C. R. Acad. Sci. Paris, Ser. I 349 (2011) 1145–1150

Contents lists available at SciVerse ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Mathematical Analysis

Solve exactly an under determined linear system by minimizing least squares regularized with an ℓ_0 penalty

Résoudre exactement un système sous-déterminé en minimisant des moindres carrés régularisés avec ℓ_0

Mila Nikolova

CMLA, ENS Cachan, CNRS, UniverSud, 61, avenue President Wilson, 94230 Cachan, France

ARTICLE INFO

Article history:
Received 23 May 2011
Accepted after revision 18 August 2011
Available online 5 October 2011

Presented by Yves Meyer

ABSTRACT

We analyze objectives $\mathcal{F}_d$ combining a quadratic data-fidelity and a weighted ℓ_0 penalty. Data d are generated using a full column rank $M \times N$ matrix A with $N > M$. We provide a detailed analysis of the minimization problem. We exhibit a criterion enabling to recover exactly an original vector $\ddot{u}$ with support shorter than $M - 1$ as a strict (local) minimizer of $\mathcal{F}_d$ where $d = A\ddot{u}$.

RÉSUMÉ

Nous analysons des objectifs $\mathcal{F}_d$ combinant une fidélité aux données quadratique et une pénalisation ℓ_0. Les données d sont générées par une matrice A de dimension $M \times N$ et de rang M où $N > M$. Nous donnons une analyse détaillée du problème de minimisation. Nous établissons un critère permettant de retrouver un vecteur original $\ddot{u}$ dont la longueur du support ne dépasse pas $M - 1$ comme un minimiseur (local) strict de $\mathcal{F}_d$ où $d = A\ddot{u}$.

© 2011 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

Version française abrégée

Introduction

La fonction objectif $\mathcal{F}_d$ est de la forme (1) où $A \in \mathbb{R}^{M \times N}$, for $M < N$, est de rang M, $\|\cdot\|$ est la norme ℓ_2, d sont les données et β est un paramètre de régularisation. Nous analysons les minimiseurs (locaux) $\hat{u}$ de $\mathcal{F}_d$, voir (2). Rappelons que le problème numérique est combinatoire, de très haute complexité. On minimise $\mathcal{F}_d$ de la forme (1) en traitement du signal et de l'image, en problèmes inverses, en apprentissage, en classification, en sélection de modèles, en échantillonnage compressé, en sélection de sous-ensembles, en compression de données, etc. Il faut noter qu'il n'y a pas d'équivalence entre le problème (1)–(2) et ses variantes contraintes formulées dans (3). La raison en est que tous ces problèmes sont non convexes.

Notations

Nous dénotons $\mathbb{I}_K = \{1, \ldots, K\}$. La cardinalité d'un sous-ensemble $\omega \subset \mathbb{I}_K$ est notée par $\sharp\omega$. Si $\omega \subset \mathbb{I}_K$ alors $\omega^c = \mathbb{I}_K \setminus \omega$. D'autres notations importantes sont indiquées dans les équations (4), (5) et (6). Noter que A_ω^T désigne la transposée de A_ω.

E-mail address: nikolova@cmla.ens-cachan.fr.

1631-073X/$ – see front matter © 2011 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.crma.2011.08.011

On dit qu'un sous-ensemble $\mathcal{S} \subset \mathbb{R}^K$ est *négligeable* s'il existe un fermé $\mathcal{Z} \subset \mathbb{R}^K$ de mesure de Lebesgue dans $\mathbb{R}^K$ $\mathbb{L}^K(\mathcal{Z}) = 0$, tel que $\mathcal{S} \subseteq \mathcal{Z}$. Si une propriété est vraie pour tout $v \in \mathbb{R}^K \setminus \mathcal{S}$, nous dirons que cette propriété est *générique* sur $\mathbb{R}^K$.

Tous les minimiseurs de $\mathcal{F}_d$

La section 2 offre une caractérisation générale des minimiseurs (locaux) de $\mathcal{F}_d$.

On y établit que si $\hat{u} \in \mathbb{R}^N$ résout un problème de la forme $(\mathcal{P}_\omega)$, alors $\mathcal{F}_d$ atteint un minimum (local) en $\hat{u}$, pour tout $\beta > 0$ (Théorème 1). La réciproque – que chaque minimiseur $\hat{u}$ de $\mathcal{F}_d$ résout le problème formulé dans $(\mathcal{P}_\omega)$ pour $\omega = \hat{\sigma}$ où $\hat{\sigma}$ est le support de $\hat{u}$ – est bien intuitive (Lemme 1). Ainsi a-t-on une équivalence entre la minimisation de $\mathcal{F}_d$ et la résolution de $(\mathcal{P}_\omega)$ pour des supports différents $\omega \subset \mathbb{I}_N$. Cette équivalence sous-tend la majorité des résultats présentés dans cette Note. Le Corollaire 1 nous dit qu'un minimiseur $\hat{u}$ de $\mathcal{F}_d$ satisfait le problème linéaire formulé dans (7).

