Stable Recovery with Analysis Decomposable Priors
Jalal M. Fadili, Gabriel Peyré, Samuel Vaiter, Charles-Alban Deledalle, Joseph Salmon

To cite this version:

HAL Id: hal-00926732
https://hal.archives-ouvertes.fr/hal-00926732

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Stable Recovery with Analysis Decomposable Priors

Jalal M. Fadili
GREYC
CNRS-ENSICAEN-Univ. Caen
Caen, France

Gabriel Peyré and Samuel Vaiter
CEREMADE
CNRS-Univ. Paris-Dauphine
Paris, France

Charles-Alban Deledalle
IMB
CNRS-Univ. Bordeaux 1CNRS-Télécom ParisTech
Bordeaux, France

Joseph Salmon
LTCI
Paris, France

Abstract—In this paper, we investigate in a unified way the structural properties of solutions to inverse problems. These solutions are regularized by the generic class of semi-norms defined as a decomposable norm composed with a linear operator, the so-called analysis type decomposable prior. This encompasses several well-known analysis-type regularizations such as the discrete total variation (in any dimension), analysis group-Lasso or the nuclear norm. Our main results establish sufficient conditions under which uniqueness and stability to a bounded measurement error that comes within a factor of the noise level $\epsilon$. This goes beyond [1] who considered identifiability under a generalized irrepresentable condition in the noiseless case with $L = \text{Id}$, $\ell_2$-stability for a class of decomposable priors closely related to Definition 1, is also studied in [8] for $L = \text{Id}$ and general sufficiently smooth data fidelity. Their stability results require however stronger assumptions than ours (typically a restricted strong convexity which becomes a type of restricted eigenvalue property for linear regression with quadratic data fidelity). The authors in [3] provide sharp estimates of the number of generic measurements required for exact and $\ell_2$-stable recovery of models from random partial information by solving a constrained form of (1) regularized by atomic norms. This is however restricted to the compressed sensing scenario. Our results generalize the stability guarantee of [7] established when the decomposable norm is $\ell_1$ and $L^*$ is the analysis operator of a frame. A stability result for general sublinear functions $R$ is given in [6]. The stability is however measured in terms of $R$, and $\ell_2$-stability can only be obtained if $R$ is coercive, i.e., $L^*$ is injective.

At this stage, we would like to point out that although we carry out our analysis on the penalized form (1), our results remain valid for the data fidelity constrained version but obviously with different constants in the bounds. We omit these results for obvious space limitations.

I. INTRODUCTION

A. Problem statement

Suppose we observe
\[ y = \Phi x_0 + w, \quad \text{where} \quad \|w\|_2 \leq \epsilon, \]
where $\Phi$ is a linear operator from $\mathbb{R}^N$ to $\mathbb{R}^M$ that may have a non-trivial kernel. We want to robustly recover an approximation of $x_0$ by solving the optimization problem
\[ x^* \in \text{Argmin}_{x \in \mathbb{R}^N} \frac{1}{2} \|y - \Phi x\|_2^2 + \lambda R(x), \quad (1) \]
where
\[ R(x) := \|L^* x\|_A, \]
with $L : \mathbb{R}^P \to \mathbb{R}^N$ a linear operator, and $\| \cdot \|_A : \mathbb{R}^P \to \mathbb{R}^+$ is a decomposable norm in the sense of [1]. Decomposable regularizers are intended to promote solutions conforming to some notion of simplicity/low complexity that complies with that of $x_0 = L^* x_0$. This motivates the following definition of these norms. Throughout the paper, given a subspace $V \subset \mathbb{R}^P$, we will use the shorthand notation $L_V = LP_V$, $L^*_V = P_V L^*$, and $\alpha_V = P_V \alpha$ for any vector $\alpha \in \mathbb{R}^P$, where $P_V$ (resp. $P_{V^\perp}$) is the orthogonal projector on $V$ (resp. on its orthogonal complement $V^\perp$).

