An iterative algorithm for recovering the phase of complex components from their mixture

Un algorithme itératif de reconstruction de phase de composantes à partir de leur mélange

Paul Magron   Roland Badeau   Bertrand David

LTCI, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France
<firstname>.<lastname>@telecom-paristech.fr

Abstract

This report addresses the problem of estimating complex components from their mixture in the Time-Frequency (TF) domain. Traditional techniques, which consist in non-iteratively optimizing a cost function measuring the difference between the mixture and the model, do not lead to satisfactorily sounding results. Thus, we propose to optimize this cost function by means of an iterative algorithm, which allows us to incorporate some prior phase information in the procedure. We provide a mathematical proof of the non-increasing property of the error function over the update rules of this algorithm. In addition, we show that the algorithm must be carefully initialized to avoid getting stuck in a local minimum and to output satisfying results.

Key words

Phase reconstruction, source separation, auxiliary function method, linear unwrapping.

Résumé

Ce rapport s'intéresse au problème de l'estimation de composantes complexes à partir de leur mélange dans le plan Temps-Fréquence. Pour ce faire, nous proposons de rechercher un minimum local d'une fonction de coût qui mesure la différence entre le mélange et la somme des sources estimées. Bien que des méthodes non-itératives fournissent des solutions à ce problème, elles ne sont pas perceptivement satisfaisantes. Aussi, nous proposons d'optimiser la fonction de coût par une procédure itérative, qui permet, via l'initialisation des composantes, d'injecter des à priori sur leurs phases. Nous fournissons une preuve mathématique que les règles de mise à jour de cet algorithme font décroître la fonction de coût associée. Nous soulignons également l'importance d'une initialisation bien choisie pour aboutir à des résultats satisfaisants, sans être bloqué dans un minimum local.

Mots clés

Reconstruction de phase, séparation de sources, méthode de la fonction auxiliaire, déroulé linéaire.

1 Introduction

This document features supplementary materials to the reference paper [1]. We address the problem of source separation, which consists in extracting the underlying components composing a mixture of audio sources in the TF domain. Since all TF bins are treated independently, the problem here reduces to finding K complex numbers $X_1,...,X_K$ such that $\sum_k X_k = X$, given their magnitude estimates $V_1,...,V_k$. To do so, we consider the problem of optimizing a cost function $|E|$ which measures the difference between the observed data and the model. We establish that this function has many global minima. For instance, the estimates provided by the Wiener filtering technique are zeros of the cost function E, although they do not verify $|\hat{X}_k| = V_k$. Since this technique does not lead to satisfactorily sounding results [2], we propose to optimize $|E[$ by means of an iterative procedure. Our approach allows us, through the initialization of the procedure, to incorporate some prior knowledge about the phase of the components, which can increase the quality of the separated components [2].

We first present in section 2 the problem setting and motivate the need for a novel optimization technique of $|E|$, different from traditional non-iterative methods. In section 3, we present the algorithm and we provide a mathematical proof that the error function $|E|$ is non-creasing under the corresponding update rules. Section 4 motivates the research of a properly-chosen initialization scheme for the algorithm, and section 5 draws some concluding remarks.

2 Problem setting

We consider the problem of estimating K complex components $X_1,...,X_K$ from their sum $X = \sum_k X_k$, assuming their magnitudes denoted $V_1,...,V_K$ are estimated beforehand. This problem can be solved by minimizing the following cost function:

$$|E| = |X - \sum_k X_k|. \tag{1}$$

This function has many global minima. For instance, if $K = 2$, let us consider the two complex numbers $X_1 = V_1 e^{i\theta_1}$ and $X_2 = V_2 e^{i\theta_2}$ which are such that $X_1 + X_2 = X = V e^{i\theta}$. The numbers generated by applying the symmetry of axis X to X_1 and X_2 are $\bar{X}_1 e^{2\theta}$ and $\bar{X}_2 e^{2\theta}$, where $\bar{z}$ denotes the complex conjugate of z. Their sum is:

$$(\bar{X}_1 + \bar{X}_2)e^{2\theta} = \bar{X}e^{2\theta} = Ve^{-\theta}e^{2\theta} = X. \tag{2}$$

Thus, from a first solution to the problem, it is easy to find another one, which motivates the research of a properly-chosen solution. When $K \geq 3$, the problem has infinitely many solutions.