Une conséquence en est qu'un minimiseur (local) $\hat{u}$ de $\mathcal{F}_d$ peut supprimer du bruit ou trouver une approximation parcimonieuse selon un schéma qui peut ressembler au seuillage dur. En présence de bruit, les coefficients non nuls peuvent être fortement bruités si $A_{\hat{\sigma}}^T A_{\hat{\sigma}}$ est mal conditionnée. La qualité du résultat dépend notablement du minimiseur local sélectionné.

Les minimiseurs stricts de $\mathcal{F}_d$

Cette question est étudiée dans la section 3. La condition nécessaire et suffisante pour qu'un minimum (local) soit strict est formulée dans (8), Théorème 2. La Proposition 1 est constructive puisqu'elle fournit des directions permettant de s'échapper d'un minimiseur non strict. Les minimiseurs globaux, dont l'existence est prouvée, sont stricts (Théorème 3).

L'ensemble $\Omega_{\max}$ dans la Définition 2 est composé de tous les sous-ensembles d'indices de colonnes de A tels que la sous matrice correspondante est de rang plein. Cet ensemble décrit tous les minimiseurs (locaux) stricts de $\mathcal{F}_d$.

Le Théorème 4 stipule que chaque minimiseur strict (local) de $\mathcal{F}_d$ est produit par une fonction minimiseur local $\mathcal{U}$ qui est linéaire sur $\mathbb{R}^M$. D'où la continuité des minimiseurs (locaux) stricts par rapport à d.

Chaque minimiseur global satisfait une condition nécessaire formulée dans (10), Théorème 5.

Une méthodologie pour trouver la solution exacte comme un minimiseur (local) de $\mathcal{F}_d$

Dans la section 4, les données sont de la forme $d = A\ddot{u}$. Si $\sharp\sigma_{\ddot{u}} \geqslant M$, on ne peut pas retrouver $\ddot{u}$. On ne considère donc que des vecteurs originaux dont les supports sont dans Ω (*cf.* Définition 2) et des données d sans bruit. On démontre que la solution de n'importe quel des problèmes formulés dans (11) fournit le vecteur d'origine $\hat{u} = \ddot{u}$ (Théorème 2). Cependant, il faut connaître un $\omega \in \Omega$ tel que $\ddot{\varpi} = \sigma_{\ddot{u}} \subseteq \omega$, ce qui n'est pas réaliste. Afin de trouver la solution exacte parmi tous les minimiseurs (locaux) de $\mathcal{F}_d$ de support dans Ω, on adopte une hypothèse sur A formulée dans H1. *Cette hypothèse n'est mise en défaut que pour un sous-ensemble négligeable de matrices A* – voir la Proposition 3.

Les vecteurs originaux dont le support est de longueur $r \leqslant M-1$ que nous ne saurons pas trouver en minimisant $\mathcal{F}_d$ sont contenus dans l'ensemble Θ_r, équation (13), qui est *négligeable* dans $\mathbb{R}^r$ (Lemme 3). Un original $\ddot{u}$ avec $\sharp\sigma_{\ddot{u}} = r \leqslant M-1$ satisfait génériquement $\ddot{u}_{\sigma_{\ddot{u}}} \in \mathbb{R}^r \setminus \Theta_r$. La Proposition 4 affirme que si H1 est satisfaite, alors tous les minimiseurs dont les supports ω sont dans Ω, mais ne contiennent pas le support original $\ddot{\varpi}$, produisent un terme d'erreur positif. Le Théorème 6 stipule que le minimiseur $\hat{u}$ de $\mathcal{F}_d$ qui coïncide avec le vecteur original $\ddot{u}$ résout (14). Donc *parmi tous les minimiseurs stricts dont la longueur du support est au maximum $M-1$*, on a $\|A\hat{u} - d\|^2 = 0$ uniquement pour $\hat{u} = \ddot{u}$.