Definition 1. A norm $\| \cdot \|_A$ is decomposable at $u \in \mathbb{R}^P$ if:

(i) there is a subspace $T \subset \mathbb{R}^P$ and a vector $e \in T$ such that
\[ \partial \| \cdot \|_A(u) = \{ \alpha \in \mathbb{R}^P \mid \alpha_T = e \quad \text{and} \quad \|\alpha_{T^\perp}\|_A \leq 1 \} \]
(ii) and for any $z \in T^\perp$, $\|z\|_A = \sup_{v \in T} \|v\|_A \leq 1 \langle v, z \rangle$, where $\| \cdot \|_A$ is the dual norm of $\| \cdot \|_A$.

From this definition, it can be easily proved, using Fenchel identity, that $u \in T$ whenever $\| \cdot \|_A$ is decomposable at $u$. Popular examples covered by decomposable regularizers are the $\ell_1$-norm, the $\ell_1-\ell_2$ group sparsity norm, and the nuclear norm [1].

B. Contributions and relation to prior work

In this paper, we give a strong sufficient condition under which (1) admits a unique minimizer. From this, sufficient uniqueness conditions are derived. Then we develop results guaranteeing a stable approximation of $x_0$ from the noisy measurements $y$ by solving (1), with an $\ell_2$-error that comes within a factor of the noise level $\epsilon$. This goes beyond [1] who considered identifiability under a generalized irrepresentable condition in the noiseless case with $L = \text{Id}$, $\ell_2$-stability for a class of decomposable priors closely related to Definition 1, is also studied in [8] for $L = \text{Id}$ and general sufficiently smooth data fidelity. Their stability results require however stronger assumptions than ours (typically a restricted strong convexity which becomes a type of restricted eigenvalue property for linear regression with quadratic data fidelity). The authors in [3] provide sharp estimates of the number of generic measurements required for exact and $\ell_2$-stable recovery of models from random partial information by solving a constrained form of (1) regularized by atomic norms. This is however restricted to the compressed sensing scenario. Our results generalize the stability guarantee of [7] established when the decomposable norm is $\ell_1$ and $L^*$ is the analysis operator of a frame. A stability result for general sublinear functions $R$ is given in [6]. The stability is however measured in terms of $R$, and $\ell_2$-stability can only be obtained if $R$ is coercive, i.e., $L^*$ is injective.

At this stage, we would like to point out that although we carry out our analysis on the penalized form (1), our results remain valid for the data fidelity constrained version but obviously with different constants in the bounds. We omit these results for obvious space limitations.

II. UNIQUENESS

A. Main assumptions

We first note that traditional coercivity and convexity arguments allow to show that the set of (global) minimizers of (1) is a non-empty compact set if, and only if, $\ker(\Phi) \cap \ker(L^*) = \{0\}$.
The following assumptions will play a pivotal role in our analysis.

**Assumption (SC$_c$)** There exist $\eta \in \mathbb{R}^M$ and $\alpha \in \partial |\cdot|_A(L^*x)$ such that the following so-called source (or range) condition is verified:

$$\Phi^* \eta = L\alpha \in \partial R(x).$$

**Assumption (INJ$_T$)** For a subspace $T \subset \mathbb{R}^p$, $\Phi$ is injective on $\ker(L_{T^\perp}^*)$.

It is immediate to see that since $\ker(L^*) \subseteq \ker(L_{T^\perp}^*)$, (INJ$_T$) implies that the set of minimizers is indeed non-empty and compact.

**B. Strong Null Space Property**

We shall now give a novel strong sufficient uniqueness condition under which problem (1) admits exactly one minimizer.

**Theorem 1.** For a minimizer $x^*$ of (1), let $T$ and $e$ be the subspace and vector in Definition 1 associated to $u^* = L^*x^*$, and denote $S = T^\perp$. $x^*$ is the unique minimizer of (1) if

$$(L_{T^\perp}^*h, e) < |L_{T^\perp}^*h|_A, \quad \forall h \in \ker(\Phi) \setminus \{0\}.$$ 

The above condition is a strong generalization of the Null Space Property well known in $l_1$ regularization [4].