Wiener-like filtering [3] is one of the most commonly-used methods to obtain roots of this function. With this technique, the estimates are:

$$\hat{X}_k = \frac{V_k^2}{\sum_l V_l^2} \odot X, \tag{3}$$

and the corresponding error is:

$$|E| = |X - \sum_k \hat{X}_k| = |X| \left| 1 - \sum_k \frac{V_k^2}{\sum_l V_l^2} \right| = |X||1 - 1| = 0. \tag{4}$$

However, the Wiener filtering estimates does not verify the condition $|\hat{X}_k| = V_k$. Besides, as it is stated in the reference document [1], these estimates do not lead to satisfactorily sounding results when the sources overlap in the TF domain.

It has been shown in a previous work [2] that incorporating phase information into a source separation framework can improve the quality of the overall separation. We then propose an iterative procedure to minimize the error function $|E|$. This novel method allows us, through its initialization, to use some prior information about the phase of the components. Indeed, our goal is to get close to a local minima of the cost function using this prior knowledge.

Algorithm 1 Estimation of complex components from their mixture
---
Inputs:
Mixture $X \in \mathbb{C}$, magnitudes $V_k \in \mathbb{R}_+$, weights λ_k, and initial values $\hat{X}_k \in \mathbb{C}$, $\forall k \in [\![1; K]\!]$,
Number of iterations N_{it}.
Compute initial error $E = X - \sum_k \hat{X}_k$.
for $it = 1$ to N_{it} **do**
  for $k = 1$ to K **do**
    $Y_k \leftarrow \hat{X}_k + \lambda_k E,$
    $\hat{X}_k \leftarrow \frac{Y_k}{|Y_k|} V_k.$
  end for
  $E \leftarrow X - \sum_k \hat{X}_k.$
end for
Ouputs: $\forall k \in [\![1; K]\!]$, $\hat{X}_k \in \mathbb{C}$.

Note that alternative methods exist, such as considering the problem of minimizing $|E|$ under some phase constraints. For instance, phase evolution cost functions have been proposed in [4] and in [5]. However, as it is shown in the experimental part of the reference paper [1], the algorithm presented in this paper provides good results with a moderate computational cost. Thus, it seems to be a good candidate for a source separation task.

3 Iterative source separation procedure

3.1 Intuitive approach

We present here the intuitive technique that was introduced in the reference paper [1] for minimizing the cost function $|E|$. At iteration (it), the error $E^{(it)} = X - \sum_k \hat{X}_k^{(it)}$ is distributed over the estimates:

$$Y_k^{(it+1)} = \hat{X}_k^{(it)} + \lambda_k E^{(it)}, \tag{5}$$

where the parameters λ_k are nonnegative weights. In order to preserve the fact that $\sum_k Y_k^{(it)} = X$, the weights must verify:

$$\sum_k \lambda_k = 1. \tag{6}$$

Intuitively, we look for λ_k such that the more important the component k (i.e. the greater V_k), the greater the corresponding weight λ_k. Indeed, the components of highest energy are expected to have more impact on the estimation error than the components of lowest energy. Thus, we propose the following definition, which obviously conforms to (6):

$$\lambda_k = \frac{V_k^2}{\sum_l V_l^2}. \tag{7}$$

We then present in Algorithm 1 (labeled Algorithm 2 in [1]) the procedure for estimating the components from a complex mixture X, knowing their magnitudes.

3.2   Proof of convergence

We present here the mathematical proof of the non-increasing property of the error in this algorithm. The error at iteration $(it + 1)$ is:

$$\left|E^{(it+1)}\right| = \left|X - \sum_k \hat{X}_k^{(it+1)}\right|$$

$$= \left|\sum_k Y_k^{(it+1)} - \sum_k \hat{X}_k^{(it+1)}\right|$$

$$= \left|\sum_k Y_k^{(it+1)} - \hat{X}_k^{(it+1)}\right|$$

$$= \left|\sum_k \hat{X}_k^{(it)} + \lambda_k E^{(it)} - \frac{\hat{X}_k^{(it)} + \lambda_k E^{(it)}}{|\hat{X}_k^{(it)} + \lambda_k E^{(it)}|} V_k\right|$$

$$= \left|\sum_k (\hat{X}_k^{(it)} + \lambda_k E^{(it)}) \left(1 - \frac{V_k}{|\hat{X}_k^{(it)} + \lambda_k E^{(it)}|}\right)\right|.$$

Applying the triangle inequality, we have:

$$\left|E^{(it+1)}\right| \leq \sum_k |\hat{X}_k^{(it)} + \lambda_k E^{(it)}| \left|1 - \frac{V_k}{|\hat{X}_k^{(it)} + \lambda_k E^{(it)}|}\right|$$

$$\leq \sum_k \left||\hat{X}_k^{(it)} + \lambda_k E^{(it)}| - V_k\right|$$

$$\leq \sum_k \left||\hat{X}_k^{(it)} + \lambda_k E^{(it)}| - |\hat{X}_k^{(it)}|\right|.$$