Exemple numérique

Un exemple numérique est proposé dans la section 5. On considère l'objectif $\mathcal{F}_d$ dans (1) pour A et d spécifiés dans (15). Par recherche combinatoire exhaustive, nous avons vérifié que A satisfait l'hypothèse H1 et calculé les minimiseurs locaux de $\mathcal{F}_d$ pour différentes valeurs de β. Pour $\beta = 0.01$ le vecteur original $\ddot{u}$ est le minimiseur global de $\mathcal{F}_d$. Pour $\beta = 10^5 \geqslant \|d\|^2$ le minimiseur global de $\mathcal{F}_d$ est $\hat{u} = 0$, ce qui confirme le Lemme 2. Dans tous les cas, quelle que soit la valeur de β, on trouve que le minimiseur de $\mathcal{F}_d$ qui coïncide avec l'original $\hat{u} = \ddot{u}$ est l'unique minimiseur strict tel que $\|A\hat{u} - d\|^2 = 0$ parmi tous les minimiseurs stricts dont la longueur du support est inférieure ou égale à $M-1$. Cela corrobore le Théorème 6.

Conclusions et perspectives

Résoudre l'équation normale issue d'une sous matrice de A à M lignes arbitraire fournit un minimiseur (local) de $\mathcal{F}_d$. On sait reconnaître les minimiseurs (locaux) *stricts* de $\mathcal{F}_d$. Ses minimizeurs globaux de $\mathcal{F}_d$ sont stricts. Le Théorème 6 utilise une hypothèse sur A qui est *génériquement* vraie. Il fournit un critère confortable permettant de savoir si un minimiseur (local) $\hat{u}$ de $\mathcal{F}_d$ est la solution exacte à partir de données $d = A\ddot{u}$. Sous les conditions du Théorème 6, un vecteur $\ddot{u}$ avec $\|\ddot{u}\|_0 \leqslant M-1$ est aussi l'unique solution (globale) du problème (3)-(a). Cette question mérite une exploration approfondie. Les minimizeurs de $\mathcal{F}_d$ en présence de bruit ou en approximation cachent bien d'inconnues.

1. Introduction

Let $A \in \mathbb{R}^{M \times N}$ meet $\operatorname{rank} A = M < N$. Given a data vector $d \in \mathbb{R}^M$, consider the objective $\mathcal{F}_d : \mathbb{R}^N \to \mathbb{R}$

$$\mathcal{F}_d(u) = \|Au - d\|^2 + \beta \|u\|_0, \quad \beta > 0, \quad \text{where } \|u\|_0 \stackrel{\text{def}}{=} \sharp\{i \in \mathbb{I}_N \colon u[i] \neq 0\} \quad \text{for } \mathbb{I}_N \stackrel{\text{def}}{=} \{1, \dots, N\}. \tag{1}$$

In this Note, $\|\cdot\|$ is the ℓ_2-norm, β is a regularization parameter, $u[i]$ is the ith entry of a vector $u \in \mathbb{R}^N$ and $\sharp$ stands for cardinality. The columns of A are denoted by a_i, $\forall i \in \mathbb{I}_N$. It is assumed that $a_i \neq 0$ for all $i \in \mathbb{I}_N$. We focus on the (local) minimizers $\hat{u}$ of $\mathcal{F}_d$:

$$\hat{u} \in \mathbb{R}^N \quad \text{such that} \quad \mathcal{F}_d(\hat{u}) = \min_{u \in \mathcal{O}} \mathcal{F}_d(u), \quad \hat{u} \in \mathcal{O} \subset \mathbb{R}^N, \ \mathcal{O} \text{ is open}. \tag{2}$$

Finding the global minimizer of $\mathcal{F}_d$ must be an *NP-hard* computational problem [3,13]. The ℓ_0 pseudo-norm in (1) served as a regularizer for a long time. The problem stated in (1)–(2) arises in image processing, data compression, sparse approximation, compressive sensing, machine learning, model selection, subset selection, *etc*. The hard-thresholding method [4] amounts to minimize (1) where A is a wavelet basis. When $M < N$ various (usually strong) conditions on $\|u\|_0$ and on A (e.g. RIP like criteria), 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, see among others [10,6–8,1,12,5].

Our goal is to analyze the (local) minimizers $\hat{u}$ of objective functions of the form (1). We provide a detailed analysis of the minimization problem. A point of attention is the possibility to recover (exactly) an original $(M-1)$-sparse[1] $\ddot{u} \in \mathbb{R}^N$, where $N > M$, from data $d = A\ddot{u} \in \mathbb{R}^M$ as a strict (local) minimizer $\hat{u}$ of $\mathcal{F}_d$.

The optimization problem (1)–(2) might seem close to its constraint variants: given $\varepsilon \geqslant 0$ or $K \in \mathbb{I}_M$,

$$\text{(a) minimize } \|u\|_0 \text{ subject to } \|Au - d\|^2 \leqslant \varepsilon \quad \text{or} \quad \text{(b) minimize } \|Au - d\|^2 \text{ subject to } \|u\|_0 \leqslant K. \tag{3}$$

The latter problems are abundantly studied in the context of sparse recovery in different fields. An excellent account is given in [2], see also the book [9]. We emphasize that in general, *there is no equivalence between the problems stated in* (3) *and the minimization of* $\mathcal{F}_d$ *in* (1). The reason is that all these problems are nonconvex.