**C. Sufficient uniqueness conditions**

**1) General case:** A direct consequence of the above theorem is the following corollary.

**Corollary 1.** For a minimizer $x^*$ of (1), let $T$ and $e$ be the subspace and vector in Definition 1 associated to $u^* = L^*x^*$, and denote $S = T^\perp$. Assume that (SC$_{x^*}$) is verified with $\|\alpha|_A < 1$, and that (INJ$_T$) holds. Then, $x^*$ is the unique minimizer of (1).

In fact, it turns out that the above two results are proved without requiring some restrictive implications of Definition 1(ii) of decomposable norms, and are therefore valid for a much larger class of regularizations. This can be clearly checked in the arguments used in the proofs.

**2) Separable case:**

**Definition 2.** The decomposable norm $\|\cdot\|_A$ is separable on the subspace $T^\perp = S = V \oplus W \subset \mathbb{R}^p$ if for any $u \in \mathbb{R}^p$, $\|u_{T^\perp}\|_A = \|uv\|_A + \|uw\|_A$.

Separability as just defined is fulfilled for several decomposable norms such as the $\ell_1$ or $\ell_{1-p}$ norms, $1 \leq p < +\infty$.

The non-saturation condition on the dual certificate required in Corollary 1 can be weakened to hold only on a subspace $V \subset S$ and the conclusions of the corollary remain valid, and assuming a stronger restricted injectivity assumption. We have the following corollary.

**Corollary 2.** Assume that $\|\cdot\|_A$ is also separable, with $S = V \oplus W$, such that (SC$_{x^*}$) is verified with $\|\alpha|_A < 1$, and (INJ$_V$) holds. Then, $x^*$ is the unique minimizer of (1).

**III. Stability to noise**

**A. Main result**

**1) General case:** We are now ready to state our main stability results.

**Theorem 2.** Let $T_0$ and $e_0$ be the subspace and vector in Definition 1 associated to $u_0 = L^*x_0$, and denote $S_0 = T_0^\perp$. Assume that (SC$_{x_0}$) is verified with $\|\alpha|_A < 1$, and that (INJ$_{T_0}$) holds. Then, choosing $\lambda = \varepsilon e$, $\varepsilon > 0$, the following holds for any minimizer $x^*$ of (1)

$$\|x^* - x_0\|_2 \leq C\varepsilon,$$

where $C = C_1 (2 + \varepsilon \|\eta\|_2) + C_2 \|\alpha\|_A^2 + C_3 \|\gamma\|_2^2$, and $C_1 > 0$ and $C_2 > 0$ are constants independent of $\eta$ and $\alpha$.

**Remark 1** (Separable case). When the decomposable norm is also separable (see Corollary 2), the stability result of Theorem 2 remains true assuming that $\|\alpha\|_A < 1$ for $V \subset S_0$. This however comes at the price of the stronger restricted injectivity assumption (INJ$_V$). To show this, the only thing to modify is the statement and the proof of Lemma 2 which can be done easily using similar arguments to those in the proof of Corollary 2.

**2) Case of frames:** Suppose that $L^*$ is the analysis operator of a frame (ker$(L^*) = \{0\}$) with lower bound $a > 0$, let $L$ be a dual frame. The following stability bound can be obtained whose proof is omitted for space limitations.

**Proposition 1.** Let $T_0$ and $e_0$ be the subspace and vector in Definition 1 associated to $u_0 = L^*x_0$, and denote $S_0 = T_0^\perp$. Assume that (SC$_{x_0}$) is verified with $\|\alpha|_A < 1$, and that $\Phi$ is injective on $\mathrm{Im}(\tilde{L}_{T_0})$. Then, choosing $\lambda = \varepsilon e$, $\varepsilon > 0$, the following holds for any minimizer $x^*$ of (1)

$$\|x^* - x_0\|_2 \leq C\varepsilon.$$

where $C = C_1 (2 + \varepsilon \|\eta\|_2) + C_2 \|\alpha\|_A^2 + C_3 \|\gamma\|_2^2$, and $C_1 > 0$ and $C_2 > 0$ are constants independent of $\eta$ and $\alpha$.

