Description of the minimizers of least squares regularized with $\ell_0$-norm.
Uniqueness of the global minimizer

Mila Nikolova*

Abstract

We have an $M \times N$ real-valued arbitrary matrix $A$ (e.g., a dictionary) with $M < N$ and data $d$ describing the sought-after object with the help of $A$. This work provides an in-depth analysis of the (local and global) minimizers of an objective function $F_d$ combining a quadratic data-fidelity term and an $\ell_0$ penalty applied to each entry of the sought after solution, weighted by a regularization parameter $\beta > 0$. For several decades, this objective focuses a ceaseless effort to conceive algorithms approaching a good minimizer. Our theoretical contributions, summarized below, shed new light on the existing algorithms and can help the conception of innovative numerical schemes.

To solve the normal equation associated with any $M$-row submatrix of $A$ is equivalent to compute a local minimizer $\hat{u}$ of $F_d$. (Local) minimizers $\hat{u}$ of $F_d$ are strict if and only if the submatrix, composed of those columns of $A$ whose indexes form the support of $\hat{u}$, has full column rank. An outcome is that strict local minimizers of $F_d$ are easily computed without knowing the value of $\beta$. Each strict local minimizer is linear in data. It is proved that $F_d$ has global minimizers and that they are always strict. They are studied in more details under the (standard) assumption that rank($A$) = $M < N$. The global minimizers with $M$-length support are seen to be impractical. Given $d$, critical values $\beta_K$ for any $K \leq M - 1$ are exhibited such that if $\beta > \beta_K$, all global minimizers of $F_d$ are $K$-sparse.

An assumption on $A$ is adopted and proved to fail only on a closed negligible subset. Then for all data $d$ beyond a closed negligible subset, the objective $F_d$ for $\beta > \beta_K$, $K \leq M - 1$, has a unique global minimizer and this minimizer is $K$-sparse.

Instructive small-size ($5 \times 10$) numerical illustrations confirm the main theoretical results.

Keywords: asymptotically level stable functions, global minimizers, local minimizers, $\ell_0$ regularization, nonconvex nonsmooth minimization, perturbation analysis, quadratic programming, solution analysis, sparse recovery, strict minimizers, underdetermined linear systems, uniqueness of the solution, variational methods


DOI. 10.1137/11085476X

1 Introduction

Let $A$ be an arbitrary matrix (e.g., a dictionary) such that

$$A \in \mathbb{R}^{M \times N} \quad \text{for} \quad M < N,$$

where the positive integers $M$ and $N$ are fixed. Given a data vector $d \in \mathbb{R}^M$, we consider an objective function $F_d : \mathbb{R}^N \to \mathbb{R}$ of the form

$$F_d(u) = \|Au - d\|_2^2 + \beta \|u\|_0, \quad \beta > 0,$$

$$\|u\|_0 = \#\sigma(u),$$

---

*CMLA, ENS Cachan, CNRS, 61 Av. Prés. Wilson, 94230 Cachan, France (nikolova@cmla.ens-cachan.fr).
*Received by the editors November 10, 2011; accepted for publication (in revised form) January 14, 2013;
where \( u \in \mathbb{R}^N \) contains the coefficients describing the sought-after object, \( \beta \) is a regularization parameter, \( \sharp \) stands for cardinality and \( \sigma(u) \) is the support of \( u \) (i.e., the set of all \( i \in \{1, \ldots, N\} \) for which the \( i \)th entry of \( u \) satisfies \( u[i] \neq 0 \)). By an abuse of language, the penalty in (1) is called the \( \ell_0 \)-norm. Define \( \phi : \mathbb{R} \to \mathbb{R} \) by

\[
\phi(t) \overset{\text{def}}{=} \begin{cases} 
0 & \text{if } t = 0, \\
1 & \text{if } t \neq 0.
\end{cases}
\]

Then \( \|u\|_0 = \sum_{i=1}^{N} \phi(u[i]) = \sum_{i \in \sigma(u)} \phi(u[i]), \) so \( \mathcal{F}_d \) in (1) equivalently reads

\[
\mathcal{F}_d(u) = \|Au - d\|_2^2 + \beta \sum_{i=1}^{N} \phi(u[i]) = \|Au - d\|_2^2 + \beta \sum_{i \in \sigma(u)} \phi(u[i]).
\]

We focus on all (local and global) minimizers \( \hat{u} \) of an objective \( \mathcal{F}_d \) of the form (1):

\[
\hat{u} \in \mathbb{R}^N \text{ such that } \mathcal{F}_d(\hat{u}) = \min_{u \in \mathcal{O}} \mathcal{F}_d(u),
\]

where \( \mathcal{O} \) is an open neighborhood of \( \hat{u} \). We note that finding a global minimizer of \( \mathcal{F}_d \) must be an NP-hard computational problem [11, 40].

The function \( \phi \) in (2) served as a regularizer for a long time. In the context of Markov random fields it was used by Geman and Geman in 1984 [20] and Besag in 1986 [5] as a prior in MAP energies to restore labeled images. The MAP objective reads as

\[
\mathcal{F}_d(u) = \|Au - d\|_2^2 + \beta \sum_{k} \phi(D_k u),
\]

where \( D_k \) is a finite difference operator and \( \phi \) is given by (2). This label-designed form is known as the Potts prior model, or as the multi-level logistic model [6, 24]. Various stochastic and deterministic algorithms have been considered to minimize (3). Leclerc [23] introduced the stochastic continuation approach and successfully used it to reconstruct 3D tomographic images. Robini and Magnin refined the method and the theory in [34]. Very recently, Robini and Reissman [35] gave theoretical results relating the probability for global convergence and the computation speed.

The problem stated in (1) and (4)—to (locally) minimize \( \mathcal{F}_d \)—arises when sparse solutions are desired. Typical application fields are signal and image processing, morphologic component analysis, compression, dictionary building, inverse problems, compressive sensing, machine learning, model selection, classification, and subset selection, among others. The original hard-thresholding method proposed by Donoho and Johnstone [15] amounts to minimizing \( \mathcal{F}_d \), where \( d \) contains the coefficients of a signal or an image expanded in a wavelet basis \( (M = N) \). When \( M < N \), various (usually strong) restrictions on \( \|u\|_0 \) (often \( \|u\|_0 \) is replaced by a less irregular function) and on \( A \) (e.g., RIP-like criteria, conditions on \( \|A\| \), etc.) are needed to conceive numerical schemes approximating a minimizer of \( \mathcal{F}_d \), to establish local convergence and derive the asymptotic of the obtained solution. In statistics the problem has been widely considered for subset selection, and numerous algorithms have been designed, with limited theoretical production, as explained in the book by Miller [30]. More recently, Haupt and Nowak [22] investigate the statistical performances of the global minimizer of \( \mathcal{F}_d \) and propose an iterative bound-optimization

\footnote{As a reminder, if \( d \) are some noisy coefficients, the restored coefficients \( \hat{u} \) minimize \( \|u - d\|_2^2 + \beta \|u\|_0 \) and read \( \hat{u}[i] = 0 \) if \( |d[i]| \leq \sqrt{\beta} \) and \( \hat{u}[i] = d[i] \) otherwise.}
procedure. Fan and Li \cite{FanLi2001} discuss a variable splitting and penalty decomposition minimization technique for (1), along with other approximations of the $\ell_0$-norm. Liu and Wu \cite{LiuWu2010} mix the $\ell_0$ and the $\ell_1$ penalties, establish some asymptotic properties of the new estimator and use mixed integer programming aimed at global minimization. For model selection, Lv and Fan \cite{LvFan2012} approximate the $\ell_0$ penalty using functions that are concave on $\mathbb{R}_+$ and prove a nonasymptotic nearly oracle property of the resultant estimator. Thiao, Dinh, and Thi \cite{Thiao2010} reformulate the problem so that an approximate solution can be found using difference-of-convex-functions programming. Blumensath and Davies \cite{BlumensathDavies2008} propose an iterative thresholding scheme to approximate a solution and prove convergence to a local minimizer of $F_d$. Lu and Zhang \cite{LuZhang2012} suggest a penalty decomposition method to minimize $F_d$. Fornasier and Ward \cite{FornasierWard2011} propose an iterative thresholding algorithm for minimizing $F_d$ where $\ell_0$ is replaced by a reasonable sparsity-promoting relaxation given by $\phi(t) = \min\{|t|, 1\}$; then convergence to a local minimizer is established. In a recent paper by Chouzenoux et al. \cite{Chouzenoux2013}, a mixed $\ell_2 - \ell_0$ regularization is considered: a slightly smoothed version of the objective is analyzed and a majorize-minimize subspace approach, satisfying a finite length property, converges to a local minimizer. Since the submission of our paper, image reconstruction methods have been designed where $\ell_0$ regularization is applied to the coefficients of the expansion of the sought-after image in a wavelet frame \cite{Chouzenoux2013, Elad2006}: the provided numerical results outperform $\ell_1$ regularization for a reasonable computational cost achieved using penalty decomposition techniques. In a general study on the convergence of descent methods for nonconvex objectives, Attouch, Bolte, and Svaiter \cite{Attouch} apply an inexact forward-backward splitting scheme to find a critical point of $F_d$. Several other references can be evoked, e.g., \cite{Bolte2010, Attouch2010}.

Even though overlooked for several decades, the objective $F_d$ was essentially considered from a numerical standpoint. The motivation naturally comes from the promising applications and the intrinsic difficulty of minimizing $F_d$.

The goal of this work is to analyze the (local and global) minimizers $\hat{u}$ of objectives $F_d$ of the form (1).

- **We provide detailed results on the minimization problem.**
- **The uniqueness of the global minimizer of $F_d$ is examined as well.**

We do not propose an algorithm. However, our theoretical results raise salient questions about the existing algorithms and can help the conception of innovative numerical schemes.

The minimization of $F_d$ in (1) might seem close to its constraint variants:

\begin{align}
\text{given } & \varepsilon \geq 0, \quad \text{minimize } \|u\|_0 \quad \text{subject to } \|Au - d\|_2^2 \leq \varepsilon, \\
\text{given } & K \in \mathbb{I}_M, \quad \text{minimize } \|Au - d\|_2^2 \quad \text{subject to } \|u\|_0 \leq K.
\end{align}

The latter problems are abundantly studied in the context of sparse recovery in different fields. An excellent account is given in \cite{.elad2010}, see also the book \cite{candes2008}. For recent achievements, we refer the reader to \cite{Yin2010}. It is worth emphasizing that in general, there is no equivalence between the problems stated in (1) and the minimization of $F_d$ in (1) because all of these problems are nonconvex.

### 1.1 Main notations and definitions

We recall that if $\hat{u}$ is a (local) minimizer of $F_d$, the value $F_d(\hat{u})$ is a (local) minimum\footnote{These two terms are often confused in the literature.} of $F_d$ reached at (possibly numerous) points $\hat{u}$. Saying that a (local) minimizer $\hat{u}$ is strict means that there is a neighborhood $\mathcal{O} \subset \mathbb{R}^N$, containing $\hat{u}$, such that $F_d(\hat{u}) < F_d(v)$ for any $v \in \mathcal{O} \setminus \{\hat{u}\}$. So $\hat{u}$ is an isolated minimizer.
Let $K$ be any positive integer. The expression $\{u \in \mathbb{R}^K : u \text{ satisfying property } \mathcal{P}\}$ designates the subset of $\mathbb{R}^K$ formed from all elements $u$ that meet $\mathcal{P}$. The identity operator on $\mathbb{R}^K$ is denoted by $I_K$. The entries of a vector $u \in \mathbb{R}^K$ read as $u[i]$, for any $i$. The $i$th vector of the canonical basis of $\mathbb{R}^K$ is denoted by $e_i \in \mathbb{R}^K$. Given $u \in \mathbb{R}^K$ and $\rho > 0$, the open ball at $u$ of radius $\rho$ with respect to the $\ell_p$-norm for $1 \leq p \leq \infty$ reads as

$$B_p(u, \rho) \overset{\text{def}}{=} \{v \in \mathbb{R}^K : \|v - u\|_p < \rho\}.$$ 

To simplify the notation, the $\ell_2$-norm is systematically denoted by

$$\| \cdot \| \overset{\text{def}}{=} \| \cdot \|_2.$$ 

We denote by $\mathbb{I}_K$ the totally and strictly ordered index set

$$\mathbb{I}_K \overset{\text{def}}{=} (\{1, \ldots, K\}, <),$$

where the symbol $<$ stands for the natural order of the positive integers. Accordingly, any subset $\omega \subseteq \mathbb{I}_N$ inherits the property of being totally and strictly ordered.

We shall often consider the index set $\mathbb{I}_N$. The complement of $\omega \subseteq \mathbb{I}_N$ in $\mathbb{I}_N$ is denoted by

$$\omega^c = \mathbb{I}_N \setminus \omega \subseteq \mathbb{I}_N.$$

**Definition 1** For any $u \in \mathbb{R}^N$, the support $\sigma(u)$ of $u$ is defined by

$$\sigma(u) = \left\{ i \in \mathbb{I}_N : u[i] \neq 0 \right\} \subseteq \mathbb{I}_N.$$

If $u = 0$, clearly $\sigma(u) = \emptyset$.

The $i$th column in a matrix $A \in \mathbb{R}^{M \times N}$ is denoted by $a_i$. It is systematically assumed that

$$a_i \neq 0 \quad \forall i \in \mathbb{I}_N.$$ 

For a matrix $A \in \mathbb{R}^{M \times N}$ and a vector $u \in \mathbb{R}^N$, with any $\omega \subseteq \mathbb{I}_N$, we associate the submatrix $A_\omega$ and the subvector $u_\omega$ given by

$$A_\omega \overset{\text{def}}{=} (a_{\omega[1]}, \ldots, a_{\omega[|\omega|]}) \in \mathbb{R}^{M \times |\omega|},$$

$$u_\omega \overset{\text{def}}{=} (u[\omega[1]], \ldots, u[\omega[|\omega|]]) \in \mathbb{R}^{|\omega|},$$

respectively, as well as the zero padding operator $Z_\omega : \mathbb{R}^{|\omega|} \to \mathbb{R}^N$ that inverts (9):

$$u = Z_\omega(u_\omega), \quad u[i] = \begin{cases} 0 & \text{if } i \notin \omega, \\ u_\omega[k] & \text{for the unique } k \text{ such that } \omega[k] = i. \end{cases}$$

Thus for $\omega = \emptyset$ one finds $u_{\emptyset} = \emptyset$ and $u = Z_\emptyset(u_{\emptyset}) = 0 \in \mathbb{R}^N$.

Using Definition 1 and the notation in (8)-(9), for any $u \in \mathbb{R}^N \setminus \{0\}$ we have

$$\omega \in \mathbb{I}_N \quad \text{and} \quad \omega \supseteq \sigma(u) \Rightarrow A u = A_\omega u_\omega.$$ 

---

3 More precisely, for any $i \in \mathbb{I}_K$, the vector $e_i \in \mathbb{R}^K$ is defined by $e_i[i] = 1$ and $e_i[j] = 0$, $\forall j \in \mathbb{I}_K \setminus \{i\}$.

4 E.g. without strict order we have $\omega = \{1, 2, 3\} = \{2, 1, 1, 3\}$ in which case the notation in (9) below is ambiguous.
To simplify the presentation, we adopt the following definitions:  

\[(a) \quad A\emptyset = [ ] \in \mathbb{R}^{M \times 0},\]
\[(b) \quad \text{rank} (A\emptyset) = 0.\]

In order to avoid possible ambiguities, we set

\[A_\omega^T \overset{\text{def}}{=} (A_\omega)^T,\]

where the superscript \(^T\) stands for transposed. If \(A_\omega\) is invertible, similarly \(A_\omega^{-1} \overset{\text{def}}{=} (A_\omega)^{-1}\).

In the course of this work, we shall frequently refer to the constrained quadratic optimization problem stated next.

