. In *European Signal Processing Conference (EUSIPCO), 2015*. IEEE, 2015.

[209] R. Ramirez. Source localization. http://www.scholarpedia.org/article/Source_localization#An_overview_of_all_the_models, 2008. Accessed: 2015-09-16.

[210] M. A. Richards. *Fundamentals of radar signal processing*. Tata McGraw-Hill Education, 2005.

[211] S. W. Rienstra and A. Hirschberg. An introduction to acoustics. *Eindhoven University of Technology*, 18:19, 2003.

[212] D. Romero, R. Lopez-Valcarce, and G. Leus. Compression limits for random vectors with linearly parameterized second-order statistics. *IEEE Transactions on Information Theory*, 61(3):1410–1425, 2015.

[213] F. H. Rose, C. D. McGregor, and P. Mathur. Method and apparatus for stone localization using ultrasound imaging, Jan. 30 1990. US Patent 4,896,673.

[214] R. Roy and T. Kailath. ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(7):984–995, 1989.

[215] R. Rubinstein, T. Peleg, and M. Elad. Analysis k-svd: A dictionary-learning algorithm for the analysis sparse model. *IEEE Transactions on Signal Processing*, 61(3):661–677, 2013.

[216] R. Rubinstein, M. Zibulevsky, and M. Elad. Double sparsity: Learning sparse dictionaries for sparse signal approximation. *IEEE Transactions on Signal Processing*, 58(3):1553–1564, 2010.

[217] M. D. Rugg and M. G. Coles. *Electrophysiology of mind: Event-related brain potentials and cognition.* Oxford University Press, 1995.

[218] Y. Saad. *Iterative methods for sparse linear systems*. Siam, 2003.

[219] S. Sanei and J. A. Chambers. *EEG signal processing*. John Wiley & Sons, 2013.

[220] J. C. Santamarina and D. Fratta. *Discrete signals and inverse problems: an introduction for engineers and scientists*. John Wiley & Sons, 2005.

[221] M. Scherg and D. Von Cramon. Two bilateral sources of the late aep as identified by a spatio-temporal dipole model. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 62(1):32–44, 1985.

[222] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational methods in imaging*, volume 2. Springer, 2009.

[223] R. O. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, 1986.

[224] M. Seibert, J. Wörmann, R. Gribonval, and M. Kleinsteuber. Learning co-sparse analysis operators with separable structures. *arXiv preprint arXiv:1503.02398*, 2015.

[225] S. Shalev-Shwartz and N. Srebro. SVM optimization: inverse dependence on training set size. In *Proceedings of the 25th international conference on Machine learning*, pages 928–935. ACM, 2008.

[226] J. Sheaffer, M. van Walstijn, and B. Fazenda. Physical and numerical constraints in source modeling for finite difference simulation of room acoustics. *The Journal of the Acoustical Society of America*, 135(1):251–261, 2014.

[227] K. Siedenburg, M. Kowalski, and M. Dorfler. Audio declipping with social sparsity. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1577–1581. IEEE, 2014.

[228] R. J. Solomonoff. A formal theory of inductive inference. part ii. *Information and control*, 7(2):224–254, 1964.

[229] P. Stoica and J. Li. Lecture notes-source localization from range-difference measurements. *Signal Processing Magazine, IEEE*, 23(6):63–66, 2006.

[230] G. Strang. Fast transforms: Banded matrices with banded inverses. *Proceedings of the National Academy of Sciences*, 107(28):12413–12416, 2010.

[231] J. C. Strikwerda. *Finite difference schemes and partial differential equations*. Siam, 2004.

[232] D. L. Sun and C. Fevotte. Alternating direction method of multipliers for non-negative matrix factorization with the beta-divergence. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014*, pages 6201–6205. IEEE, 2014.

[233] J.-F. Synnevåg, A. Austeng, and S. Holm. Adaptive beamforming applied to medical ultrasound imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 54(8):1606–1613, 2007.

[234] Y. Tachioka, T. Narita, and J. Ishii. Speech recognition performance estimation for clipped speech based on objective measures. *Acoustical Science and Technology*, 35(6):324–326, 2014.

[235] M. Teplan. Fundamentals of eeg measurement. *Measurement science review*, 2(2):1–11, 2002.

[236] A. N. Tikhonov and A. V. Ia. *Metody resheniia nekorrectnykh zadach*. Fizmatizdat, 1979.

[237] A. M. Tillmann, R. Gribonval, and M. E. Pfetsch. Projection onto the cosparse set is NP-hard. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014*, pages 7148–7152. IEEE, 2014.

[238] A. M. Tillmann and M. E. Pfetsch. The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing. *IEEE Transactions on Information Theory*, 60(2):1248–1259, 2014.

[239] I. Tošić and P. Frossard. Dictionary learning. *IEEE Signal Processing Magazine*, 28(2):27–38, 2011.