Since ker$(L_{T_0}^*) \subseteq \mathrm{Im}(\tilde{L}_{T_0})$, the required restricted injectivity assumption is more stringent than (INJ$_{T_0}$). On the positive side, the constant $C_2$ is in general better than $C_2$. More precisely, the constant $C_1$, see the proof of Theorem 2, is replaced with $\sqrt{a}$. Note also that coercivity of $R$ in this case allows to derive a bound similar to ours from the results in [6]. His restricted injectivity assumption is however different and our constants are sharper.

**B. Generalized irrepresentable condition**

In the following corollary, we provide a stronger sufficient stability condition that can be viewed as a generalization of the irrepresentable condition introduced in [5] when $R$ is the $\ell_1$ norm. It allows to construct dual vectors $\eta$ and $\alpha$ which obey the source condition and are computable, which in turn yield explicit constants in the bound.
Definition 3. Let \( T \subset \mathbb{R}^P \) and \( e \in \mathbb{R}^P \), and denote \( S = T^\perp \). Suppose that \((\text{INJ}_T)\) is verified. Define for any \( u \in \ker(L_S) \) and \( z \in \mathbb{R}^M \) such that \( \Phi^* z \in \text{Im}(L_S) \)

\[
\text{IC}_{u,z}(T,e) = \|\Gamma e + u_S + (L_S)^+ \Phi^* z\|_A
\]

where
\[
\Gamma = (L_S)^+ (\Phi^* \Phi \Xi - \text{Id}) L_T \theta,
\]

\( \Xi : h \mapsto \Xi h = \arg\min_{x \in \ker(L_S)} \frac{1}{2} \|\Phi x\|_2^2 - \langle h, x \rangle \),

and \( M^+ \) is the Moore-Penrose pseudoinverse of \( M \). Let \( \bar{u}, \bar{z} \) and \( u \) defined as

\[
(\bar{u}, \bar{z}) = \arg\min_{u \in \ker(L_S), \{z \mid \Phi^* z \in \text{Im}(L_S)\}} \text{IC}_{u,z}(T,e)
\]

and \( u = \arg\min_{u \in \ker(L_S)} \text{IC}_{u,0}(T,e) \).

Obviously, we have

\[
\text{IC}_{u,z}(T,e) \leq \text{IC}_{u,0}(T,e) = \text{IC}_{0,0}(T,e).
\]

The convex programs defining \( \text{IC}_{u,z}(T,e) \) and \( \text{IC}_{u,0}(T,e) \) can be solved using primal-dual proximal splitting algorithms whenever the proximity operator of \( \| \cdot \|_A \) can be easily computed [2]. The criterion \( \text{IC}_{u,0}(T,e) \) specializes to the one developed in [10] when \( \| \cdot \|_A \) is the \( \ell_1 \) norm. \( \text{IC}_{0,0}(T,e) \) is a generalization of the coefficient involved in the irrepresentable condition introduced in [5] when \( R \) is the \( \ell_1 \) norm, and to the one in [1] for decomposable priors with \( L = \text{Id} \).

Corollary 3. Assume that \((\text{INJ}_T)\) is verified and \( \text{IC}_{u,z}(T_0,e_0) < 1 \). Then, taking \( \eta = \Phi \Xi L_{T_0} e_0 + \bar{z} \), one can construct \( \alpha \) such that \((\text{SC}_{e_0})\) is satisfied and \( \|\alpha_{S_0}\|_A < 1 \). Moreover, the conclusion of Theorem 2 remains true substituting \( 1 - \text{IC}_{u,z}(T_0,e_0) \) for \( 1 - \|\alpha_{S_0}\|_A \).