Given \(d \in \mathbb{R}^M\) and \(\omega \subseteq \mathbb{I}_N\), problem \((P_\omega)\) reads as:

\[
\begin{dcases}
\min_{u \in \mathbb{R}^N} \|Au - d\|^2, \\
\text{subject to} \quad u[i] = 0, \quad \forall \, i \in \omega^c.
\end{dcases}
\]

Clearly, problem \((P_\omega)\) always admits a solution.

The definition below will be used to evaluate the extent of some subsets and assumptions.

**Definition 2** A property (an assumption) is called generic on \(\mathbb{R}^K\) if it holds true on a dense open subset of \(\mathbb{R}^K\).

As usual, a subset \(S \subset \mathbb{R}^K\) is said to be negligible in \(\mathbb{R}^K\) if there exists \(Z \subset \mathbb{R}^K\) whose Lebesgue measure in \(\mathbb{R}^K\) is \(L^K(Z) = 0\) and \(S \subseteq Z\). If a property fails only on a negligible set, it is said to hold almost everywhere, meaning “with probability one”. Definition 2 requires much more than almost everywhere. Let us explain.

If a property holds true for all \(v \in \mathbb{R}^K \setminus S\), where \(S \subseteq Z \subset \mathbb{R}^K\), \(Z\) is closed in \(\mathbb{R}^K\) and \(L^K(Z) = 0\), then this property is generic on \(\mathbb{R}^K\). Indeed, \(\mathbb{R}^K \setminus Z\) contains a dense open subset of \(\mathbb{R}^K\). So if a property is generic on \(\mathbb{R}^K\), then it holds true almost everywhere on \(\mathbb{R}^K\). But the converse is false: an almost everywhere true property is not generic if the closure of its negligible subset has a positive measure because then \(\mathbb{R}^K \setminus Z\) does not contain an open subset of \(\mathbb{R}^K\). In this sense, a generic property is stable with respect to the objects to which it applies.

The elements of a set \(S \subset \mathbb{R}^K\) where a generic property fails are highly exceptional in \(\mathbb{R}^K\). The chance that a truly random \(v \in \mathbb{R}^K\)—i.e., a \(v\) following a non singular probability distribution on \(\mathbb{R}^K\)—comes across such an \(S\) can be ignored in practice.

**1.2 Content of the paper**

The main result in section 2 tells us that finding a solution of \((P_\omega)\) for \(\omega \subset \mathbb{I}_N\) is equivalent to computing a (local) minimizer of \(F_d\). In section 3 we prove that the (local) minimizers \(\hat{u}\) of \(F_d\) are strict if and only if the submatrix \(A_{\sigma}(\hat{u})\) has full column rank. The strict minimizers of \(F_d\) are shown to be linear in data \(d\). The importance of the \((M - 1)\)-sparse strict minimizers is emphasized. The global minimizers of \(F_d\) are studied in section 4. Their existence is proved. They are shown to be strict for any \(d\) and for any \(\beta > 0\). More details are provided under the

---

5 Note that (a) corresponds to the zero mapping on \(\mathbb{R}^0\) and that (b) is the usual definition for the rank of an empty matrix.

6 In the light of (8), \(A_\omega^T\) could also mean \((A^T)_\omega\).

7 There are many examples—e.g. \(Z = \{ x \in [0,1] : x \text{ is rational}\}\), then \(L^1(Z) = 0\) and \(L^1(\text{closure}(Z)) = 1\).
standard assumption that \( \text{rank}(A) = M < N \). Given \( d \in \mathbb{R}^M \), critical values \( \beta_K \) for \( K \in \mathbb{I}_{M-1} \) are exhibited such that all global minimizers of \( F_d \) are \( K \)-sparse\(^8\) if \( \beta > \beta_K \).

In section 5 a gentle assumption on \( A \) is shown to be generic for all \( M \times N \) real matrices. Under this assumption, for all data \( d \in \mathbb{R}^M \) beyond a closed negligible subset, the objective \( F_d \) for \( \beta > \beta_K \), \( K \in \mathbb{I}_{M-1} \), has a unique global minimizer and this minimizer is \( K \)-sparse.

Small size \((A \text{ is } 5 \times 10)\) numerical tests in section 6 illustrate the main theoretical results.

### 2 All minimizers of \( F_d \)

#### 2.1 Preliminary results

First, we give some basic facts on problem \((\mathcal{P}_\omega)\) as defined in (13) that are needed for later use. If \( \omega = \emptyset \), then \( \omega^c = \mathbb{I}_N \), so the unique solution of \((\mathcal{P}_\omega)\) is \( \hat{u} = 0 \). For an arbitrary \( \omega \subset \mathbb{I}_N \) meeting \( \sharp \omega \geq 1 \), \((\mathcal{P}_\omega)\) amounts to minimizing a quadratic term with respect to only \( \sharp \omega \) components of \( u \), the remaining entries being null. This quadratic problem \((\mathcal{Q}_\omega)\) reads as

\[
(14) \quad \min_{v \in \mathbb{R}^\sharp \omega} \|A_\omega v - d\|^2, \quad \sharp \omega \geq 1, \quad \mathcal{Q}_\omega
\]

and it always admits a solution. Using the zero-padding operator \( Z_\omega \) in (10), we have

\[
\left[ \hat{u}_\omega \in \mathbb{R}^{\sharp \omega} \text{ solves } (\mathcal{Q}_\omega) \text{ and } \hat{u} = Z_\omega(\hat{u}_\omega) \right] \iff \left[ \hat{u} \in \mathbb{R}^N \text{ solves } (\mathcal{P}_\omega), \quad \sharp \omega \geq 1 \right].
\]

The optimality conditions for \((\mathcal{Q}_\omega)\) , combined with the definition in (12)(a), give rise to the following equivalence, which holds true for any \( \omega \subseteq \mathbb{I}_N \):

\[
(15) \quad \left[ \hat{u} \in \mathbb{R}^N \text{ solves } (\mathcal{P}_\omega) \right] \iff \left[ \hat{u}_\omega \in \mathbb{R}^{\sharp \omega} \text{ solves } A_\omega^T A_\omega v = A_\omega^T d \text{ and } \hat{u} = Z_\omega(\hat{u}_\omega) \right].
\]

Note that \( A_\omega^T A_\omega v = A_\omega^T d \) in (15) is the normal equation associated with \( A_\omega v = d \). The remark below shows that the optimal value of \((\mathcal{P}_\omega)\) in (13) can also be seen as an orthogonal projection problem.

**Remark 1** Let \( r \stackrel{\text{def}}{=} \text{rank}(A_\omega) \) and \( B_\omega \in \mathbb{R}^{M \times r} \) be an orthonormal basis for \( \text{range}(A_\omega) \). Then \( A_\omega = B_\omega H_\omega \) for a uniquely defined matrix \( H_\omega \in \mathbb{R}^{r \times \sharp \omega} \) with \( \text{rank}(H_\omega) = r \). Using (15), we have

\[
A_\omega^T A_\omega \hat{u}_\omega = A_\omega^T d \iff H_\omega^T H_\omega \hat{u}_\omega = H_\omega^T B_\omega^T d \iff H_\omega \hat{u}_\omega = B_\omega^T d \iff A_\omega \hat{u}_\omega = B_\omega B_\omega^T d.
\]

In addition, \( \Pi_{\text{range}(A_\omega)} = B_\omega B_\omega^T \) is the orthogonal projector onto the subspace spanned by the columns of \( A_\omega \), see e.g. (29). The expression above combined with (15) shows that

\[
\left[ \hat{u} \in \mathbb{R}^N \text{ solves } (\mathcal{P}_\omega) \right] \iff \left[ \hat{u}_\omega \in \mathbb{R}^{\sharp \omega} \text{ meets } A_\omega \hat{u}_\omega = \Pi_{\text{range}(A_\omega)} d \text{ and } \hat{u} = Z_\omega(\hat{u}_\omega) \right].
\]

*Obviously, \( \hat{u} = A_\omega \hat{u}_\omega \) is the orthogonal projection of \( d \) onto the basis \( B_\omega \).*

For \( \omega \subseteq \mathbb{I}_N \), let \( K_\omega \) denote the vector subspace

\[
(16) \quad K_\omega \stackrel{\text{def}}{=} \{ v \in \mathbb{R}^N : v[i] = 0, \forall i \in \omega^c \}.
\]

This notation enables problem \((\mathcal{P}_\omega)\) in (13) to be rewritten as

\[
(17) \quad \min_{u \in K_\omega} \|Au - d\|^2.
\]

The technical lemma below will be used in what follows. We emphasize that its statement is independent of the vector \( \hat{u} \in \mathbb{R}^N \setminus \{0\} \).

---

\(^8\) As usual, a vector \( u \) is said to be \( K \)-sparse if \( \|u\|_0 \leq K \).
Lemma 1 Let $d \in \mathbb{R}^M$, $\beta > 0$, and $\hat{u} \in \mathbb{R}^N \setminus \{0\}$ be arbitrary. For $\hat{\sigma} \overset{\text{def}}{=} \sigma(\hat{u})$, set
\[
(18) \quad \rho \overset{\text{def}}{=} \min \left\{ \min_{i \in \hat{\sigma}} |\hat{u}[i]|, \frac{\beta}{2\|A^T(A\hat{u} - d)\|_1 + 1} \right\}.
\]

Then $\rho > 0$.

(i) For $\phi$ as defined in (2), we have
\[
 v \in B_\infty(0, \rho) \Rightarrow \sum_{i \in \mathbb{N}} \phi(\hat{u}[i] + v[i]) = \sum_{i \in \hat{\sigma}} \phi(\hat{u}[i]) + \sum_{i \in \hat{\sigma}^c} \phi(v[i]) .
\]

(ii) For $K_\hat{\sigma}$ defined according to (16), $F_d$ satisfies
\[
 v \in B_\infty(0, \rho) \cap (\mathbb{R}^N \setminus K_\hat{\sigma}) \Rightarrow F_d(\hat{u} + v) \geq F_d(\hat{u}) ,
\]
where the inequality is strict whenever $\hat{\sigma}^c \neq \emptyset$.

The proof is outlined in Appendix 8.1.

2.2 The (local) minimizers of $F_d$ solve quadratic problems

It is worth emphasizing that no special assumptions on the matrix $A$ are adopted.

We begin with an easy but cautionary result.

Lemma 2 For any $d \in \mathbb{R}^M$ and for all $\beta > 0$, $F_d$ has a strict (local) minimum at $\hat{u} = 0 \in \mathbb{R}^N$.

Proof Using the fact that $F_d(0) = \|d\|^2 \geq 0$, we have
\[
(19) \quad F_d(v) = \|Av - d\|^2 + \beta\|v\|_0 = F_d(0) + R_d(v) ,
\]
where $R_d(v) = \|Av\|^2 - 2\langle v, A^T d \rangle + \beta\|v\|_0$.

Noticing that $\beta\|v\|_0 \geq \beta > 0$ for $v \neq 0$ leads to
\[
 v \in B_2 \left( 0, \frac{\beta}{2\|A^T d\| + 1} \right) \setminus \{0\} \Rightarrow R_d(v) \geq -2\|v\| \|A^T d\| + \beta > 0 .
\]
Inserting this implication into (19) proves the lemma. \(\square\)

For any $\beta > 0$ and $d \in \mathbb{R}^M$, the sparsest strict local minimizer of $F_d$ reads $\hat{u} = 0$. Initialization with zero of a suboptimal algorithm should generally be a bad choice. Indeed, experiments have shown that such an initialization can be harmful; see, e.g., [30, 7].

The next proposition states a result that is often evoked in this work.

Proposition 1 Let $d \in \mathbb{R}^M$. Given an $\omega \subseteq \mathbb{N}$, let $\hat{u}$ solve problem ($\mathcal{P}_\omega$) as formulated in (13). Then for any $\beta > 0$, the objective $F_d$ in (11) reaches a (local) minimum at $\hat{u}$ and
\[
(20) \quad \sigma(\hat{u}) \subseteq \omega ,
\]
where $\sigma(\hat{u})$ is given in Definition 7.
Proof Let \( \hat{u} \) solve problem \((P_\omega)\), and let \( \beta > 0 \). The constraint in \((P_\omega)\) entails \((20)\).

Consider that \( \hat{u} \neq 0 \), in which case for \( \hat{\sigma} \) we have \( 1 \leq \hat{\sigma} \leq \omega \). Using the equivalent formulation of \((P_\omega)\) given in \((16)-(17)\), yields

\[
(21) \quad v \in K_\omega \Rightarrow \|A(\hat{u} + v) - d\|^2 \geq \|A\hat{u} - d\|^2.
\]

The inclusion in \((20)\) is equivalent to \( \omega^c \subseteq \hat{\sigma}^c \). Let \( K_{\hat{\sigma}} \) be defined according to \((16)\) as well. Then

\[
\hat{u} \in K_{\hat{\sigma}} \subseteq K_\omega.
\]

Combining the latter relation with \((21)\) leads to

\[
(22) \quad v \in K_{\hat{\sigma}} \Rightarrow \|A(\hat{u} + v) - d\|^2 \geq \|A\hat{u} - d\|^2.
\]

Let \( \rho \) be defined as in \((18)\) Lemma 1. Noticing that by \((2)\) and \((16)\)

\[
(23) \quad v \in K_{\hat{\sigma}} \Rightarrow \phi(v[i]) = 0 \quad \forall \ i \in \hat{\sigma}^c,
\]

the following inequality chain is derived:

\[
\begin{align*}
v \in B_\infty(0, \rho) \cap K_{\hat{\sigma}} & \Rightarrow \ F_d(\hat{u} + v) & = \|A(\hat{u} + v) - d\|^2 + \beta \sum_{i \in \hat{\sigma}} \phi(\hat{u}[i] + v[i]) \\
& \quad \text{[by Lemma 1 (i)]} & = \|A(\hat{u} + v) - d\|^2 + \beta \sum_{i \in \hat{\sigma}} \phi(\hat{u}[i]) + \beta \sum_{i \in \hat{\sigma}^c} \phi(v[i]) \\
& \quad \text{[by \((23)\)]} & = \|A(\hat{u} + v) - d\|^2 + \beta \sum_{i \in \hat{\sigma}} \phi(\hat{u}[i]) \\
& \quad \text{[by \((22)\)]} & \geq \|A\hat{u} - d\|^2 + \beta \sum_{i \in \hat{\sigma}} \phi(\hat{u}[i]) \\
& \quad \text{[by \((3)\)]} & = F_d(\hat{u}).
\end{align*}
\]

Combining the obtained implication with Lemma 1 (ii) shows that

\[
F_d(\hat{u} + v) \geq F_d(\hat{u}) \quad \forall \ v \in B_\infty(0, \rho).
\]

If \( \hat{u} = 0 \), this is a (local) minimizer of \( F_d \) by Lemma 2.

Many authors mention that initialization is paramount for the success of approximate algorithms minimizing \( F_d \). In view of Proposition 1, if one already has a well-elaborated initialization, it could be enough to solve the relevant problem \((P_\omega)\).

The statement reciprocal to Proposition 1 is obvious but it helps the presentation.

Lemma 3 For \( d \in \mathbb{R}^M \) and \( \beta > 0 \), let \( F_d \) have a (local) minimum at \( \hat{u} \). Then \( \hat{u} \) solves \((P_{\hat{\sigma}})\) for \( \hat{\sigma} \) defined by \( \sigma(\hat{u}) \).

Proof Let \( \hat{u} \) be a (local) minimizer of \( F_d \). Denote \( \hat{\sigma} \) defined by \( \sigma(\hat{u}) \). Then \( \hat{u} \) solves the problem

\[
\min_{u \in \mathbb{R}^N} \left\{ \|Au - d\|^2 + \beta \sum_\hat{\sigma} \right\} \text{ subject to } u[i] = 0 \quad \forall \ i \in \hat{\sigma}^c.
\]

Since \( \sum_\hat{\sigma} \) is a constant, \( \hat{u} \) solves \((P_{\hat{\sigma}})\).

□
Remark 2  By Proposition 1 and Lemma 3 solving \((P_\omega)\) for some \(\omega \subseteq \mathbb{I}_N\) is equivalent to finding a (local) minimizer of \(F_d\).