We then use the following property (which is easy to demonstrate using the triangle inequality) : $\forall (a,b) \in \mathbb{C}^2$,

$$||a| - |b|| \leq |a - b|. \tag{8}$$

Let $a = \hat{X}_k^{(it)} + \lambda_k E^{(it)}$ and $b = \hat{X}_k^{(it)}$. We then have:

$$\left||\hat{X}_k^{(it)} + \lambda_k E^{(it)}| - |\hat{X}_k^{(it)}|\right| \leq \left|\lambda_k E^{(it)}\right|. \tag{9}$$

Using this inequality in the previous calculation, we have:

$$\left|E^{(it+1)}\right| \leq \sum_k \left|\lambda_k E^{(it)}\right| \leq \left|E^{(it)}\right| \sum_k \lambda_k. \tag{10}$$

Since the weights are chosen such that

$$\sum_k \lambda_k = 1, \tag{11}$$

we then obtain:

$$\left|E^{(it+1)}\right| \leq \left|E^{(it)}\right|. \tag{12}$$

This result traduces that the error function $|E|$ is non increasing over iterations. Since $|E^{(it)}|$ is non-increasing and minored by 0, it converges.

3.3 Auxiliary function method

A more rigorous method for obtaining the procedure described in Algorithm 1 consists in using the *auxiliary function* technique. Such technique has been used for estimating the Complex Nonnegative Matrix Factorization (CNMF) model in [6]. We provide here a full mathematical derivation of the procedure using this method.

We consider the following cost function:

$$f(\theta) = |E|^2 = |X - \sum_k \hat{X}_k|^2, \tag{13}$$

with $\theta = \{\hat{X}_k, k \in [\![1; K]\!]\}$, under the constraints $|\hat{X}_k| = V_k$ for all k. The idea is then to introduce a function $g(\theta, \tilde{\theta})$ which depends on some new parameters $\tilde{\theta}$, and verify:

$$f(\theta) = \min_{\tilde{\theta}} g(\theta, \tilde{\theta}). \tag{14}$$

Such a function is called an *auxiliary function*. It can be shown (for instance in [6]) that f is non-increasing under the following update rules:

$$\tilde{\theta} \leftarrow \arg\min_{\tilde{\theta}} g(\theta, \tilde{\theta}) \text{ and } \theta \leftarrow \arg\min_{\theta} g(\theta, \tilde{\theta}). \tag{15}$$

We propose to obtain an auxiliary function for our problem. We introduce the auxiliary variables $\tilde{\theta} = \{Y_k, k \in [\![1; K]\!]\}$ such that $\sum_k Y_k = X$. We have:

$$|X - \sum_k \hat{X}_k|^2 = |\sum_k (Y_k - \hat{X}_k)|^2. \tag{16}$$

We then introduce the nonnegative weights λ_k which verify $\sum_k \lambda_k = 1$, and we can write:

$$|X - \sum_k \hat{X}_k|^2 = \left| \sum_k \lambda_k \left(\frac{Y_k - \hat{X}_k}{\lambda_k} \right) \right|^2. \tag{17}$$

Applying the Jensen inequality to the convex function $|.|^2$, we obtain:

$$|X - \sum_k \hat{X}_k|^2 \leq \sum_k \frac{|Y_k - \hat{X}_k|^2}{\lambda_k}. \tag{18}$$

Thus, $f(\theta) \leq g(\theta, \tilde{\theta})$ with:

$$g(\theta, \tilde{\theta}) = \sum_k \frac{|Y_k - \hat{X}_k|^2}{\lambda_k}, \tag{19}$$

and the problem becomes that of minimizing g under the constraints $\sum_k Y_k = X$ and $\forall k, |\hat{X}_k| = V_k$. Let us prove that g is an auxiliary function of the objective cost function f, i.e. that it satisfies (14). To do so, we introduce the constraint on the auxiliary variables $\sum_k Y_k = X$ by means of the Lagrange multipliers:

$$\mathcal{L}(\theta, \tilde{\theta}, \gamma) = g(\theta, \tilde{\theta}) + \gamma(\sum_k \bar{Y}_k - \bar{X}). \tag{20}$$