[240] E. Treister and I. Yavneh. A multilevel iterated-shrinkage approach to penalized least-squares minimization. *IEEE Transactions on Signal Processing*, 60(12):6319–6329, 2012.

[241] K. Uutela, M. Hämäläinen, and E. Somersalo. Visualization of magnetoencephalographic data using minimum current estimates. *NeuroImage*, 10(2):173–180, 1999.

[242] S. Vaiter, G. Peyré, C. Dossal, and J. Fadili. Robust sparse analysis regularization. *IEEE Transactions on Information Theory*, 59(4):2001–2016, 2013.

[243] J.-M. Valin, F. Michaud, J. Rouat, and D. Létourneau. Robust sound source localization using a microphone array on a mobile robot. In *Proceedings. 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2003.(IROS 2003).*, volume 2, pages 1228–1233. IEEE, 2003.

[244] G. Van Hoey, R. Van de Walle, B. Vanrumste, M. D'Havse, I. Lemahieu, and P. Boon. Beamforming techniques applied in eeg source analysis. *Proc. ProRISC99*, 10:545–549, 1999.

[245] P. Vary and R. Martin. *Digital speech transmission: Enhancement, coding and error concealment*. John Wiley & Sons, 2006.

[246] N. Vaswani. Kalman filtered compressed sensing. In *15th IEEE International Conference on Image Processing, 2008. ICIP 2008.*, pages 893–896. IEEE, 2008.

[247] E. Vincent, M. Jafari, and M. Plumbley. Preliminary guidelines for subjective evalutation of audio source separation algorithms. In *UK ICA Research Network Workshop*, 2006.

[248] J. Vorwerk. Comparison of numerical approaches to the eeg forward problem. *Westfälische Wilhelms-Universität Münster*, 2011.

[249] M. Wagner, M. Fuchs, H.-A. Wischmann, R. Drenckhahn, and T. Köhler. Smooth reconstruction of cortical sources from eeg or meg recordings. *NeuroImage*, 3(3):S168, 1996.

[250] C. Wang and M. S. Brandstein. A hybrid real-time face tracking system. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, 1998.*, volume 6, pages 3737–3740. IEEE, 1998.

[251] Y. Wang, W. Yin, and J. Zeng. Global convergence of ADMM in nonconvex nonsmooth optimization. *arXiv preprint arXiv:1511.06324*, 2015.

[252] D. B. Ward, E. Lehmann, R. C. Williamson, et al. Particle filtering algorithms for tracking an acoustic source in a reverberant environment. *IEEE Transactions on Speech and Audio Processing*, 11(6):826–836, 2003.

[253] R. Warren and S. Osher. Hyperspectral unmixing by the alternating direction method of multipliers. *Inverse Problems and Imaging*, 2015.

[254] L. T. Watson, M. Bartholomew-Biggs, and J. A. Ford. *Optimization and nonlinear equations*, volume 4. Gulf Professional Publishing, 2001.

[255] M. Wax and T. Kailath. Optimum localization of multiple sources by passive arrays. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 31(5):1210–1217, 1983.

[256] A. J. Weinstein and M. B. Wakin. Recovering a clipped signal in Sparseland. *arXiv preprint arXiv:1110.5063*, 2011.

[257] J. G. Witwer, G. J. Trezek, and D. L. Jewett. The effect of media inhomogeneities upon intracranial electrical fields. *IEEE Transactions on Biomedical Engineering*, (5):352–362, 1972.

[258] J. Wormann, S. Hawe, and M. Kleinsteuber. Analysis based blind compressive sensing. *IEEE Signal Processing Letters*, 20(5):491–494, 2013.

[259] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan. Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*, 98(6):1031–1044, 2010.

[260] M. Yaghoobi, T. Blumensath, and M. E. Davies. Dictionary learning for sparse approximations with the majorization method. *IEEE Transactions on Signal Processing*, 57(6):2178–2191, 2009.

[261] M. Yaghoobi, S. Nam, R. Gribonval, and M. E. Davies. Analysis operator learning for overcomplete cosparse representations. In *19th European Signal Processing Conference, 2011*, pages 1470–1474. IEEE, 2011.

[262] H. Yan. *Signal processing for magnetic resonance imaging and spectroscopy*, volume 15. CRC Press, 2002.

[263] M. Yannakakis. Computing the minimum fill-in is NP-complete. *SIAM Journal on Algebraic Discrete Methods*, 2(1):77–79, 1981.

[264] X. Zhang and J. H. Hansen. Csa-bf: A constrained switched adaptive beamformer for speech enhancement and recognition in real car environments. *IEEE Transactions on Speech and Audio Processing*, 11(6):733–745, 2003.

[265] L. Zhukov. Physics of EEG/MEG. http://leonidzhukov.net/content/eeg_meg_physics/node1.html, 1999. Accessed: 2015-09-11.