This equivalence underlies most of the theory developed in this work.

Corollary 1  For \(d \in \mathbb{R}^M\) and \(\beta > 0\), let \(\hat{u}\) be a (local) minimizer of \(F_d\). Set \(\hat{\sigma} \overset{\text{def}}{=} \sigma(\hat{u})\). Then
\[
\hat{u} = Z_{\hat{\sigma}}(\hat{u}_{\hat{\sigma}}) \quad \text{where} \quad \hat{u}_{\hat{\sigma}} \quad \text{satisfies} \quad A_{\hat{\sigma}}^T A_{\hat{\sigma}} \hat{u}_{\hat{\sigma}} = A_{\hat{\sigma}}^T d .
\]
Conversely, if \(\hat{u} \in \mathbb{R}^N\) satisfies (24) for \(\hat{\sigma} = \sigma(\hat{u})\), then \(\hat{u}\) is a (local) minimizer of \(F_d\).

Proof  By Lemma 3, \(\hat{u}\) solves \((P_{\hat{\sigma}})\). The equation in (24) follows directly from (15). The last claim is a straightforward consequence of (15) and Proposition 1. \(\square\)

Remark 3  Equation (24) shows that a (local) minimizer \(\hat{u}\) of \(F_d\) follows a pseudo-hard thresholding scheme: the nonzero part \(\hat{u}_{\hat{\sigma}}\) of \(\hat{u}\) is the least squares solution with respect to the submatrix \(A_{\hat{\sigma}}\) and the whole data vector \(d\) is involved in its computation. Unlike the hard thresholding scheme in [15], unsignificant or purely noisy data entries can hardly be discarded from \(\hat{u}\) and they threaten to pollute its nonzero part \(\hat{u}_{\hat{\sigma}}\). See also Remark 6.

Noisy data \(d\) should degrade \(\hat{u}_{\hat{\sigma}}\) and this effect is stronger if \(A_{\hat{\sigma}}^T A_{\hat{\sigma}}\) is ill-conditioned [13]. The quality of the outcome critically depends on the selected (local) minimizer and on the pertinence of \(A\).

It may be interesting to evoke another consequence of Proposition 1:

Remark 4  Given \(d \in \mathbb{R}^M\), for any \(\omega \subseteq \mathbb{I}_N\), \(F_d\) has a (local) minimizer \(\hat{u}\) defined by (24) and obeying \(\sigma(\hat{u}) \subseteq \omega\).

3  The strict minimizers of \(F_d\)

We remind, yet again, that no special assumptions on \(A \in \mathbb{R}^{M \times N}\) are adopted.

Strict minimizers of an objective function enable unambiguous solutions of inverse problems. The definition below is useful in characterizing the strict minimizers of \(F_d\).

Definition 3  Given a matrix \(A \in \mathbb{R}^{M \times N}\), for any \(r \in \mathbb{I}_M\) we define \(\Omega_r\) as the subset of all \(r\)-length supports that correspond to full column rank \(M \times r\) submatrices of \(A\), i.e.,
\[
\Omega_r = \left\{ \omega \subseteq \mathbb{I}_N : \# \omega = r = \text{rank}(A_{\omega}) \right\}.
\]
Set \(\Omega_0 = \emptyset\) and define as well
\[
\Omega = \bigcup_{r=0}^{M-1} \Omega_r \quad \text{and} \quad \Omega_{\text{max}} = \Omega \cup \Omega_M .
\]
Definition 3 shows that for any \(r \in \mathbb{I}_M\),
\[
\text{rank}(A) = r \geq 1 \iff \Omega_r \neq \emptyset \quad \text{and} \quad \Omega_t = \emptyset \quad \forall \ t \geq r + 1 .
\]

\(^9\)In a Bayesian setting, the quadratic data fidelity term in \(F_d\) models data corrupted with Gaussian i.i.d. noise.
3.1 How to recognize a strict minimizer of $F_d$?

The theorem below gives an exhaustive answer to this question.

**Theorem 1** Given $d \in \mathbb{R}^M$ and $\beta > 0$, let $\hat{u}$ be a (local) minimizer of $F_d$. Define

$$\hat{\sigma} \overset{\text{def}}{=} \sigma(\hat{u}) .$$

The following statements are equivalent:

(i) The (local) minimum that $F_d$ has at $\hat{u}$ is strict;

(ii) $\text{rank} (A_{\hat{\sigma}}) = \# \hat{\sigma}$;

(iii) $\hat{\sigma} \in \Omega_{\text{max}}$.

If $\hat{u}$ is a strict (local) minimizer of $F_d$, then it reads

$$\hat{u} = Z_{\hat{\sigma}} (\hat{u}_{\hat{\sigma}}) \quad \text{for} \quad \hat{u}_{\hat{\sigma}} = (A_{\hat{\sigma}}^T A_{\hat{\sigma}})^{-1} A_{\hat{\sigma}}^T d$$

and satisfies $\hat{\sigma} = \| \hat{u} \|_0 \leq M$.

**Proof** We break the proof into four parts.

[(i)$\Rightarrow$(ii)] We recall that by the rank-nullity theorem [21] [29]

$$\dim \ker (A_{\hat{\sigma}}) = \# \hat{\sigma} - \text{rank} (A_{\hat{\sigma}}) .$$

Let $\hat{u} \neq 0$ be a strict (local) minimizer of $F_d$. Assume that (ii) fails. Then (26) implies that

$$\dim \ker (A_{\hat{\sigma}}) \geq 1 .$$

By Lemma 3, $\hat{u}$ solves $(P_{\hat{\sigma}})$ . Let $\rho$ read as in (18) and let $K_{\hat{\sigma}}$ be defined according to (16). Noticing that

$$v \in K_{\hat{\sigma}}, \quad \hat{\sigma} \neq \emptyset \quad \Rightarrow \quad Av = A_{\hat{\sigma}} v_{\hat{\sigma}} ,$$

Lemma 1(i) shows that

\[
\begin{cases}
  v \in B_\infty(0, \rho) \cap K_{\hat{\sigma}} , \\
  v_{\hat{\sigma}} \in \ker (A_{\hat{\sigma}})
\end{cases}
\Rightarrow \quad F_d(\hat{u} + v) = \| A_{\hat{\sigma}} (\hat{u}_{\hat{\sigma}} + v_{\hat{\sigma}}) - d \|^2 + \beta \sum_{i \in \hat{\sigma}} \phi(\hat{u}[i] + v[i])
\]

\[
\begin{aligned}
\left[\text{by Lemma 1(i)}\right] &\Rightarrow \quad \| A_{\hat{\sigma}} \hat{u}_{\hat{\sigma}} - d \|^2 + \beta \sum_{i \in \hat{\sigma}} \phi(\hat{u}[i]) + \beta \sum_{i \in \hat{\sigma}^c} \phi(v[i]) \\
&\left[\text{by (23)}\right] = \| A_{\hat{\sigma}} \hat{u}_{\hat{\sigma}} - d \|^2 + \beta \sum_{i \in \hat{\sigma}} \phi(\hat{u}[i]) \\
&\left[\text{by (3)}\right] = F_d(\hat{u}) ,
\end{aligned}
\]

i.e., that $\hat{u}$ is not a strict minimizer, which contradicts (i). Hence the assumption in (27) is false. Therefore (ii) holds true.

If $\hat{u} = 0$, then $\hat{\sigma} = \emptyset$; hence $A_{\hat{\sigma}} \in \mathbb{R}^{M \times 0}$ and $\text{rank} (A_{\hat{\sigma}}) = 0 = \# \hat{\sigma}$ according to (12).

\[\text{This part can alternatively be proven using Remark 1.}\]
(ii) ⇒ (i)] Let \( \hat{u} \) be a minimizer of \( \mathcal{F}_d \) that satisfies (ii). To have \( \hat{\sigma} = 0 \) is equivalent to \( \hat{u} = 0 \). By Lemma 2, \( \hat{u} \) is a strict minimizer. Focus on \( \hat{\sigma} \geq 1 \). Since rank \( (A_{\hat{\sigma}}) = \hat{\sigma} \leq M \) and problem (\( Q_\omega \)) in (14) is strictly convex for \( \omega = \hat{\sigma} \), its unique solution \( \hat{u}_{\hat{\sigma}} \) satisfies

\[
v \in \mathbb{R}^{\hat{\sigma}} \setminus \{0\} \implies \| A_{\hat{\sigma}} (\hat{u}_{\hat{\sigma}} + v) - d \|_2^2 > \| A_{\hat{\sigma}} \hat{u}_{\hat{\sigma}} - d \|_2^2 .
\]

Using (28), this is equivalent to

\[
(29) \; v \in K_{\hat{\sigma}} \setminus \{0\} \implies \| A(\hat{u} + v) - d \|_2^2 = \| A_{\hat{\sigma}} (\hat{u}_{\hat{\sigma}} + v_{\hat{\sigma}}) - d \|_2^2 > \| A_{\hat{\sigma}} \hat{u}_{\hat{\sigma}} - d \|_2^2 = \| A\hat{u} - d \|_2^2 .
\]

Lemma 1(i), along with (23), yields

\[
v \in B_{\infty}(0, \rho) \cap K_{\hat{\sigma}} \setminus \{0\} \implies F_d(\hat{u} + v) = \| A(\hat{u} + v) - d \|_2^2 + \beta \sum_{i \in \hat{\sigma}} \phi(\hat{u}[i]) \]

by (29),

\[
> \| A\hat{u} - d \|_2^2 + \beta \sum_{i \in \hat{\sigma}} \phi(\hat{u}[i]) = F_d(\hat{u}) .
\]

Since \( \hat{\sigma} \leq M \leq N - 1 \), we have \( \hat{\sigma}^\complement \neq \emptyset \). So Lemma 1(ii) tells us that

\[
v \in B_{\infty}(0, \rho) \setminus K_{\hat{\sigma}} \implies F_d(\hat{u} + v) > F_d(\hat{u}) .
\]

Combining the last two implications proves (i).

[(ii) ⇒ (iii)] Comparing (iii) with Definitions 1 and 3 proves the claim.

[Equation (25)] The proof follows from equation (24) in Corollary 1 where \( A_{\hat{\sigma}}^T A_{\hat{\sigma}} \) is invertible.

\[\square\]

Theorem 1 provides a simple rule enabling one to verify whether or not a numerical scheme has reached a strict (local) minimizer of \( \mathcal{F}_d \).

The notations \( \Omega_r \), \( \Omega \) and \( \Omega_{\max} \) are frequently used in this paper. Their interpretation is obvious in the light of Theorem 1. For any \( d \in \mathbb{R}^M \) and for all \( \beta > 0 \), the set \( \Omega_{\max} \) is composed of the supports of all possible strict (local) minimizers of \( \mathcal{F}_d \), while \( \Omega \) is is the subset of those that are \( (M - 1) \)-sparse.

An easy and useful corollary is presented next.

**Corollary 2** Let \( d \in \mathbb{R}^M \). Given an arbitrary \( \omega \in \Omega_{\max} \), let \( \hat{u} \) solve (\( P_\omega \)). Then

(i) \( \hat{u} \) reads as

\[
(30) \quad \hat{u} = Z_{\omega}(\hat{u}_\omega) , \quad \text{where} \quad \hat{u}_\omega = (A_{\omega}^T A_{\omega})^{-1} A_{\omega}^T d ,
\]

and obeys \( \hat{\sigma} \overset{\text{def}}{=} \sigma(\hat{u}) \subseteq \omega \) and \( \hat{\sigma} \in \Omega_{\max} \);

(ii) for any \( \beta > 0 \), \( \hat{u} \) is a strict (local) minimizer of \( \mathcal{F}_d \);

(iii) \( \hat{u} \) solves (\( P_{\hat{\sigma}} \)).

\[\text{11 For } \hat{\sigma} = \emptyset, \text{Eq. } 10 \text{ and Eq. } \text{(a)} \text{ show that } 25 \text{ yields } \hat{u} = 0.\]

11
Proof Using \([15]\), \(\hat{u}\) fulfills (i) since \(A_T^\dagger A_\omega\) is invertible and \(\sigma(\hat{u}) \subseteq \omega\) by the constraint in \((P_\omega)\). If \(\hat{\sigma} = \emptyset\), (ii) follows from Lemma \(2\) for \(\hat{\sigma} \gg 1\). \(A_\hat{\sigma}\) is an \(M \times \hat{\sigma}\) submatrix of \(A_\omega\). Since \(\text{rank}(A_\omega) = \hat{\sigma}\omega\), we have \(\text{rank}(A_\hat{\sigma}) = \hat{\sigma}\) and so \(\hat{\sigma} \in \Omega_{\max}\). By Proposition \(1\) \(\hat{u}\) is a (local) minimizer of \(\mathcal{F}_d\), and Theorem \(1\) leads to (ii). Lemma \(3\) and Corollary \(2\) (ii) yield (iii). \(\square\)

Remark 5 One can easily compute a strict (local) minimizer \(\hat{u}\) of \(\mathcal{F}_d\) without knowing the value of the regularization parameter \(\beta\). Just data \(d\) and an \(\omega \in \Omega_{\max}\) are needed.

This consequence of Corollary \(2\) might be striking.

Clearly, the support \(\sigma(\hat{u})\) of a nonstrict local minimizer \(\bar{u}\) of \(\mathcal{F}_d\) contains some subsupports yielding strict (local) minimizers of \(\mathcal{F}_d\). It is easy to see that among them, there is \(\hat{\sigma} \subseteq \sigma(\bar{u})\) such that the corresponding \(\hat{u}\) given by \((25)\) strictly decreases the value of \(\mathcal{F}_d\); i.e., \(\mathcal{F}_d(\hat{u}) < \mathcal{F}_d(\bar{u})\).

3.2 Every strict (local) minimizer of \(\mathcal{F}_d\) is linear in \(d\)

Here we explore the behavior of the strict (local) minimizers of \(\mathcal{F}_d\) with respect to variations of \(d\). An interesting sequel of Theorem \(1\) is presented in the following corollary.

Corollary 3 For \(d \in \mathbb{R}^M\) and \(\beta > 0\), let \(\hat{u}\) be a (local) minimizer of \(\mathcal{F}_d\) satisfying \(\hat{\sigma} \overset{\text{def}}{=} \sigma(\hat{u}) \in \Omega\). Define
\[
N_{\hat{\sigma}} \overset{\text{def}}{=} \ker (A_\hat{\sigma}^T) \subset \mathbb{R}^M.
\]
We have \(\dim N_{\hat{\sigma}} = M - \hat{\sigma} \gg 1\) and
\[
d' \in N_{\hat{\sigma}} \implies \mathcal{F}_{d+d'} \text{ has a strict (local) minimum at } \hat{u}.
\]

Proof Since \(\hat{\sigma} \in \Omega\), the minimizer \(\hat{u}\) is strict by Theorem \(1\). By \(d' \in \ker (A_\hat{\sigma}^T)\) we find \(A_\hat{\sigma}^T(d + d') = A_\hat{\sigma}^Td\) for any \(d' \in N_{\hat{\sigma}}\). Inserting this into \((25)\) in Theorem \(1\) yields the result. \(\square\)

All data located in the vector subspace \(N_{\hat{\sigma}} \nsubseteq \{0\}\) yield the same strict (local) minimizer \(\hat{u}\).

Remark 6 If data contain noise \(n\), it can be decomposed in a unique way as \(n = n_{N_{\hat{\sigma}}} + n_{N_{\hat{\sigma}}^\perp}\) where \(n_{N_{\hat{\sigma}}} \in N_{\hat{\sigma}}\) and \(n_{N_{\hat{\sigma}}^\perp} \in N_{\hat{\sigma}}^\perp\). The component \(n_{N_{\hat{\sigma}}}\) is removed (Corollary \(3\)), while \(n_{N_{\hat{\sigma}}^\perp}\) is transformed according to \((25)\) and added to \(\hat{u}_{\hat{\sigma}}\).