Minimizing g with respect to $\tilde{\theta}$ under the constraint $\sum_k Y_k = X$ leads to finding a saddle point for $\mathcal{L}$. We then calculate the partial derivatives of $\mathcal{L}$ with respect to the complex variables Y_k (the so-called Wirtinger derivatives). In practice, this is computed by taking the derivative with respect to $\bar{Y}_k$ which is treated as a standard real variable, while Y_k is treated as a constant [7]. We have:

$$\frac{\partial \mathcal{L}}{\partial \bar{Y}_k}(\theta, \tilde{\theta}, \gamma) = \frac{1}{\lambda_k}(Y_k - \hat{X}_k) + \gamma, \tag{21}$$

which is set at 0 and leads to:

$$Y_k = \hat{X}_k + \lambda_k \gamma. \tag{22}$$

Besides, setting the derivative with respect to the Lagrange multiplier γ at zero leads to the constraint $\sum_k Y_k = X$. By summing (22) over k and using this constraint, we have:

$$X = \sum_k Y_k = \sum_k \hat{X}_k + \gamma \sum_k \lambda_k, \tag{23}$$

and since the weights λ_k add up to 1, we obtain:

$$\gamma = X - \sum_k \hat{X}_k, \tag{24}$$

which leads to:

$$Y_k = \hat{X}_k + \lambda_k (X - \sum_k \hat{X}_k). \tag{25}$$

Thus, $g(\theta, \tilde{\theta})$ is minimized for a set of auxiliary parameters $\tilde{\theta}_m$ defined by (25), and is then equal to:

$$\begin{aligned}
g(\theta, \tilde{\theta}_m) &= \sum_k \frac{|\hat{X}_k + \lambda_k (X - \sum_k \hat{X}_k) - \hat{X}_k|^2}{\lambda_k} \\
&= \sum_k \lambda_k |X - \sum_k \hat{X}_k|^2 \\
&= |X - \sum_k \hat{X}_k|^2 \sum_k \lambda_k \\
&= |X - \sum_k \hat{X}_k|^2 \\
&= f(\theta),
\end{aligned}$$

which shows that g is an auxiliary function of f. In accordance with (15), we obtain the update rules on θ and $\tilde{\theta}$ by alternatively minimizing g with respect to these variables. As it has already been shown, the update rule on Y_k is given by (25). To obtain the update rule on $\hat{X}_k$, we introduce the constraints $|\hat{X}_k| = V_k$, $\forall k$, by means of the Lagrange multipliers:

$$\mathcal{H}(\theta, \tilde{\theta}, \delta_1, ..., \delta_K) = g(\theta, \tilde{\theta}) + \sum_k \delta_k (|\hat{X}_k|^2 - V_k^2). \tag{26}$$

Minimizing g with respect to θ under the constraints $|\hat{X}_k| = V_k$ leads to finding a saddle point for $\mathcal{H}$. We then calculate the partial derivatives of $\mathcal{H}$ with respect to the complex variables $\hat{X}_k$:

$$\frac{\partial \mathcal{H}}{\partial \hat{X}_k}(\theta, \tilde{\theta}, \delta_1, ..., \delta_K) = \frac{1}{\lambda_k}(\hat{X}_k - Y_k) + \delta_k \hat{X}_k, \tag{27}$$

and setting this derivative at 0 leads to:

$$\hat{X}_k = \frac{Y_k}{1 + \lambda_k \delta_k}. \tag{28}$$

Besides, setting the derivatives with respect to the Lagrange multipliers δ_k at zero leads to the constraint $|X_k| = V_k$. By taking the modulus in (28) and using these constraints, we have:

$$V_k = |\hat{X}_k| = \frac{|Y_k|}{|1 + \lambda_k \delta_k|}, \tag{29}$$

so

$$1 + \lambda_k \delta_k = \pm \frac{|Y_k|}{V_k}, \tag{30}$$

6

and finally, combining this relation and (28), we have:

$$\hat{X}_k = \pm V_k \frac{Y_k}{|Y_k|}. \tag{31}$$

To avoid any ambiguity on the sign in (31), we calculate the value of g for both cases. We have:

$$|Y_k - V_k \frac{Y_k}{|Y_k|}| = ||Y_k| - V_k|, \tag{32}$$

and :

$$|Y_k + V_k \frac{Y_k}{|Y_k|}| = ||Y_k| + V_k|. \tag{33}$$

Besides, $|Y_k| \geq 0$ and $V_k \geq 0$, so we obviously have $||Y_k| - V_k| \leq ||Y_k| + V_k|$. Then, g is minimized with respect to θ when $\forall k$:

$$\hat{X}_k = V_k \frac{Y_k}{|Y_k|}. \tag{34}$$

Ultimately, the objective function f is minimized by alternatively applying the update rules (25) and (34). This leads to the iterative procedure that is summarized in Algorithm 1.