IV. PROOFS

A. Proof of Theorem 1

A key observation is that by strong (hence strict) convexity of \( x \mapsto \|y - \Phi x\|_2^2 \), all minimizers of (1) share the same image under \( \Phi \). Therefore any minimizer of (1) takes the form \( x^* + h \) where \( h \in \ker(\Phi) \). Furthermore, it can be shown by arguments from convex analysis that any proper convex function \( R \) has a unique minimizer \( x^* \) (if any) over a convex set \( C \) if its directional derivative satisfies

\[
R'(x^*; x - x^*) > 0, \quad x \in C, x \neq x^*.
\]

Applying this to (1) with \( C = x^* + \ker(\Phi) \), and using the fact that the directional derivative is the support function of the subdifferential, we get that \( x^* \) is the unique minimizer of (1) if, and only if, \( \forall h \in \ker(\Phi) \setminus \{0\} \),

\[
0 < R'(x^*; h) = \sup_{v \in \partial R(x^*)} \langle v, h \rangle = \sup_{\alpha \in \partial \|L^* h\|_A} \langle \alpha, L^* h \rangle = \langle e, L^* h \rangle + \sup_{\|\alpha_S\|_A \leq 1} \langle \alpha_S, L_S h \rangle \leq \langle e, L^* h \rangle + \|L_S h\|_A.
\]

We conclude using symmetry of the norm and the fact that \( \ker(\Phi) \) is a subspace.

B. Proof of Corollary 1

The source condition \((\text{SC}_{x^*})\) implies that \( \forall h \in \ker(\Phi) \setminus \{0\} \),

\[
\langle h, L h \rangle = \langle \Phi^* \eta, h \rangle = \langle \Phi h, \eta \rangle = 0.
\]

Moreover

\[
\langle h, L h \rangle = \langle L^* h, \alpha \rangle = \langle \Phi^* h, \alpha \rangle = \langle L^* h, \alpha S \rangle.
\]

Thus, applying the dual-norm inequality we get

\[
\|L^* h\|_A \leq \|L^* h\|_A \|\alpha_S\|_A < \|L^* h\|_A,
\]

where the last inequality is strict since \( L_S^* h \) does not vanish owing to \((\text{INJ}_T)\), and \( \|\alpha_{S_0}\|_A < 1 \).

C. Proof of Corollary 2

We follow the same lines as the proof of Corollary 1 and get

\[
\langle L^* h, \alpha \rangle = \langle L^* h, \alpha S \rangle + \langle L^* h, \alpha V \rangle + \langle L^* h, \alpha W \rangle.
\]

We therefore obtain

\[
\|L^* h\|_A \leq \|L^* h\|_A \|\alpha_V\|_A + \|\alpha_W\|_A \|L^* h\|_A + \|\alpha_W\|_A \|L^* h\|_A = \|L^* h\|_A,
\]

where we used that \( h \not\in \ker(L^*_T) \), \( \|\alpha_{V_0}\|_A < 1 \), separability and \( \|\alpha_W\|_A \leq \|\alpha_V\|_A + \|\alpha_W\|_A = \|\alpha_{S_0}\|_A \leq 1 \).

D. Proof of Theorem 2

We first define the Bregman distance/divergence.

Definition 4. Let \( D_B^R(x,x_0) \) be the Bregman distance associated to \( R \) with respect to \( s \in \partial R(x_0) \),

\[
D_B^R(x,x_0) = R(x) - R(x_0) - \langle s, x - x_0 \rangle.
\]

Define \( D_B^A(u,u_0) = R(x) - R(x_0) - \langle s, x - x_0 \rangle \) as the Bregman distance associated to \( \| \cdot \|_A \) with respect to \( \alpha \in \partial A \) \( \| \cdot \|_A(u_0) \).

Observe that by convexity, the Bregman distance is non-negative.

Preparatory lemmas

We first need the following key lemmas.

Lemma 1 (Prediction error and Bregman distance convergence rates). Suppose that \((\text{SC}_{x^*})\) is satisfied. Then, for any minimizer \( x^* \) of (1), and with \( \lambda = \varepsilon \) for \( c > 0 \), we have

\[
D_{\Phi^* \eta}^R(x^*,x_0) = D_{\Phi^* \eta}^R(L^* x^*,L^* x_0) \leq \frac{\varepsilon (1+c\|\eta\|_2/2)}{c},
\]

\[
\|\Phi^* x - \Phi^* x_0\|_2 \leq \varepsilon (2+\|\eta\|_2).\]

The proof follows the same lines as that for any sublinear regularizer, see e.g. [9], where we additionally use the source condition \((\text{SC}_{x_0})\) and \( D_{\Phi^* \eta}^R(x,x_0) = D_{\Phi^* \eta}^R(L^* x,L^* x_0) \).