We shall use the following definition.

Definition 4 Let \(O \subset \mathbb{R}^M\) be an open domain. We say that \(\mathcal{U} : O \to \mathbb{R}^N\) is a local minimizer function for the family of objectives \(\mathcal{F}_O \overset{\text{def}}{=} \{\mathcal{F}_d : d \in O\}\) if for any \(d \in O\), the function \(\mathcal{F}_d\) reaches a strict (local) minimum at \(\mathcal{U}(d)\).

Corollary \(2\) shows that for any \(d \in \mathbb{R}^M\), each strict (local) minimizer of \(\mathcal{F}_d\) is entirely described by an \(\omega \in \Omega_{\max}\) via equation \((31)\) in the same corollary. Consequently, a local minimizer function \(\mathcal{U}\) is associated with every \(\omega \in \Omega_{\max}\).

Lemma 4 For some arbitrarily fixed \(\omega \in \Omega_{\max}\) and \(\beta > 0\), the family of functions \(\mathcal{F}_{\mathbb{R}M}\) has a linear (local) minimizer function \(\mathcal{U} : \mathbb{R}^M \to \mathbb{R}^N\) that reads as
\[
\forall d \in \mathbb{R}^M, \quad \mathcal{U}(d) = Z_\omega(U_\omega d), \quad \text{where } U_\omega = (A_\omega^T A_\omega)^{-1} A_\omega^T \in \mathbb{R}^{\hat{\omega} \times M}.
\]
Proof The function $\mathcal{U}$ in (31) is linear in $d$. From Corollary 2 for any $\beta > 0$ and for any $d \in \mathbb{R}^{M}$, $\mathcal{F}_d$ has a strict (local) minimum at $\hat{\mathcal{U}} = \mathcal{U}(d)$. Hence $\mathcal{U}$ fits Definition 1.

Thus, even if $\mathcal{F}_d$ has many strict (local) minimizers, each is linear in $d$.

Next we exhibit a closed negligible subset of $\mathbb{R}^{M}$, associated with a nonempty $\omega \in \Omega_{\text{max}}$, whose elements are data $d$ leading to $\|\mathcal{U}(d)\|_0 < \frac{\beta}{\# \omega}$.

Lemma 5 For any $\omega \in \Omega_{\text{max}}$, define the subset $D_\omega \subset \mathbb{R}^M$ by

$$(32) \quad D_\omega \overset{\text{def}}{=} \bigcup_{i=1}^{\# \omega} \left\{ g \in \mathbb{R}^M : \left\langle e_i, \left(A^T_\omega A_\omega\right)^{-1} A^T_\omega g \right\rangle = 0 \right\}.$$  

Then $D_\omega$ is closed in $\mathbb{R}^M$ and $\mathbb{I}^M(D_\omega) = 0$.

Proof If $\omega = \emptyset$ then $D_\omega = \emptyset$, hence the claim. Let $\# \omega \geq 1$. For some $i \in \mathbb{I}^\omega$, set

$$D \overset{\text{def}}{=} \left\{ g \in \mathbb{R}^M : \left\langle e_i, \left(A^T_\omega A_\omega\right)^{-1} A^T_\omega g \right\rangle = 0 \right\} = \left\{ g \in \mathbb{R}^M : \left\langle A_\omega \left(A^T_\omega A_\omega\right)^{-1} e_i, g \right\rangle = 0 \right\}.$$  

Since $\text{rank} \left(A_\omega \left(A^T_\omega A_\omega\right)^{-1}\right) = \# \omega$, $\ker \left(A_\omega \left(A^T_\omega A_\omega\right)^{-1}\right) = \{0\}$. Hence $A_\omega \left(A^T_\omega A_\omega\right)^{-1} e_i \neq 0$. Therefore $D$ is a vector subspace of $\mathbb{R}^M$ of dimension $M - 1$ and so $\mathbb{I}^M(D) = 0$. The conclusion follows from the fact that $D_\omega$ in (32) is the union of $\# \omega$ subsets like $D$ (see, e.g., [36] [16]).

Proposition 2 For some arbitrarily fixed $\omega \in \Omega_{\text{max}}$ and $\beta > 0$, let $\mathcal{U} : \mathbb{R}^M \to \mathbb{R}^N$ be the relevant (local) minimizer function for $\mathcal{F}_{\mathcal{R}^M}$ as given in (31) (Lemma 4). Let $D_\omega$ read as in (32). Then the function $d \mapsto \mathcal{F}_d(\mathcal{U}(d))$ is $C^\infty$ on $\mathbb{R}^M \setminus D_\omega$ and

$$d \in \mathbb{R}^M \setminus D_\omega \implies \sigma(\mathcal{U}(d)) = \omega,$$

where the set $\mathbb{R}^M \setminus D_\omega$ contains an open and dense subset of $\mathbb{R}^M$.

Proof The statement about $\mathbb{R}^M \setminus D_\omega$ is a direct consequence of Lemma 5.

If $\omega = \emptyset$, then $\mathcal{U}(d) = 0$ for all $d \in \mathbb{R}^M$, the result is trivial. Consider that $\# \omega \geq 1$. For any $i \in \mathbb{I}^\omega$, the $\omega[i]$th component of $\mathcal{U}(d)$ reads as (see Lemma 1)

$$\mathcal{U}_{\omega[i]}(d) = \left\langle e_i, \left(A^T_\omega A_\omega\right)^{-1} A^T_\omega d \right\rangle.$$  

The definition of $D_\omega$ shows that

$$d \in \mathbb{R}^M \setminus D_\omega \text{ and } i \in \mathbb{I}^\omega \implies \mathcal{U}_{\omega[i]}(d) \neq 0,$$

whereas $\mathcal{U}_i(d) = 0$ for all $i \in \omega^c$. Consequently,

$$\omega \in \Omega_{\text{max}} \text{ and } d \in \mathbb{R}^M \setminus D_\omega \implies \sigma(\mathcal{U}(d)) = \omega.$$  

Then $\|\mathcal{U}(d)\|_0$ is constant on $\mathbb{R}^M \setminus D_\omega$ and

$$\omega \in \Omega_{\text{max}} \text{ and } d \in \mathbb{R}^M \setminus D_\omega \implies \mathcal{F}_d(\mathcal{U}(d)) = \|\mathcal{A}\mathcal{U}(d) - d\|^2 + \beta \# \omega.$$  

We infer from (31) that $d \mapsto \|\mathcal{A}\mathcal{U}(d) - d\|^2$ is $C^\infty$ on $\mathbb{R}^M$, so $d \mapsto \mathcal{F}_d(\mathcal{U}(d))$ is $C^\infty$ on $\mathbb{R}^M \setminus D_\omega$. □

A generic property is that a local minimizer function corresponding to $\mathcal{F}_d$ produces solutions sharing the same support. The application $d \mapsto \mathcal{F}_d(\mathcal{U}(d))$ is discontinuous on the closed negligible subset $D_\omega$, where the support of $\mathcal{U}(d)$ is shrunk.
3.3 Strict minimizers with an $M$-length support

Here we explain why minimizers with an $M$-length support are useless in general.

**Proposition 3** Let $\text{rank}(A) = M$, $\beta > 0$ and for $d \in \mathbb{R}^M$ set

$$U_M \overset{\text{def}}{=} \{ \hat{u} \in \mathbb{R}^N : \hat{u} \text{ is a strict (local) minimizer of } F_d \text{ meeting } \|\hat{u}\|_0 = M \}.$$ 

Put

$$Q_M \overset{\text{def}}{=} \bigcup_{\omega \in \Omega_M} \bigcup_{i \in \mathbb{N}} \{ g \in \mathbb{R}^M : \langle e_i, A_\omega^{-1} g \rangle = 0 \}.$$ 

Then $\mathbb{R}^M \setminus Q_M$ contains a dense open subset of $\mathbb{R}^M$ and

$$d \in \mathbb{R}^M \setminus Q_M \implies \nexists \ U_M \ni \omega \text{ and } F_d(\hat{u}) = \beta M \ \forall \hat{u} \in U_M.$$ 

**Proof** Using the notation in [32], $Q_M$ reads as

$$Q_M = \bigcup_{\omega \in \Omega_M} D_\omega.$$ 

The claim on $\mathbb{R}^M \setminus Q_M$ follows from Lemma 5. Since $\text{rank}(A) = M$, we have $\nexists \Omega_M \ni \omega \geq 1$. 

Consider that $d \in \mathbb{R}^M \setminus Q_M$. By Proposition 2

$$d \in \mathbb{R}^M \setminus Q_M \text{ and } \omega \in \Omega_M \implies F_d \text{ has a strict (local) minimizer } \hat{u} \text{ obeying } \sigma(\hat{u}) = \omega.$$ 

Hence $\hat{u} \in U_M$. Therefore, we have a mapping $b : \Omega_M \to U_M$ such that $\hat{u} = b(\omega) \in U_M$. Using Lemma 3 and Corollary 2 it reads as

$$b(\omega) = Z_\omega(A_\omega^{-1} d).$$ 

For $(\omega, \varpi) \in \Omega_M \times \Omega_M$ with $\varpi \neq \omega$ one obtains $\hat{u} = b(\omega) \in U_M$, $\hat{u} = b(\varpi) \in U_M$ and $\hat{u} \neq \hat{u}$, hence $b$ is one-to-one. Conversely, for any $\hat{u} \in U_M$ there is $\omega \in \Omega_M$ such that $\hat{u} = b(\omega)$ and $\sigma(\hat{u}) = \omega$ (because $d \notin Q_M$). It follows that $b$ maps $\Omega_M$ onto $U_M$. Therefore, $\Omega_M$ are $U_M$ in one-to-one correspondence, i.e. $\nexists \Omega_M = \nexists U_M$. 

Last, it is clear that $\omega \in \Omega_M$ and $d \notin Q_M$ lead to $\|\hat{u} - d\|^2 = 0$ and $F_d(\hat{u}) = \beta M$. 

For any $\beta > 0$, a generic property of $F_d$ is that it has $\nexists \Omega_M$ strict minimizers $\hat{u}$ obeying $\|\hat{u}\|_0 = M$ and $F_d(\hat{u}) = \beta M$. It is hard to discriminate between all these minimizers. Hence the interest in minimizers with supports located in $\Omega$, i.e., strict $(M-1)$-sparse minimizers of $F_d$.

4 On the global minimizers of $F_d$

The next proposition gives a necessary condition for a global minimizer of $F_d$. It follows directly from [32, Proposition 3.4] where the regularization term is $\|Du\|_0$ for a full row rank matrix $D$. For $F_d$ in [4] with $\|a_i\|_2 = 1$, $\forall i \in \mathbb{N}$, a simpler condition was derived later in [40, Theorem 12], using different tools. For completeness, the proof for a general $A$ is outlined in Appendix 8.2.

**Proposition 4** For $d \in \mathbb{R}^M$ and $\beta > 0$, let $F_d$ have a global minimum at $\hat{u}$. Then

$$(33) \quad i \in \sigma(\hat{u}) \implies |\hat{u}[i]| \geq \frac{\sqrt{\beta}}{\|a_i\|}.$$ 

---

12 Just set $g_i = e_i$, $P = I_M$ and $H = I_N$ in [32, Proposition 3.4].

13 Recall that $a_i \neq 0$ for all $i \in \mathbb{N}$ by [4] and that $\|\cdot\|_1 = \|\cdot\|_2$.
Observe that the lower bound on \(\{|\hat{u}[i]| : i \in \sigma(\hat{u})\}\) given in (33) is independent of \(d\). This means that in general, (33) provides a pessimistic bound.

The proof of the statement shows that (33) is met also by all (local) minimizers of \(F_d\) satisfying

\[ F_d(\hat{u}) \leq F_d(\hat{u} + \rho e_i) \quad \forall \rho \in \mathbb{R}, \quad \forall i \in \mathbb{N}. \]

4.1 The global minimizers of \(F_d\) are strict

**Remark 7** Let \(d \in \mathbb{R}^M\) and \(\beta > \|d\|^2\). Then \(F_d\) has a strict global minimum at \(\hat{u} = 0\). Indeed,

\[ v \neq 0 \Rightarrow \|v\|_0 \geq 1 \Rightarrow F_d(0) = \|d\|^2 < \beta \leq \|Av - d\|^2 + \beta\|v\|_0. \]

For least-squares regularized with a more regular \(\phi\), one usually gets \(\hat{u} = 0\) asymptotically as \(\beta \to +\infty\) but \(\hat{u} \neq 0\) for finite values of \(\beta\). This does not hold for \(F_d\) by Remark 7.

Some theoretical results on the global minimizers of \(F_d\) have been obtained [32, 22, 40, 7]. Surprisingly, the question about the existence of global minimizers of \(F_d\) has never been raised.

We answer this question using the notion of asymptotically level stable functions introduced by Auslender [2] in 2000. As usual,

\[ \text{lev} (F_d, \lambda) \overset{\text{def}}{=} \{v \in \mathbb{R}^N : F_d(v) \leq \lambda\} \quad \text{for} \quad \lambda > \inf F_d. \]

The following definition is taken from [3, p. 94].

**Definition 5** Let \(F_d : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}\) be lower semicontinuous and proper. Then \(F_d\) is said to be asymptotically level stable if for each \(\rho > 0\), each bounded sequence \(\{\lambda_k\} \subseteq \mathbb{R}\) and each sequence \(\{v_k\} \subseteq \mathbb{R}^N\) satisfying

\[ v_k \in \text{lev} (F_d, \lambda_k), \quad \|v_k\| \to +\infty, \quad v_k \|v_k\|^{-1} \to \bar{v} \in \ker ((F_d)_\infty), \]

where \((F_d)_\infty\) denotes the asymptotic (or recession) function of \(F_d\), there exists \(k_0\) such that

\[ v_k - \rho \bar{v} \in \text{lev} (F_d, \lambda_k) \quad \forall k \geq k_0. \]

One can note that a coercive function is asymptotically level stable, since (34) is empty. We prove that our discontinuous noncoercive objective \(F_d\) is asymptotically level stable as well.

**Proposition 5** Let \(F_d : \mathbb{R}^N \to \mathbb{R}\) be of the form (1). Then \((F_d)_\infty = \ker(A)\) and \(F_d\) is asymptotically level stable.

The proof is outlined in Appendix 8.3.

**Theorem 2** Let \(d \in \mathbb{R}^M\) and \(\beta > 0\). Then

(i) the set of the global minimizers of \(F_d\)

\[ \hat{U} \overset{\text{def}}{=} \left\{ \hat{u} \in \mathbb{R}^N : \hat{u} = \min_{u \in \mathbb{R}^N} F_d(u) \right\} \]

is nonempty;

(ii) every \(\hat{u} \in \hat{U}\) is a strict minimizer of \(F_d\), i.e.,

\[ \sigma(\hat{u}) \in \Omega_{\text{max}}, \]

hence \(\|\hat{u}\|_0 \leq M\).
In order to simplify the presentation, in what follows we consider that $\hat{u}$ is a strict minimizer of $F$. This is a nice property that fails for many convex nonsmooth optimization problems.

It follows that \[ \dim \ker (A_{\hat{\sigma}}) \geq 1. \]

Choose $v_\hat{\sigma} \in \ker (A_{\hat{\sigma}}) \setminus \{0\}$ and set $v = Z_{\hat{\sigma}} (v_\hat{\sigma})$. Select an $i \in \hat{\sigma}$ obeying $v[i] \neq 0$. Define $\tilde{u}$ by \[ \tilde{u} \overset{\text{def}}{=} \hat{u} - \hat{u}[i] \frac{v}{v[i]}. \]

We have $\tilde{u}[i] = 0$ and $\tilde{u}[i] \neq 0$. Set $\tilde{\sigma} \overset{\text{def}}{=} \sigma(\tilde{u})$. Then \[ \tilde{\sigma} \subsetneq \hat{\sigma} \quad \text{hence} \quad \sharp(\tilde{\sigma}) \leq \sharp(\hat{\sigma}) - 1. \]

From $v_\hat{\sigma} / v[i] \in \ker (A_{\hat{\sigma}})$, using \[ (\hat{\sigma}, \hat{\sigma}) \]
and Remark \[ (\hat{\sigma}, \hat{\sigma}) \]
shows that $A\hat{u} = A_{\hat{\sigma}} \tilde{u}_{\hat{\sigma}} = A_{\hat{\sigma}} \tilde{u}_{\hat{\sigma}} = A\tilde{u}$.