4   Initialization of the algorithm

The initialization in Algorithm 1 is crucial to ensure that the solution will be physically correct. As stated in section 2, the keystone of our approach is to properly initialize the algorithm in order to exploit prior phase knowledge for improving the quality of the separation. In addition, it contributes to significantly decrease the required number of iterations until convergence.

Intuitively, one can initialize the algorithm with the following scheme, inspired from the Wiener filtering technique: the phase of the mixture is given to each source. However, it corresponds to a fixed point of the algorithm: the components will not be modified over iterations. Indeed, with this initialization:

$$\hat{X}_k^{(0)} = V_k \frac{X}{|X|}, \tag{35}$$

and the error is equal to:

$$E^{(0)} = X - \sum_k V_k \frac{X}{|X|} = \frac{X}{|X|} \left(|X| - \sum_l V_l \right), \tag{36}$$

which leads to:

$$Y_k^{(1)} = \hat{X}_k^{(0)} + \lambda_k E^{(0)} = V_k \frac{X}{|X|} + \lambda_k \frac{X}{|X|} \left(|X| - \sum_l V_l \right) = \frac{X}{|X|} \left(V_k + \lambda_k (|X| - \sum_l V_l) \right). \tag{37}$$

Therefore, the auxiliary variables $Y_k^{(1)}$ have the seem direction than X if the quantity $V_k + \lambda_k(|X| - \sum_l V_l)$ is positive, and the opposite direction otherwise. If $|X| - \sum_l V_l \geq 0$, it is obvious that this quantity is positive, then after normalization we have :

$$\hat{X}_k^{(1)} = V_k \frac{X}{|X|} = \hat{X}_k^{(0)}. \tag{38}$$

Now, let us assume that $|X| - \sum_l V_l < 0$. Using some simple algebra, we can write:

$$\lambda_k = \frac{V_k^2}{\sum_l V_l^2} \leq \frac{V_k}{\sum_l V_l} \leq \frac{V_k}{\sum_l V_l - |X|}, \tag{39}$$

and since we assumed that $|X| - \sum_l V_l < 0$, this leads to:

$$\lambda_k (\sum_l V_l - |X|) \leq V_k, \tag{40}$$

which is equivalent to $V_k + \lambda_k(|X| - \sum_l V_l) \geq 0$ Finally, after normalization, we still have:

$$\hat{X}_k^{(1)} = V_k \frac{X}{|X|} = \hat{X}_k^{(0)}. \tag{41}$$

In other words, in both cases, the procedure does not modify the components $\hat{X}_k$ from it initial values.

For those reasons, we propose to initialize this procedure with the phase unwrapping algorithm [1]. Such an initialization is, in general, not a fixed point of the algorithm, and it leads to a fast convergence and good quality solutions from a perceptual point of view.

5 Conclusion

In this report, we have addressed the problem of source separation by seeking to minimize the cost function $|E|$. To overcome the limitation of existing non-iterative techniques, such as the Wiener filtering, we have introduced an iterative procedure, under which the cost function is non-increasing. We have pointed out the need for a non-trivial initialization of this algorithm, in order to take advantage of some prior phase information about the components. We suggest that the components estimated with the phase unwrapping algorithm is a good candidate for this task, since under this initialization, the error is expected to converge fast, and the solution should have some temporal continuity. Experimental validation is conducted in the reference paper, and show that better results than with the traditional Wiener filtering technique can be reached with this algorithm.

References

[1] P. Magron, R. Badeau, and B. David, "Model-based STFT phase recovery for audio source separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 6, pp. 1095–1105, June 2018.

[2] ——, "Phase recovery in NMF for audio source separation: an insightful benchmark," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015.

[3] A. Liutkus and R. Badeau, "Generalized Wiener filtering with fractional power spectrograms," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015.

[4] J. Bronson and P. Depalle, "Phase constrained complex NMF: Separating overlapping partials in mixtures of harmonic musical sources," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014.

[5] P. Magron, R. Badeau, and B. David, "Complex NMF under phase constraints based on signal modeling: application to audio source separation," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016.

[6] H. Kameoka, N. Ono, K. Kashino, and S. Sagayama, "Complex NMF: A new sparse representation for acoustic signals," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2009.

[7] P. Bouboulis, "Wirtinger's calculus in general hilbert spaces," *arXiv preprint arXiv:1005.5170*, 2010.