Now since \( \| \cdot \|_A \) is a norm, it is coercive, and thus

\[
\exists C_A > 0 \quad \text{s.t.} \quad \forall x \in \mathbb{R}^P, \|x\|_A \geq C_A |x|_2.
\]
We get the following inequality.

**Lemma 2** (From Bregman to $\ell_2$ bound). Suppose that $(\mathcal{SC}_{\alpha})$ holds with $\|\alpha_{S_0}\|_A < 1$. Then,

$$\|L_{S_0}^*(x^* - x_0)\|_2 \leq \frac{D_A^\alpha(L\star x^*, L\star x_0)}{C_A(1 - \|\alpha_{S_0}\|_A)}.$$

**Proof:** Decomposability of $L$ implies that $\exists v \in S_0$ such that $\|v\|_A^2 \leq 1\text{ and } \langle L_{S_0}^*(x^* - x_0), v \rangle = 0$. Moreover, $v + c_0 \in \partial L(L\star x_0)$. Thus

$$D_A^\alpha(L\star x^*, L\star x_0) \geq D_A^\alpha(L\star x^*, L\star x_0) - D_A^\alpha(L\star x^*, L\star x_0)$$

$$= \langle v + c_0 - \alpha, L(x^* - x_0) \rangle$$

$$= \langle v - \alpha_{S_0}, L_{S_0}^*(x^* - x_0) \rangle$$

$$= \|L_{S_0}^*(x^* - x_0)\|_A$$

$$\geq \|v\|_A^2 \leq 1\text{ and } \|\alpha_{S_0}\|_A < 1,$$

we get the following inequality.

Now, we observe by definition of $\Xi$ that $P_{\ker(L_{S_0}^*)}(\Phi^*\Phi\Xi - \text{Id})L_{T_0} = 0$, which implies that $\|\Phi^*\Phi\Xi - \text{Id}\|_{L_{T_0}} = \|P_{\ker(L_{S_0}^*)}(\Phi^*\Phi\Xi - \text{Id})L_{T_0}\|$. In turn, $L_{T_0} = L_{S_0}(L_{S_0}^*)^\dagger(\Phi^*\Phi\Xi - \text{Id})L_{T_0} = P_{\ker(L_{S_0}^*)}(\Phi^*\Phi\Xi - \text{Id})L_{T_0}$. This, together with the fact that $\hat{u} \in \ker(L_{S_0})$ and $\Phi^*\Xi \in \Im(L_{S_0})$ yields

$$L_{S_0}\alpha = (\Phi^*\Phi\Xi - \text{Id})L_{T_0}c_0 + \Phi^*\Xi \hat{u} = \Phi^*\Xi - \text{Id}\Leftrightarrow \Phi^*\Xi = L\alpha,$$

which implies that $\Phi^*\Xi = L\alpha \in \partial R(x_0)$. We have just shown that the vectors $\alpha$ and $\eta$ as given above satisfy the source condition $(\mathcal{SC}_{\alpha})$ and the dual non-saturation condition. We conclude by applying Theorem 2 using $(\text{INJ}_{T_0})$.

**V. CONCLUSION**

We provided a unified analysis of the structural properties of regularized solutions to linear inverse problems through a class of semi-norms formed by composing decomposable norms with a linear operator. We provided conditions that guarantee uniqueness, and also those ensuring stability to bounded noise. The stability bound was achieved without requiring (even partial) recovery of $T_0$ and $c_0$. Recovery of $T_0$ and $c_0$ for analysis-type decomposable priors and beyond is currently under investigation. Another perspective concerns whether the $\ell_2$ bound on $x^* - x_0$ can be extended to cover more general low-complexity-inducing regularizers beyond decomposable norms.

**REFERENCES**