Then \[ \mathcal{F}_d (\tilde{u}) = \| A\tilde{u} - d \|^2 + \beta \sharp(\tilde{\sigma}) \leq \mathcal{F}_d (\hat{u}) - \beta = \| A\hat{u} - d \|^2 + \beta (\sharp(\hat{\sigma}) - 1) \].

It follows that $\hat{u}$ is not a global minimizer, hence this is false. Therefore $\text{rank}(A_{\hat{\sigma}}) = \sharp(\hat{\sigma})$ and $\hat{u}$ is a strict minimizer of $\mathcal{F}_d$ (Theorem \[ (\hat{\sigma}, \hat{\sigma}) \]).

One can note that if $\text{rank}(A) = M$, any global minimizer $\hat{u}$ of $\mathcal{F}_d$ obeys \[ \mathcal{F}_d (\hat{u}) \leq \beta M. \]

According to Theorem \[ (\hat{\sigma}, \hat{\sigma}) \] the global minimizers of $\mathcal{F}_d$ are strict and their number is finite: this is a nice property that fails for many convex nonsmooth optimization problems.

### 4.2 $K$-sparse global minimizers for $K \leq M - 1$

In order to simplify the presentation, in what follows we consider that \[ \text{rank}(A) = M < N. \]

Since $\mathcal{F}_d$ has a large number (typically equal to $\sharp(\Omega_M)$) of strict minimizers with $\| \hat{u} \|_0 = M$ yielding the same value $\mathcal{F}_d (\hat{u}) = \beta M$ (see Proposition \[ (\hat{\sigma}, \hat{\sigma}) \] and the comments given after its proof), it is important to be sure that the global minimizers of $\mathcal{F}_d$ are $(M - 1)$-sparse.

We introduce a notation which is used in the rest of this paper. For any $K \in \mathbb{I}_{M-1}$, put \[ \Omega_K \overset{\text{def}}{=} \bigcup_{r=0}^{K} \Omega_r. \]

---

\[14\] This result was originally exhibited in \[ (\hat{\sigma}, \hat{\sigma}) \] (without the notion of asymptotically level stable functions).

\[15\] In detail we have $A\tilde{u} = A_{\hat{\sigma}} \tilde{u}_{\hat{\sigma}} = A_{\hat{\sigma}} \left( \tilde{u}_{\hat{\sigma}} - v_\hat{\sigma} \frac{\bar{v}[i]}{v[i]} \right) = A_{\hat{\sigma}} \tilde{u}_{\hat{\sigma}} = A_{\hat{\sigma}} \tilde{u}_{\hat{\sigma}} = A\tilde{u}$. 

where $\Omega_r$ was set up in Definition 3. Theorem 1 gives a clear meaning of the set $\overline{\Omega}_K$. For any $d \in \mathbb{R}^M$ and any $\beta > 0$, for any fixed $K \in \mathbb{I}_{M-1}$, the set $\overline{\Omega}_K$ is composed of the supports of all possible $K$-sparse strict (local) minimizers of $F_d$.

The next propositions checks the existence of $\beta > 0$ ensuring that all the global minimizers of $F_d$ are $K$-sparse, for some $K \in \mathbb{I}_{M-1}$.

**Proposition 6** Let $d \in \mathbb{R}^M$. For any $K \in \mathbb{I}_{M-1}$, there exists $\beta_K \geq 0$ such that if $\beta > \beta_K$, then each global minimizer $\hat{u}$ of $F_d$ satisfies

$$||\hat{u}||_0 \leq K \quad \text{and} \quad \sigma(\hat{u}) \in \overline{\Omega}_K.$$

One can choose $\beta_K = \|A\hat{u} - d\|^2$ where $\tilde{u}$ solves $(P_\omega)$ for some $\omega \in \Omega_K$.

The proof is given in Appendix 4. The value of $\beta_K$ in the statement is easy to compute, but in general it is not sharp.

5 Uniqueness of the global minimizer of $F_d$

The presentation is simplified using the notation introduced next. Given a matrix $A \in \mathbb{R}^{M \times N}$, with any $\omega \in \Omega$ (see Definition 3), we associate the $M \times M$ matrix $\Pi_\omega$ that yields the orthogonal projection onto the subspace spanned by the columns of $A_\omega$:

$$\Pi_\omega \overset{\text{def}}{=} A_\omega (A_\omega^T A_\omega)^{-1} A_\omega^T.$$

For $\omega \in \Omega$, the projector in Remark 1 reads $\Pi_{\text{range}(A_\omega)} = \Pi_\omega$.

Checking whether a global minimizer $\hat{u}$ of $F_d$ is unique requires us to compare its value $F_d(\hat{u})$ with the values $F_d(\bar{u})$ of the concurrent strict minimizers $\bar{u}$. Let $\hat{u}$ be an $(M - 1)$-sparse strict (local) minimizer of $F_d$. Then $\hat{\sigma} \overset{\text{def}}{=} \sigma(\hat{u}) \in \Omega$. Using Remark 1 shows that

$$F_d(\hat{u}) = \|A\hat{\sigma} \hat{\sigma} - d\|^2 + \beta \hat{\sigma} \hat{\sigma} = \|\Pi_\sigma d - d\|^2 + \beta \hat{\sigma} \hat{\sigma} = d^T (I_M - \Pi_\sigma) d + \beta \hat{\sigma} \hat{\sigma}.$$

Let $\bar{u}$ be another $(M - 1)$-sparse strict minimizer of $F_d$; set $\bar{\sigma} \overset{\text{def}}{=} \sigma(\bar{u})$. Then

$$F_d(\hat{u}) - F_d(\bar{u}) = d^T (\Pi_\sigma - \Pi_\bar{\sigma}) d + \beta (\hat{\sigma} \hat{\sigma} - \bar{\sigma} \bar{\sigma}).$$

If both $\hat{u}$ and $\bar{u} \neq \hat{u}$ are global minimizers of $F_d$, the previous equality yields

$$F_d(\hat{u}) = F_d(\bar{u}) \iff d^T (\Pi_\sigma - \Pi_\bar{\sigma}) d = -\beta (\hat{\sigma} \hat{\sigma} - \bar{\sigma} \bar{\sigma}).$$

Equation (39) reveals that the uniqueness of the global minimizer of $F_d$ cannot be guaranteed without suitable assumptions on $A$ and on $d$.

5.1 A generic assumption on $A$

We adopt an assumption on the matrix $A$ in $F_d$ in order to restrict the cases when (39) takes place for some supports $\hat{\sigma} \neq \bar{\sigma}$ obeying $\hat{\sigma} \bar{\sigma} = \bar{\sigma} \hat{\sigma}$.

**H 1** The matrix $A \in \mathbb{R}^{M \times N}$, where $M < N$, is such that for some given $K \in \mathbb{I}_{M-1}$,

$$r \in \mathbb{I}_K \quad \text{and} \quad (\omega, \varpi) \in (\Omega_r \times \Omega_r), \quad \omega \neq \varpi \Rightarrow \Pi_\omega \neq \Pi_\varpi.$$

16 Clearly, $\overline{\Omega}_{M-1} = \Omega$.
17 For $\beta \geq \beta_K$ the global minimizers of $F_d$ might be $k$-sparse for $k \ll K$.
18 If $\omega = \varnothing$, we have $A_\omega \in \mathbb{R}^{M \times 0}$ and so $\Pi_\omega$ is an $M \times M$ null matrix.
Assumption \( H_1 \) means that the angle (or the gap) between the equidimensional subspaces range \((A_\omega)\) and range \((A_\varpi)\) must be nonzero \([29]\). For instance, if \((i,j) \in \mathbb{I}_N \times \mathbb{I}_N\) meet \(i \neq j\), \( H_1 \) implies that \(a_i \neq \kappa a_j\) for any \(\kappa \in \mathbb{R} \setminus \{0\}\) since \(\Pi_{\{i\}} = a_i a_i^T / \|a_i\|^2\).

Checking whether \( H_1 \) holds for a given matrix \( A \) requires a combinatorial search over all possible couples \((\varpi, \omega) \in (\Omega_r \times \Omega_r)\) satisfying \(\varpi \neq \omega, \forall r \in \mathbb{I}_K\). This is hard to do. Instead, we wish to know whether or not \( H_1 \) is a practical limitation. Using some auxiliary claims, we shall show that \( H_1 \) fails only for a closed negligible subset of matrices in the space of all \( M \times N \) real matrices.

**Lemma 6** Given \( r \in \mathbb{I}_{M-1} \) and \( \varpi \in \Omega_r \), define the following set of submatrices of \( A \):

\[
\mathcal{H}_{\varpi} = \left\{ A_\omega : \omega \in \Omega_r \quad \text{and} \quad \Pi_\omega = \Pi_{\varpi} \right\}.
\]

Then \( \mathcal{H}_{\varpi} \) belongs to an \((r \times r)\)-dimensional subspace of the space of all \( M \times r \) matrices.

**Proof** Using the fact that \( \varpi \in \Omega_r \) and \( \omega \in \Omega_r \), we hav\(^{19}\)

\[
\Pi_\omega = \Pi_{\varpi} \quad \Leftrightarrow \quad A_\omega = A_{\varpi} \quad A_\omega.
\]

Therefore \( \mathcal{H}_{\varpi} \) equivalently reads

\[
\mathcal{H}_{\varpi} = \left\{ \begin{array}{l}
A_\omega : \omega \in \Omega_r \quad \text{and} \quad A_\omega = A_{\varpi} A_\omega.
\end{array} \right\}
\]

Let \( A_\omega \in \mathcal{H}_{\varpi} \). Denote the columns of \( A_\omega \) by \( \tilde{a}_i \) for \( i \in \mathbb{I}_r \). Then \( \Pi_{\varpi} \) yields

\[
\Pi_{\varpi} \tilde{a}_i = \tilde{a}_i, \quad \forall \ i \in \mathbb{I}_r \quad \Rightarrow \quad \tilde{a}_i \in \text{range}(A_{\varpi}), \quad \forall \ i \in \mathbb{I}_r.
\]

Hence all \( \tilde{a}_i, \ i \in \mathbb{I}_r \), live in the \( r \)-dimensional vector subspace range \((A_{\varpi})\). All the columns of each matrix \( A_\omega \in \mathcal{H}_{\varpi} \) belong to range \((A_{\varpi})\) as well. It follows that \( \mathcal{H}_{\varpi} \) belongs to a (closed) subspace of dimension \( r \times r \) in the space of all \( M \times r \) matrices, where \( r \leq M-1 \). \( \square \)

More details on the submatrices of \( A \) living in \( \mathcal{H}_{\varpi} \) are given next.

**Remark 8** The closed negligible subset \( \mathcal{H}_{\varpi} \) in Lemma 6 is formed from all the submatrices of \( A \) that are column equivalent to \( A_{\varpi} \) (see \([29\), p. 171]), that is,

\[
A_\omega \in \mathcal{H}_{\varpi} \quad \Leftrightarrow \quad \exists \ P \in \mathbb{R}^{r \times r} \quad \text{such that} \quad \text{rank}(P) = r \quad \text{and} \quad A_\omega = A_{\varpi} P.
\]

Observe that \( P \) has \( r^2 \) unknowns that must satisfy \( Mr \) equations and that \( P \) must be invertible. It should be quite unlikely that such a matrix \( P \) does exist.

This remark can help to discern whether or not structured dictionaries satisfy \( H_1 \).

Next we inspect the set of all matrices \( A \) failing assumption \( H_1 \).

**Lemma 7** Consider the set \( \mathcal{H} \) formed from \( M \times N \) real matrices described next:

\[
\mathcal{H} \quad \text{def} \quad \left\{ A \in \mathbb{R}^{M \times N} : \exists \ r \in \mathbb{I}_{M-1}, \ \exists (\varpi, \omega) \in \Omega_r \times \Omega_r, \ \varpi \neq \omega \quad \text{and} \quad \Pi_{\varpi} = \Pi_{\omega} \right\}.
\]

Then \( \mathcal{H} \) belongs to a finite union of vector subspaces in \( \mathbb{R}^{M \times N} \) whose Lebesgue measure in \( \mathbb{R}^{M \times N} \) is null.

\(^{19}\)Using \([28]\), as well as the fact that \( A_{\omega} = \Pi_{\omega} A_{\omega}, \forall \omega \in \Omega_r \), one easily derives (40) since

\[
\Pi_{\omega} = \Pi_{\varpi} \quad \Leftrightarrow \quad \begin{cases}
A_{\omega} \left( A_{\omega}^T A_{\omega} \right)^{-1} A_{\omega}^T = \Pi_{\omega} & \Rightarrow \quad \begin{cases}
A_{\omega} = \Pi_{\omega} A_{\omega} \\
A_{\omega} = \Pi_{\varpi} A_{\omega} \\
\Pi_{\omega} = \Pi_{\omega} \Pi_{\varpi} \Rightarrow \Pi_{\omega} = \Pi_{\varpi}.
\end{cases}
\end{cases}
\]

18
Proof Let \( A \in \mathcal{H} \). Then there exist at least one integer \( r \in \mathbb{I}_{M-1} \) and at least one pair \((\varpi, \omega) \in \Omega_r \times \Omega_r\) such that \( \varpi \neq \omega \) and \( \Pi_{\varpi} = \Pi_{\omega} \). Using Lemma 6, \( A \) contains (at least) one \( M \times r \) submatrix \( A_{\varpi} \) belonging to an \( r \times r \) vector subspace in the space of all \( M \times r \) real matrices. Identifying \( A \) with an \( MN \)-length vector, its entries are included in a vector subspace of \( \mathbb{R}^{MN} \) of dimension no larger than \( MN - 1 \). The claim of the lemma is straightforward. \( \square \)

We can now clarify assumption \( \mathbb{H} \) and show that it is really good.

**Theorem 3** Given an arbitrary \( K \in \mathbb{I}_{M-1} \), consider the set of \( M \times N \) real matrices below

\[
\mathcal{A}_K = \left\{ A \in \mathbb{R}^{M \times N} : A \text{ satisfies } \mathbb{H} \text{ for } K \right\}.
\]

Then \( \mathcal{A}_K \) contains an open and dense subset in the space of all \( M \times N \) real-valued matrices.

**Proof** The complement of \( \mathcal{A}_K \) in the space of all \( M \times N \) real matrices reads as

\[
\mathcal{A}_K^c = \left\{ A \in \mathbb{R}^{M \times N} : \mathbb{H} \text{ fails for } A \text{ and } K \right\}.
\]

It is clear that \( \mathcal{A}_K^c \subset \mathcal{H} \), where \( \mathcal{H} \) is described in Lemma 7. By the same lemma, \( \mathcal{A}_K^c \) is included in a closed subset of vector subspaces in \( \mathbb{R}^{M \times N} \) whose Lebesgue measure in \( \mathbb{R}^{M \times N} \) is null. Consequently, \( \mathcal{A}_K \) satisfies the statement of the theorem. \( \square \)

For any \( K \in \mathbb{I}_{M-1} \), \( \mathbb{H} \) is a generic property of all \( M \times N \) real matrices meeting \( M < N \). This is the meaning of Theorem 3 in terms of Definition 2.

We can note that \( \mathcal{A}_K \) is weakened as \( K \) decreases. This issue is illustrated in section 6.

### 5.2 A generic assumption on \( d \)

A preliminary result is stated next.

**Lemma 8** Let \((\omega, \varpi) \in \overline{\Omega}_K \times \overline{\Omega}_K\) for \( \omega \neq \varpi \) and let \( \mathbb{H} \) hold for \( K \in \mathbb{I}_{M-1} \). Given \( \kappa \in \mathbb{R} \), define

\[
T_\kappa = \left\{ g \in \mathbb{R}^M : g^T (\Pi_\omega - \Pi_{\varpi}) g = \kappa \right\}.
\]

Then \( T_\kappa \) is a closed subset of \( \mathbb{R}^M \) and \( L^M (T_\kappa) = 0 \).

**Proof** Define \( f : \mathbb{R}^M \to \mathbb{R} \) by \( f(g) = g^T (\Pi_\omega - \Pi_{\varpi}) g \). Then

\[
T_\kappa = \left\{ g \in \mathbb{R}^M : f(g) = \kappa \right\}.
\]

Using \( \mathbb{H} \), \( T_\kappa \) is closed in \( \mathbb{R}^M \). Set

\[
Q = \left\{ g \in \mathbb{R}^M : \nabla f(g) \neq 0 \right\} \quad \text{and} \quad Q^c = \mathbb{R}^M \setminus Q.
\]

Consider an arbitrary \( g \in T_\kappa \cap Q \). From \( \mathbb{H} \) \( \text{rank}(\nabla f(g)) = 1 \). For simplicity, assume that

\[
\nabla f(g)[M] = \frac{df(g)}{dg[M]} \neq 0.
\]
By the implicit functions theorem, there are open neighborhoods $\mathcal{O}_g \subset Q \subset \mathbb{R}^M$ and $\mathcal{V} \subset \mathbb{R}^{M-1}$ of $g$ and $g_{|\mathbb{R}^{M-1}}$, respectively, and a unique $C^1$-function $h_g : \mathcal{V} \to \mathbb{R}$ with $\nabla h_g$ bounded, such that

\[(44)\quad \gamma = (\gamma_{M-1}, \gamma[M]) \in \mathcal{O}_g \quad \text{and} \quad f(\gamma) = \kappa \iff \gamma_{M-1} \in \mathcal{V} \quad \text{and} \quad \gamma[M] = h_g(\gamma_{M-1}).\]

From (43) and (44) it follows that\footnote{From (43), $\mathcal{V}$ is the restriction of $\mathcal{O}_g$ to $\mathbb{R}^{M-1}$.}

\[
\mathcal{O}_g \cap T_k = \psi^g(\mathcal{O}_g \cap (\mathbb{R}^{M-1} \times \{0\})),
\]

where $\psi^g$ is a diffeomorphism on $\mathcal{O}_g$ given by

\[
\psi^g_i(\gamma) = \gamma[i], \quad 1 \leq i \leq M - 1 \quad \text{and} \quad \psi^g_M(\gamma) = h_g(\gamma_{M-1}) + \gamma[M].
\]

Since $\mathbb{L}^M(\mathcal{O}_g \cap (\mathbb{R}^{M-1} \times \{0\})) = 0$ and $\nabla\psi^g$ is bounded on $\mathcal{O}_g$, it follows from (37) Lemma 7.25 that\footnote{The same result follows from the change-of-variables theorem for the Lebesgue integral, see e.g. \cite{37}.}

\[(45)\quad S \subset Q \quad \text{and} \quad S \text{ bounded } \Rightarrow \mathbb{L}^M(S \cap T_k) = 0.
\]

Using that every open subset of $\mathbb{R}^M$ can be written as a countable union\footnote{From (37), adjacent cubes can also intersect in our case.} of cubes in $\mathbb{R}^M$ \cite{36,16,38}, the result in (45) entails that $\mathbb{L}^M(T_k \cap Q) = 0$.

Next, $Q^c = \ker (\Pi_\omega - \Pi_{\varpi})$. By \cite{11} dim ker $(\Pi_\omega - \Pi_{\varpi}) \leq M - 1$. Hence $\mathbb{L}^M(T_k \cap Q^c) = 0$. The proof follows from the equality $\mathbb{L}^M(T_k) = \mathbb{L}^M(T_k \cap Q) + \mathbb{L}^M(T_k \cap Q^c)$. \qed

We exhibit a \textit{closed negligible} subset of data in $\mathbb{R}^M$ that can still meet the equality in (39).

**Proposition 7** \textit{For} $\beta > 0$ and $K \in \mathbb{I}_{M-1}$, put

\[(46)\quad \Sigma_K \overset{\text{def}}{=} \bigcup_{n=-K}^{K} \bigcup_{\omega \in \overline{\Omega}_K} \bigcup_{\varpi \in \overline{\Omega}_K} \left\{g \in \mathbb{R}^M : \omega \neq \varpi \quad \text{and} \quad g^T(\Pi_\omega - \Pi_{\varpi})g = n^2\right\},
\]

where $\overline{\Omega}_K$ is given in \cite{37}. \textit{Let (11) hold for} $K$. \textit{Then} $\Sigma_K$ \textit{is closed in} $\mathbb{R}^M$ \textit{and} $\mathbb{L}^M(\Sigma_K) = 0$.

**Proof** \textit{For} some $n \in \{-K, \cdots, K\}$ and $(\omega, \varpi) \in (\overline{\Omega}_K \times \overline{\Omega}_K)$ such that $\omega \neq \varpi$, put

\[
\Sigma \overset{\text{def}}{=} \left\{g \in \mathbb{R}^M : g^T(\Pi_\omega - \Pi_{\varpi})g = n^2\right\}.
\]

If $\not{\parallel} \omega \neq \not{\parallel} \varpi$, then rank$(\Pi_\omega - \Pi_{\varpi}) \geq 1$. If $\not{\parallel} \omega = \not{\parallel} \varpi$, \cite{11} guarantees that rank$(\Pi_\omega - \Pi_{\varpi}) \geq 1$, yet again. The number $n^2 \in \mathbb{R}$ is given. According to Lemma \cite{8} $\Sigma$ is a closed subset of $\mathbb{R}^M$ and $\mathbb{L}^M(\Sigma) = 0$. The conclusion follows from the fact that $\Sigma_K$ is a finite union of subsets like $\Sigma$. \qed

\textit{We assume hereafter that if (11) holds for some} $K \in \mathbb{I}_{M-1}$, \textit{data} $d$ \textit{satisfy}

\[
d \in \left\{g \in \mathbb{R}^M : g \notin \Sigma_K\right\} = \mathbb{R}^M \setminus \Sigma_K.
\]
5.3 The unique global minimizer of $\mathcal{F}_d$ is $K$-sparse for $K \leq (M - 1)$

We are looking for guarantees that $\mathcal{F}_d$ has a unique global minimizer $\hat{u}$ obeying

$$\|\hat{u}\|_0 \leq K$$

for some fixed $K \in \mathbb{I}_{M-1}$.

This is the aim of the next theorem.

**Theorem 4** Given $K \in \mathbb{I}_{M-1}$, let $I[\square]$ hold for $K$, $\beta > \beta_K$ where $\beta_K$ meets Proposition 6 and $\Sigma_K \subset \mathbb{R}^M$ reads as in (46). Consider that $d \in \mathbb{R}^M \setminus \Sigma_K$.

Then

(i) the set $\mathbb{R}^M \setminus \Sigma_K$ is open and dense in $\mathbb{R}^M$;

(ii) $\mathcal{F}_d$ has a unique global minimizer $\hat{u}$ and $\|\hat{u}\|_0 \leq K$.

**Proof** Statement (i) follows from Proposition 7.

Since $\beta > \beta_K$, all global minimizers of $\mathcal{F}_d$ have their support in $\Omega_K$ (Proposition 6). Using the fact that $d \in \mathbb{R}^M \setminus \Sigma_K$, the definition of $\Sigma_K$ in (46) shows that

$$(47) \quad -K \leq n \leq K \quad \text{and} \quad (\omega, \varpi) \in (\overline{\Omega}_K \times \overline{\Omega}_K), \quad \omega \neq \varpi \implies d^T (\Pi_\omega - \Pi_\varpi) d \neq n\beta.$$

The proof is conducted by contradiction. Let $\hat{u}$ and $\bar{u} \neq \hat{u}$ be two global minimizers of $\mathcal{F}_d$. Then

$$\hat{\sigma} \overset{\text{def}}{=} \sigma(\hat{u}) \in \overline{\Omega}_K \quad \text{and} \quad \bar{\sigma} \overset{\text{def}}{=} \sigma(\bar{u}) \in \overline{\Omega}_K,$$

and $\hat{\sigma} \neq \bar{\sigma}$. By $\mathcal{F}_d(\hat{u}) = \mathcal{F}_d(\bar{u})$, (39) yields

$$(48) \quad d^T (\Pi_{\hat{\sigma}} - \Pi_{\bar{\sigma}}) d = \beta (\hat{\sigma} - \bar{\sigma}).$$

An enumeration of all possible values of $\hat{\sigma} - \bar{\sigma}$ shows that

$$\beta (\hat{\sigma} - \bar{\sigma}) = n\beta \quad \text{for some} \quad n \in \{-K, \cdots, K\}.$$

Inserting this equation into (48) leads to

$$d^T (\Pi_{\hat{\sigma}} - \Pi_{\bar{\sigma}}) d = n\beta \quad \text{for some} \quad n \in \{-K, \cdots, K\}.$$

The last result contradicts (47); hence it violates the assumptions $I[\square]$ and $d \in \mathbb{R}^M \setminus \Sigma_K$. Consequently, $\mathcal{F}_d$ cannot have two global minimizers. Since $\mathcal{F}_d$ always has global minimizers (Theorem 2(i)), it follows that $\mathcal{F}_d$ has a unique global minimizer, say $\hat{u}$. And $\|\hat{u}\|_0 \leq K$ because $\sigma(\hat{u}) \in \overline{\Omega}_K$. \qed

For $\beta > \beta_K$, the objective $\mathcal{F}_d$ in (1) has a unique global minimizer and it is $K$-sparse for $K \leq M - 1$. For all $K \in \mathbb{I}_{M-1}$, the claim holds true in a generic sense. This is the message of Theorem 4 using Definition 2.
6 Numerical illustrations

6.1 On assumption H1

Assumption H1 requires that \( \Pi_\omega \neq \Pi_{\overline{\omega}} \) when \( (\omega, \overline{\omega}) \in \Omega_r \times \Omega_r, \omega \neq \overline{\omega} \) for all \( r \leq K \in \mathbb{I}_{M-1} \).

From a practical viewpoint, the magnitude of \( (\Pi_\omega - \Pi_{\overline{\omega}}) \) should be discernible. One way to assess the viability of H1 for a matrix \( A \) and \( K \in \mathbb{I}_{M-1} \) is to calculate

\[
\xi_K(A) = \min_{r \in \mathbb{I}_K} \mu_r(A),
\]

where \( \mu_r(A) = \min_{(\omega, \overline{\omega}) \in \Omega_r \times \Omega_r, \omega \neq \overline{\omega}} \| \Pi_\omega - \Pi_{\overline{\omega}} \|_2, \ \forall r \in \mathbb{I}_K \).

In fact, \( \| \Pi_\omega - \Pi_{\overline{\omega}} \|_2 = \sin(\theta) \), where \( \theta \in [0, \pi/2] \) is the maximum angle between range \( (A_\omega) \) and range \( (A_{\overline{\omega}}) \); see \cite{29} p. 456. These subspaces have the same dimension and \( \Pi_\omega \neq \Pi_{\overline{\omega}} \) when \( (\omega, \overline{\omega}) \in \Omega_r \times \Omega_r, \omega \neq \overline{\omega} \) and \( r \in \mathbb{I}_K \), hence \( \theta \in (0, \pi/2) \). Consequently,

\[
H1 \text{ holds for } K \in \mathbb{I}_{M-1} \implies \mu_r(A) \in (0, 1] \ \forall r \in \mathbb{I}_K \implies \xi_K(A) \in (0, 1].
\]

According to \((49)\), we have \( \xi_K \geq \xi_{K+1} \), \( \forall K \in \mathbb{I}_{M-2} \). Our guess that assumption H1 is lightened when \( K \) decreases (see the comments following the proof of Theorem \(3\)) means that

\[
(50) \quad \xi_1(A) > \cdots > \xi_{M-1}(A).
\]

We provide numerical tests on two subsets of real-valued random matrices for \( M = 5 \) and \( N = 10 \), denoted by \( A_{20}^N \) and \( U_{1000}^U \). The values of \( \xi_K(\cdot), K \in \mathbb{I}_{M-1} = \mathbb{I}_4 \), for every matrix in \( A_{20}^N \) and in \( U_{1000}^U \), were calculated using an exhaustive combinatorial search. All tested matrices satisfy assumption H1 which confirms Theorem 3 and its consequences. In order to evaluate the extent of H1, we computed the worst and the best values of \( \xi_K(\cdot) \) over these sets:

\[
\begin{cases}
\xi_K^{\text{worst}} &= \min_{A \in \mathcal{A}} \xi_K(A) \\
\xi_K^{\text{best}} &= \max_{A \in \mathcal{A}} \xi_K(A)
\end{cases} \quad \forall K \in \mathbb{I}_{M-1}, \ A \in \{A_{20}^N, U_{1000}^U\}.
\]

**Set \( A_{20}^N \)** This set was formed from 20 matrices \( A^n, n \in \mathbb{I}_{20} \) of size \( 5 \times 10 \). The components of each matrix \( A^n \) were independent and uniformly drawn from the standard normal distribution with mean zero and variance one. The values of \( \xi_K(\cdot) \) are depicted in Fig. 11. We have\(23\) \( \xi_1(A^{10}) = \xi_2(A^{10}) \) and \( \xi_1(A^{17}) = \xi_2(A^{17}) \). In all other cases \((50)\) is satisfied. Fig. 11 clearly shows that \( \xi_K(\cdot) \) increases as \( K \) decreases (from \( M - 1 \) to \( 1 \)).

The worst and the best values of \( \xi_K(\cdot) \) over the whole set \( A_{20}^N \) are displayed in Table 1

<table>
<thead>
<tr>
<th>( \xi_K^{\text{worst}} )</th>
<th>( K = 1 )</th>
<th>( K = 2 )</th>
<th>( K = 3 )</th>
<th>( K = 4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K = 1 )</td>
<td>0.3519</td>
<td>0.1467</td>
<td>0.0676</td>
<td>0.0072</td>
</tr>
<tr>
<td>( K = 2 )</td>
<td>0.8666</td>
<td>0.5881</td>
<td>0.3966</td>
<td>0.0785</td>
</tr>
</tbody>
</table>

**Set \( U_{1000}^U \)** The set \( U_{1000}^U \) was composed of one thousand \( 5 \times 10 \) matrices \( A^n, n \in \mathbb{I}_{1000} \). The entries of each matrix \( A^n \) were independent and uniformly sampled on \([-1, 1]\). The obtained values for \( \xi_K^{\text{worst}} \) and \( \xi_K^{\text{best}} \), calculated according to \((51)\), are shown in Table 2

\(23\) This is why on the figure, in columns 10 and 17, the green “o” and the red “φ” overlap.
Figure 1: x-axis: the list of the 20 random matrices in $A_{20}^N$. (a) y-axis: the value $\xi_K(A^n)$ according to (49) for all $K \in \mathbb{I}_{M-1}$ and for all $n \in \mathbb{I}_{20}$. The plot in (b) is a zoom of (a) along the y-axis.

Table 2: The worst and the best values of $\xi_K(A)$, for $K \in \mathbb{I}_{M-1}$, over the set $A_{1000}^U$.

<table>
<thead>
<tr>
<th>$K$</th>
<th>$\xi_{\text{worst}}$</th>
<th>$\xi_{\text{best}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1085</td>
<td>0.9526</td>
</tr>
<tr>
<td>2</td>
<td>0.0235</td>
<td>0.8625</td>
</tr>
<tr>
<td>3</td>
<td>0.0045</td>
<td>0.5379</td>
</tr>
<tr>
<td>4</td>
<td>0.0001</td>
<td>0.1152</td>
</tr>
</tbody>
</table>

For $K \in \mathbb{I}_3$, the best values of $\xi_K(\cdot)$ were obtained for the same matrix, $A^{964}$. Note that $\xi_4(A^{964}) = 0.0425 \gg \xi_1_{\text{worst}}$. The worst values in Table 2 are smaller than those in Table 1 while the best values in Table 2 are larger than those in Table 1, one credible reason is that $A_{1000}^U$ is much larger than $A_{20}^N$.

Table 3: Percentage of the cases in $A_{1000}^U$ when (50) fails to hold.

<table>
<thead>
<tr>
<th>$\xi_1(A^n) = \xi_2(A^n)$</th>
<th>$\xi_2(A^n) = \xi_3(A^n)$</th>
<th>$\xi_3(A^n) = \xi_4(A^n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>occurrences ${n}$</td>
<td>5 %</td>
<td>1.6 %</td>
</tr>
</tbody>
</table>

Overall, (50) is satisfied on $A_{1000}^U$—the percentages in Table 3 are pretty small. All the three tables and Figure 1 agree with our guess that $H_1$ is more viable for smaller values of $K$.

Based on the magnitudes for $\xi_{\text{best}}$ in Tables 1 and 2, one can expect that there are some classes of matrices (random or not) that fit $H_1$ for larger values of $\xi_K(\cdot)$.

6.2 On the global minimizers of $\mathcal{F}_d$

Here we summarize the outcome of a series of experiments corresponding to several matrices $A \in \mathbb{R}^{M \times N}$ where $M = 5$ and $N = 10$, satisfying $H_1$ for $K = M - 1$, different original vectors $\tilde{u} \in \mathbb{R}^N$ and data samples $d = A\tilde{u} + \text{noise}$, for various values of $\beta > 0$. In each experiment, we computed the complete list of all different strict (local) minimizers of $\mathcal{F}_d$, say $(\tilde{u}^j)_{j=1}^n$. Then the sequence of values $(\mathcal{F}_d(\tilde{u}^j))_{j=1}^n$ was sorted in increasing order, $\mathcal{F}_d(\tilde{u}^{i_1}) \leq \mathcal{F}_d(\tilde{u}^{i_2}) \leq \cdots \leq \mathcal{F}_d(\tilde{u}^{i_n})$.
global minimizer \( \hat{u}\) is unique provided that \( F_d(\hat{u}) < F_d(\hat{u}') \). In order to discard numerical errors, we also checked whether \( |F_d(\hat{u}) - F_d(\hat{u}')| \) is easy to detect.

In all experiments we carried out, the following facts were observed:

- The global minimizer of \( F_d \) was unique—manifestly data \( d \) never did belong to the closed negligible subset \( \Sigma_K \) in Proposition 7. This confirms Theorem 4.
- The global minimizers of \( F_d \) remained unchanged under large variations of \( \beta \).
- The necessary condition for a global minimizer in Proposition 4 was met.

Next we present in detail two of these experiments where \( F_d \) is defined using

\[
A = \begin{bmatrix}
7 & 2 & 4 & 9 & 0 & 3 & 3 & 6 & 6 & 7 \\
3 & 4 & 9 & 3 & 3 & 9 & 1 & 3 & 1 & 5 \\
5 & 4 & 2 & 4 & 0 & 7 & 1 & 9 & 2 & 9 \\
8 & 4 & 0 & 9 & 6 & 0 & 4 & 2 & 3 & 7 \\
6 & 3 & 6 & 5 & 0 & 9 & 0 & 0 & 3 & 8 \\
\end{bmatrix}
\]

\( d = A\hat{u} + n \),

where \( n \) is noise and \( \hat{u} = (0, 1, 8, 0, 3, 0, 0, 0, 0, 9)^T \).

Only integers appear in (52) for better readability. We have \( \text{rank}(A) = M = 5 \). An exhaustive combinatorial test shows that the arbitrary matrix \( A \) in (52) satisfies H1 for \( K = M - 1 \). The values of \( \xi_K(A) \) are seen in Table 4. One notes that \( \mu_2(A) > \mu_1(A) \); hence \( \xi_1(A) = \xi_2(A) \).

Table 4: The values of \( \xi_K(A) \) and \( \mu_K(A) \), \( \forall K \in \mathbb{I}_{M-1} \), for the matrix \( A \) in (52).

<table>
<thead>
<tr>
<th>(K)</th>
<th>(\xi_K(A))</th>
<th>(\mu_K(A))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.2737</td>
<td>0.2737</td>
</tr>
<tr>
<td>2</td>
<td>0.2737</td>
<td>0.2799</td>
</tr>
<tr>
<td>3</td>
<td>0.2008</td>
<td>0.2008</td>
</tr>
<tr>
<td>4</td>
<td>0.0564</td>
<td>0.0564</td>
</tr>
</tbody>
</table>

One expects (at least when data are noise-free) that the global minimizer \( \hat{u} \) of \( F_d \) obeys \( \hat{\sigma} \subseteq \sigma(\hat{u}) \), where \( \hat{u} \) is the original in (52), and that the vanished entries of \( \hat{u} \) correspond to the least entries of \( \hat{u} \). This provides a partial way to rate the quality of the solution provided by a global minimizer \( \hat{u} \) of \( F_d \).

The experiments described hereafter correspond to two data samples relevant to (52)—without and with noise—and to several values of \( \beta > 0 \).

**Noise-free data** The noise-free data in (52) read as:

\[
(53) \quad d = A\hat{u} = (97, 130, 101, 85, 123)^T.
\]

For different values of \( \beta \), the global minimizer \( \hat{u} \) is given in Table 5.

Table 5: The global minimizer \( \hat{u} \) of \( F_d \) and its value \( F_d(\hat{u}) \) for the noise-free data \( d \) in (53) for different values of \( \beta \). Last row: the original \( \hat{u} \) in (52).

<table>
<thead>
<tr>
<th>(\beta)</th>
<th>The global minimizer ( \hat{u} ) of ( F_d ) (row vector)</th>
<th>(|\hat{u}|_0)</th>
<th>(F_d(\hat{u}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 1 8 0 3</td>
<td>0 0 0 0 9</td>
<td>4</td>
</tr>
<tr>
<td>(10^2)</td>
<td>0 0 8.12 0 3.31</td>
<td>0 0 0 0 9.33</td>
<td>3 301.52</td>
</tr>
<tr>
<td>(10^3)</td>
<td>0 0 0 0 0</td>
<td>12.58 20.28 0 0 0</td>
<td>2 2179.3</td>
</tr>
<tr>
<td>(10^4)</td>
<td>0 29.95 0 0 0</td>
<td>0 0 0 0 0</td>
<td>1 14144</td>
</tr>
<tr>
<td>(7 \cdot 10^4)</td>
<td>0 0 0 0 0</td>
<td>0 0 0 0 0</td>
<td>0 58864</td>
</tr>
</tbody>
</table>

\( \hat{u} = 0 1 8 0 3 \)
Since $\sigma(\hat{u}) \in \Omega$ and the data are noise-free, $F_d$ does not have global minimizers with $\|\hat{u}\|_0 = 5$. Actually, applying Proposition 4 for $\bar{u} = \hat{u}$ yields $\beta_{M-1} = 0$, hence for any $\beta > 0$ all global minimizers of $F_d$ have a support in $\Omega = \Omega_{M-1}$ (see Definition 3 and (57)). The global minimizer $\hat{u}$ for $\beta = 1$ meets $\bar{u} = \hat{u}$. For $\beta = 100$, the global minimizer $\hat{u}$ obeys $\hat{\sigma} \equiv \sigma(\hat{u}) = \{3, 5, 10\} \subsetneq \sigma(\bar{u})$ and $\|\hat{u}\|_0 = 3$—the least nonzero entry of the original $\bar{u}$ is canceled, which is reasonable. The global minimizers corresponding to $\beta > 300$ are meaningless. We could not find any positive value of $\beta$ giving better 2-sparsers global minimizers. Recalling that data are noise-free, this confirms Remark 3: the global minimizers of $F_d$ realize a only pseudo-hard thresholding. For $\beta \geq 7 \cdot 10^4 > \|d\|^2$, the global minimizer of $F_d$ is $\hat{u} = 0$ which confirms Remark 7.

**Noisy data** Now we consider noisy data in (52) for

\[(54) \quad n = (4, -1, 2, -3, 5)^T.\]

This arbitrary noise yields a signal-to-noise ratio (SNR) equal to 14.07 dB. If $\beta \leq 0.04$, $F_d$ has 252 different strict global minimizers $\hat{u}$ obeying $\|\hat{u}\|_0 = M$ and $F_d(\hat{u}) = \beta M$ (recall Proposition 3). For $\beta > 0.05$, the global minimizer $\hat{u}$ of $F_d$ is unique and satisfies $\sigma(\hat{u}) \in \Omega$. It is given in Table 6 for several values of $\beta \geq 0.05$.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>The global minimizer $\hat{u}$ of $F_d$ (row vector)</th>
<th>$|\hat{u}|_0$</th>
<th>$F_d(\hat{u})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 6.02 2.66 6.43 0 6.85 0 0 0 0</td>
<td>4</td>
<td>4.0436</td>
</tr>
<tr>
<td>$10^2$</td>
<td>0 0 8.23 0 0 2.3 0 0 0 0 9.71</td>
<td>3</td>
<td>301.94</td>
</tr>
<tr>
<td>$10^3$</td>
<td>0 0 8.14 0 0 0 0 0 0 0 10.25</td>
<td>2</td>
<td>2174.8</td>
</tr>
<tr>
<td>$10^4$</td>
<td>0 0 0 0 0 0 0 0 0 0 14.47</td>
<td>1</td>
<td>14473</td>
</tr>
<tr>
<td>$7 \cdot 10^4$</td>
<td>0 0 0 0 0 0 0 0 0 0 60559</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For $\beta = 1$, the global minimizer is meaningless. We could not find any positive value of $\beta$ yielding a better global minimizer with a 4-length support. For the other values of $\beta$, the global minimizer $\hat{u}$ meets $\hat{\sigma} \equiv \sigma(\hat{u}) \subsetneq \sigma(\bar{u})$, and its vanished entries correspond to the least entries in the original $\bar{u}$. For $\beta = 100$, the global minimizer seems to furnish a good approximation to $\hat{u}$. Observe that the last entry of the global minimizer $\hat{u}[10]$, corresponding to the largest magnitude in $\hat{u}$, freely increases when $\beta$ increases from $10^2$ to $10^4$. We tested a tight sequence of intermediate values of $\beta$ without finding better results. Yet again, $\beta \geq 7 \cdot 10^4 > \|d\|^2$ leads to a unique null global minimizer (see Remark 7).

Figure 2 shows the value $F_d(\hat{u})$ of all the strict local minimizers of $F_d$ for $\beta = 100$. In the zoom in Figure 2(b) it is easily seen that the global minimizer is unique (remember Theorem 1). It obeys $\|\hat{u}\|_0 = 3$ and $F_d(\hat{u}) = 301.94$. One observes that $F_d$ has 252 $\not\in \Omega_M$ different strict local minimizers $\hat{u}$ with $\|\hat{u}\|_0 = 5 = M$ and $F_d(\hat{u}) = 500 = \beta M$. This confirms Proposition 4—obviously $d$ does not belong to the closed negligible subset $Q_M$ described in the proposition.

### 7 Conclusions and perspectives

We provided a detailed analysis of the (local and global) minimizers of a regularized objective $F_d$ composed of a quadratic data fidelity term and an $\ell_0$ penalty weighted by a parameter $\beta > 0$.

---

24 Let us denote $\tilde{d} = A\hat{u}$ and $d = \tilde{d} + n$. The SNR reads $\text{SNR}(\tilde{d}, d) = 10 \log_{10} \frac{\sum_{i=1}^{M} (\tilde{d}_i - d_i)^2}{\sum_{i=1}^{M} (d_i - \hat{d}_i)^2}$. 

---

25
as given in \((1)\). We exhibited easy necessary and sufficient conditions ensuring that a (local) minimizer \(\hat{u}\) of \(F_d\) is strict (Theorem 1). The global minimizers of \(F_d\) (whose existence was proved) were shown to be strict as well (Theorem 2). Under very mild conditions, \(F_d\) was shown to have a unique global minimizer (Theorem 4). Other interesting results were listed in the abstract. Below we pose some perspectives and open questions raised by this work.

- The relationship between the value of the regularization parameter \(\beta\) and the sparsity of the global minimizers of \(F_d\) (Proposition 6) can be improved.

- The generic linearity in data \(d\) of each strict (local) minimizer of \(F_d\) (subsection 3.2) should be exploited to better characterize the global minimizers of \(F_d\).

- Is there a simple way to check whether assumption \(H1\) is satisfied by a given matrix \(A \in \mathbb{R}^{M \times N}\) when \(N\) and \(M < N\) are large? Remark 8 and in particular (42) could help to discard some nonrandom matrices. Conversely, one can ask whether there is a systematic way to construct matrices \(A\) that satisfy \(H1\).

  An alternative would be to exhibit families of matrices that satisfy \(H1\) for large values of \(\xi_K(\cdot)\), where the latter quantifiers are defined in equation (49).

- A proper adaptation of the results to matrices \(A\) and data \(d\) with complex entries should not present inherent difficulties.

- The theory developed here can be extended to MAP energies of the form evoked in \((5)\). This is important for the imaging applications mentioned there.
Based on Corollary 1 and Remarks 3 and 6 and the numerical tests in subsection 6.2, one is justified in asking for conditions ensuring that the global minimizers of $F_d$ perform a valid work. Given the high quality of the numerical results provided in many papers, see e.g., [33, 34], the question deserves attention.

There exist numerous algorithms aimed at approximating a (local) minimizer of $F_d$. As a by-product of our research, we obtained simple rules to verify whether or not an algorithm could find

- a (local) minimizer $\hat{u}$ of $F_d$—by checking whether $\hat{u}$ satisfies (24) in Corollary 1
- and whether this local minimizer is strict by testing whether the submatrix whose columns are indexed by the support of $\hat{u}$ (i.e., $A_{\sigma(\hat{u})}$) has full column rank (Theorem 1).

Some properties of the minimizers of $F_d$ given in this work can be inserted in numerical schemes in order to quickly escape from shallow local minimizers.

Many existing numerical methods involve a studious choice of the regularization parameter $\beta$, and some of them are proved to converge to a local minimizer of $F_d$. We have seen that finding a (strict or nonstrict) local minimizer of $F_d$ is easy and that it is independent of the value of $\beta$ (Corollaries 7 and 3). It is therefore obscure what meaning to attach to “choosing a good $\beta$ and proving (local) convergence”. Other successful algorithms are not guaranteed to converge to a local minimizer of $F_d$. Whenever algorithms do a good job, the choice of $\beta$, the assumptions on $A$ and on $\|\hat{u}\|_0$, and the iterative scheme and its initialization obviously provide a tool for selecting a meaningful solution by minimizing $F_d$. There is a theoretical gap that needs clarification.

The connection between the existing algorithms and the description of the minimizers exposed in this paper deserves deep exploration. What conditions ensure that an algorithm minimizing $F_d$ yields meaningful solutions? Clearly, showing local convergence does not answer this important question.

One can expect such research to give rise to innovative and more efficient algorithms enabling one to compute relevant solutions by minimizing the tricky objective $F_d$.

8 Appendix

8.1 Proof of Lemma 1

Since $\hat{u} \neq 0$, the definition of $\hat{\sigma}$ shows that $\min_{\hat{\sigma}} |\hat{u}[i]| > 0$. Then $\rho$ in (18) fulfills $\rho > 0$.

(i) Since $\hat{\sigma} \geq 1$, we have

$$
\begin{align*}
i \in \hat{\sigma}, \quad v \in B_\infty(0, \rho) \quad &\Rightarrow \quad \max_{j \in \hat{\sigma}} |v[j]| < \rho \\
&\Rightarrow \quad \max_{j \in \hat{\sigma}} |v[j]| < \min_{j \in \hat{\sigma}} |\hat{u}[j]| \\
&\Rightarrow \quad |\hat{u}[i] + v[i]| \geq |\hat{u}[i]| - |v[i]| \\
&\quad \geq \min_{j \in \hat{\sigma}} |\hat{u}[j]| - \max_{j \in \hat{\sigma}} |v[j]| \geq \rho - \max_{j \in \hat{\sigma}} |v[j]| > 0 \\
&\Rightarrow \quad \hat{u}[i] + v[i] \neq 0
\end{align*}
$$

(55) 

(by (22))

$$
\Rightarrow \quad \phi(\hat{u}[i] + v[i]) = \phi(\hat{u}[i]) = 1 .
$$

If $\hat{\sigma}^c = \emptyset$ the result is proved. Let $\hat{\sigma}^c \neq \emptyset$. Then $\hat{u}[i] = 0 = \phi(\hat{u}[i])$, $\forall i \in \hat{\sigma}^c$. Inserting this and (55) into

$$
\sum_{i \in \mathbb{N}} \phi(\hat{u}[i] + v[i]) = \sum_{i \in \hat{\sigma}} \phi(\hat{u}[i] + v[i]) + \sum_{i \in \hat{\sigma}^c} \phi(\hat{u}[i] + v[i])
$$

27
proves claim (i).

(ii) Using the fact that \( \|A(\hat{u} + v) - d\|^2 = \|A\hat{u} - d\|^2 + \|Av\|^2 + 2\langle Av, A\hat{u} - d \rangle \), one obtains

\[
v \in B_{\infty}(0, \rho) \setminus K_{\hat{\sigma}} \Rightarrow \mathcal{F}_d(\hat{u} + v) = \|A\hat{u} - d\|^2 + \|Av\|^2 + 2\langle Av, A\hat{u} - d \rangle
\]

(by Lemma 1.1)

[using (3)]

\[
= \mathcal{F}_d(\hat{u}) + \|Av\|^2 + 2\langle Av, A\hat{u} - d \rangle + \beta \sum_{i \in \hat{\sigma}} \phi(\hat{u}[i])
\]

\[
\geq \mathcal{F}_d(\hat{u}) - 2\|v\|\|A^T(A\hat{u} - d)\|_1 + \beta \|v_{\hat{\sigma}}\|_0
\]

(56) [by Hölder’s inequality]

\[
\geq \mathcal{F}_d(\hat{u}) - 2\|v\|_\infty \|A^T(A\hat{u} - d)\|_1 + \beta \|v_{\hat{\sigma}}\|_0
\]

If \( \hat{\sigma}^c = 0 \), then \( K_{\hat{\sigma}} = \mathbb{R}^N \), so \( v \in \mathbb{R}^0 \) and \( \|v\|_\infty = 0 \); hence we have the inequality.

Let \( \hat{\sigma}^c \geq 1 \). For \( v \not\in K_{\hat{\sigma}} \), there at least one index \( i \in \hat{\sigma}^c \) such that \( v[i] \neq 0 \); hence \( \|v_{\hat{\sigma}}\|_0 \geq 1 \).

The definition of \( \rho \) in (18) shows that

\[
v \in B_{\infty}(0, \rho) \setminus K_{\hat{\sigma}} \Rightarrow -\|v\|_\infty > -\rho > -\frac{\beta}{2\|A^T(A\hat{u} - d)\|_1 + 1}
\]

\[
\Rightarrow -2\|v\|_\infty \|A^T(A\hat{u} - d)\|_1 + \beta \|v_{\hat{\sigma}}\|_0 > -\frac{2\beta\|A^T(A\hat{u} - d)\|_1}{2\|A^T(A\hat{u} - d)\|_1 + 1} + \beta > 0
\]

Introducing the last inequality into (56) shows that for \( \hat{\sigma}^c \geq 1 \), the inequality in (ii) is strict.

8.2 Proof of Proposition 4

If \( \hat{u} = 0 \), the statement is obvious. We focus on \( \hat{u} \neq 0 \). For an arbitrary \( i \in I_N \), define

\[
\hat{u}^{(i)} \overset{\text{def}}{=} (\hat{u}[1], \ldots, \hat{u}[i - 1], 0, \hat{u}[i + 1], \ldots, \hat{u}[N]) \in \mathbb{R}^N
\]

We shall use the equivalent formulation of \( \mathcal{F}_d \) given in (3). Clearly\(^{25}\)

\[
\mathcal{F}_d(\hat{u}) = \mathcal{F}_d(\hat{u}^{(i)} + e_\hat{u}[i]) = \|A\hat{u}^{(i)} + a_i\hat{u}[i] - d\|^2 + \beta \sum_{j \in \mathbb{N}} \phi(\hat{u}^{(i)}[j]) + \phi(\hat{u}[i])
\]

Consider \( f : \mathbb{R} \rightarrow \mathbb{R} \) as given below

\[
(57) \quad f(t) \overset{\text{def}}{=} \mathcal{F}_d(\hat{u}^{(i)} + e_i t)
\]

Since \( \hat{u} \) is a global minimizer of \( \mathcal{F}_d \), for any \( i \in I_N \), we have

\[
f(\hat{u}[i]) = \mathcal{F}_d(\hat{u}^{(i)} + e_i\hat{u}[i]) \leq \mathcal{F}_d(\hat{u}^{(i)} + e_i t) = f(t) \quad \forall t \in \mathbb{R}
\]

Equivalently, for any \( i \in I_N \), \( f(\hat{u}[i]) \) is the global minimum of \( f(t) \) on \( \mathbb{R} \). Below we will determine the global minimizer(s) \( \hat{t} = \hat{u}[i] \) of \( f \) as given in (57), i.e.,

\[
\hat{t} = \hat{u}[i] = \arg \min_{t \in \mathbb{R}} f(t)
\]

\(^{25}\)Using the definition of \( \hat{u}^{(i)} \), we have \( \hat{u}^{(i)} = A_{(I_N \setminus \{i\})}(\hat{u} \setminus \{i\}) \), hence \( A\hat{u}^{(i)} \) is independent of \( \hat{u}[i] \).
In detail, the function $f$ reads as

$$
\begin{align*}
    f(t) &= \|A\hat{u}^{(i)} + a_i t - d\|^2 + \beta \sum_{j \in \mathbb{I}_N} \phi \left( \hat{u}^{(i)}[j] \right) + \beta \phi(t) \\
    &= \|A\hat{u}^{(i)} - d\|^2 + \|a_i\|^2 t^2 + 2t\langle a_i, A\hat{u}^{(i)} - d \rangle + \beta \sum_{j \in \mathbb{I}_N} \phi \left( \hat{u}^{(i)}[j] \right) + \beta \phi(t) \\
    &= \|a_i\|^2 t^2 + 2t\langle a_i, A\hat{u}^{(i)} - d \rangle + \beta \phi(t) + C \, ,
\end{align*}
$$

where

$$
C = \|A\hat{u}^{(i)} - d\|^2 + \beta \sum_{j \in \mathbb{I}_N} \phi \left( \hat{u}^{(i)}[j] \right) .
$$

Note that $C$ does not depend on $t$. The function $f$ has two local minimizers denoted by $\hat{t}_0$ and $\hat{t}_1$. The first is

$$
\hat{t}_0 = 0 \quad \Rightarrow \quad f(\hat{t}_0) = C .
$$

The other one, $\hat{t}_1 \neq 0$, corresponds to $\phi(t) = 1$. From (58), $\hat{t}_1$ solves

$$
2\|a_i\|^2 t + 2\langle a_i, A\hat{u}^{(i)} - d \rangle = 0 .
$$

Recalling that $a_i \neq 0$, $\forall \, i \in \mathbb{I}_N$ (see (7)), it follows that

$$
\hat{t}_1 = -\frac{\langle a_i, A\hat{u}^{(i)} - d \rangle}{\|a_i\|^2} \quad \Rightarrow \quad f(\hat{t}_1) = \frac{-\langle a_i, A\hat{u}^{(i)} - d \rangle^2}{\|a_i\|^2} + \beta + C .
$$

Next we check whether $\hat{t}_0$ or $\hat{t}_1$ is a global minimizer of $f$. From (59) and (60) we get

$$
f(\hat{t}_0) - f(\hat{t}_1) = \frac{\langle a_i, A\hat{u}^{(i)} - d \rangle^2}{\|a_i\|^2} - \beta .
$$

Furthermore,

$$
f(\hat{t}_0) < f(\hat{t}_1) \quad \Rightarrow \quad \hat{u}[i] = \hat{t}_0 = 0 ,
$$

$$
f(\hat{t}_1) < f(\hat{t}_0) \quad \Rightarrow \quad \hat{u}[i] = \hat{t}_1 = -\frac{\langle a_i, A\hat{u}^{(i)} - d \rangle}{\|a_i\|^2} ,
$$

$$
f(\hat{t}_0) = f(\hat{t}_1) \quad \Rightarrow \quad \hat{t}_0 \text{ and } \hat{t}_1 \text{ are global minimizers of } f .
$$

In particular, we have

$$
f(\hat{t}_1) \leq f(\hat{t}_0) \quad \iff \quad \langle a_i, A\hat{u}^{(i)} - d \rangle^2 \geq \beta \|a_i\|^2
$$

$$
\left[ \text{by (61)} \right] \quad \Rightarrow \quad |\hat{u}[i]| = \left| \frac{\langle a_i, A\hat{u}^{(i)} - d \rangle}{\|a_i\|^2} \right| .
$$

$$
\left[ \text{by (62)} \right] \quad \geq \sqrt{\beta} \frac{\|a_i\|}{\|a_i\|^2} = \frac{\sqrt{\beta}}{\|a_i\|} .
$$

It is clear that the conclusion holds true for any $i \in \mathbb{I}_N$. 

29
8.3 Proof of Proposition 5

The asymptotic function \((F_d)_\infty (v)\) of \(F_d\) can be calculated according to \(^{26}\) [3, Theorem 2.5.1]

\[
(F_d)_\infty (v) = \liminf_{v' \to v} \frac{F_d(tv')}{t}
\]

Then

\[
(F_d)_\infty (v) = \liminf_{v' \to v} \frac{\|Av' - d\|^2 + \beta \|v'\|_0}{t}
\]

\[
= \liminf_{v' \to v} \left( t\|Av'\|^2 - 2\langle d, Av' \rangle + \frac{\|d\|^2 + \beta \|v'\|_0}{t} \right)
\]

\[
= \begin{cases} 
0 & \text{if } v \in \text{ker}(A), \\
+\infty & \text{if } v \notin \text{ker}(A).
\end{cases}
\]

Hence

\[
\text{ker}((F_d)_\infty) = \text{ker}(A),
\]

where \(\text{ker}((F_d)_\infty) = \{v \in \mathbb{R}^N : (F_d)_\infty (v) = 0\}\).

Let \(\{v_k\}\) satisfy \(^{34}\) with \(v_k \|v_k\|^{-1} \to \bar{v} \in \text{ker}(A)\). Below we compare the numbers \(\|v_k\|_0\) and \(\|v_k - \rho \bar{v}\|_0\) where \(\rho > 0\). There are two options.

1. Consider that \(i \in \sigma(\bar{v})\), that is, \(\bar{v}[i] = \lim_{k \to \infty} v_k[i] \|v_k\|^{-1} \neq 0\). Then \(\|v_k[i]\| > 0\) for all but finitely many \(k\) as otherwise, \(v_k[i] \|v_k\|^{-1}\) would converge to 0. Therefore, there exists \(k_i\) such that

\[
|v_k[i] - \rho \bar{v}[i]| \geq 0 \quad \text{and} \quad |v_k[i]| > 0 \quad \forall k \geq k_i.
\]

2. If \(i \in (\sigma(\bar{v}))^c\), i.e. \(\bar{v}[i] = 0\), then clearly

\[
v_k[i] - \rho \bar{v}[i] = v_k[i].
\]

Combining \(^{64}\) and \(^{65}\), the definition of \(\| \cdot \|_0\) using \(\phi\) in \(^{2}\) shows that

\[
\|v_k - \rho \bar{v}\|_0 \leq \|v_k\|_0 \quad \forall k \geq k_0 \overset{\text{def}}{=} \max_{i \in \sigma(\bar{v})} k_i.
\]

By \(^{63}\), \(A\bar{v} = 0\). This fact, jointly with \(^{66}\), entails that

\[
F_d(v_k - \rho \bar{v}) = \|A(v_k + \rho \bar{v}) - d\|^2 + \beta \|v_k - \rho \bar{v}\|_0
\]

\[
= \|Av_k - d\|^2 + \beta \|v_k - \rho \bar{v}\|_0
\]

\[
\leq \|Av_k - d\|^2 + \beta \|v_k\|_0 = F_d(v_k) \quad \forall k \geq k_0.
\]

It follows that for any \(k \geq k_0\) we have

\[
v_k \in \text{lev} (F_d, \lambda_k) \quad \Rightarrow \quad v_k - \rho \bar{v} \in \text{lev} (F_d, \lambda_k),
\]

and thus \(F_d\) satisfies Definition \(^{5}\).

\(^{26}\)In the nonconvex case, the notion of asymptotic functions and the representation formula were first given by J.P. Dedieu \(^{12}\).
8.4 Proof of Proposition 6

Given $K \in \mathbb{I}_{M-1}$, set

$$U_{K+1} \overset{\text{def}}{=} \bigcup_{\omega \subset \mathbb{I}_N} \{ \bar{u} : \bar{u} \text{ solves } (P_{\omega}) \text{ and } \|\bar{u}\|_0 \geq K + 1 \} .$$

- Let $U_{K+1} \neq \emptyset$. By Proposition 1 for any $\beta > 0$, $F_d$ has a (local) minimum at each $\bar{u} \in U_{K+1}$. Thus

$$\bar{u} \text{ is a (local) minimizer of } F_d \text{ and } \|\bar{u}\|_0 \geq K + 1 \iff \bar{u} \in U_{K+1} .$$

Then for any $\beta > 0$

$$F_d(\bar{u}) \geq \beta(K + 1) \ \forall \bar{u} \in U_{K+1} .$$

Let $\bar{u}$ be defined by

$$\bar{u} \text{ solves } (P_{\omega}) \text{ for some } \omega \in \Omega_K .$$

Then

$$\|\bar{u}\|_0 \leq K .$$

Set $\beta$ and $\beta_K$ according to

$$\beta > \beta_K \overset{\text{def}}{=} \|A\bar{u} - d\|^2 .$$

For such a $\beta$ we have

$$F_d(\bar{u}) = \|A\bar{u} - d\|^2 + \beta\|\bar{u}\|_0$$

$$\leq \beta + \beta K = \beta(K + 1)$$

$$\leq F_d(\bar{u}) \ \forall \bar{u} \in U_{K+1} .$$

Let $\hat{u}$ be a global minimizer of $F_d$. Then

$$F_d(\hat{u}) \leq F_d(\bar{u}) < F_d(\bar{u}) \ \forall \bar{u} \in U_{K+1} .$$

Using (67)-(68), we find

$$\|\hat{u}\|_0 \leq K .$$

- $U_{K+1} = \emptyset$ entails that

$$\bar{u} \text{ solves } (P_{\omega}) \text{ for } \omega \subset \mathbb{I}_N, \ \#\omega \geq K + 1 \Rightarrow \|\bar{u}\|_0 \leq K .$$

Let $\hat{u}$ be a global minimizer of $F_d$. By (72) we have

$$\|\hat{u}\|_0 \leq K .$$

According to Theorem 2(ii), any global minimizer of $F_d$ is strict, hence $\sigma(\hat{u}) \in \mathbb{T}_K .$

---

27 Such a $\tilde{u}$ always exists, see subsection 1.1. By Proposition 1 and Theorem 1, it is uniquely defined.
28 Let $A = (e_1, e_2, e_3, e_4, e_1) \in \mathbb{R}^{4\times 5}$ and $d = e_1 \in \mathbb{R}^4$. For $K = M - 1 = 3$ one can check that $U_{K+1} = \emptyset$. 

---

31
Acknowledgments

The author would like to thank the anonymous Reviewer 2 for very helpful remarks and suggestions.

References


33