1.1. Notations and definitions

A local minimizer $\hat{u}$ is *strict* if there is an open neighborhood $\mathcal{O} \subset \mathbb{R}^N$ of $\hat{u}$ such that $\mathcal{F}_d(\hat{u}) < \mathcal{F}_d(v)$ for any $v \in \mathcal{O}$. We use the notation $\mathbb{I}_K = \{1, \dots, K\}$. The *support* σ_u of any $u \in \mathbb{R}^N$ is defined as

$$\sigma_u \stackrel{\text{def}}{=} \{i \in \mathbb{I}_N \colon u[i] \neq 0\}, \quad \sigma_u[1] < \cdots < \sigma_u[\sharp\sigma_u]. \tag{4}$$

Remind that $\|u\|_0 = \sharp\sigma_u$. Given $\omega \subset \mathbb{I}_N$, we denote $\omega^c \stackrel{\text{def}}{=} \mathbb{I}_N \setminus \omega$. With any strictly increasing subsequence $\omega \subset \mathbb{I}_N$ and for any $u \in \mathbb{R}^N$ we associate the submatrix A_ω and the subvector u_ω given below

$$A_\omega \stackrel{\text{def}}{=} (a_{\omega[1]}, \dots, a_{\omega[\sharp\omega]}) \in \mathbb{R}^{M \times \sharp\omega} \quad \text{and} \quad u_\omega \stackrel{\text{def}}{=} \big(u\big[\omega[1]\big], \dots, u\big[\omega[\sharp\omega]\big]\big) \in \mathbb{R}^{\sharp\omega}, \tag{5}$$

as well as the zero padding operator $Z_\omega : \mathbb{R}^{\sharp\omega} \to \mathbb{R}^N$ defined by

$$u = Z_\omega(u_\omega), \qquad u[i] = \begin{cases} 0 & \text{if } i \notin \omega, \\ u_\omega[k] & \text{for the unique } k \text{ such that } \omega[k] = i. \end{cases} \tag{6}$$

For $\omega = \varnothing$ we have $u_\varnothing = \varnothing$ and $Z_\varnothing(u_\varnothing) = 0 \in \mathbb{R}^N$. We denote by A_ω^T the transposed of A_ω.

Definition 1. A subset $\mathcal{S} \subset \mathbb{R}^K$ is said to be *negligible* (in $\mathbb{R}^K$) if there exists $\mathcal{Z} \subset \mathbb{R}^K$ which is closed in $\mathbb{R}^K$, its Lebesgue measure in $\mathbb{R}^K$ is $\mathbb{L}^K(\mathcal{Z}) = 0$ and $\mathcal{S} \subseteq \mathcal{Z}$.

The chance that a random $v \in \mathbb{R}^K$, following a non-singular probability distribution, comes across such an $\mathcal{S}$ is null.
If a property holds true for all $v \in \mathbb{R}^K \setminus \mathcal{S}$, where $\mathcal{S}$ is negligible in $\mathbb{R}^K$, we say that this property is *generic* on $\mathbb{R}^K$.

2. All minimizers

First we show that finding a local minimizer of $\mathcal{F}_d$ is equivalent to solving a quadratic minimization problem.

[1] As usual, a vector u is said to be k-sparse if $\|u\|_0 \leqslant k$.

Theorem 1. *Given $d \in \mathbb{R}^M$ and an arbitrary $\omega \subset \mathbb{I}_N$, consider the problem $(\mathcal{P}_\omega)$ given below:*

$$\boxed{\min_{v \in \mathbb{R}^N} \|Av - d\|^2 \quad \textit{subject to} \quad v[i] = 0 \; \forall i \in \omega^c.} \qquad (\mathcal{P}_\omega)$$

Let $\hat{u}$ solve problem $(\mathcal{P}_\omega)$. Then for any $\beta > 0$, $\hat{u}$ is a (local) minimizer of $\mathcal{F}_d$ in (1) and $\sigma_{\hat{u}} \subseteq \omega$.

The reciprocal statement is quite obvious.

Lemma 1. *Let $\mathcal{F}_d$ have a (local) minimum at $\hat{u}$. Set $\hat{\sigma} \stackrel{\text{def}}{=} \sigma_{\hat{u}}$. Then $\hat{u}$ solves $(\mathcal{P}_\omega)$* for $\omega = \hat{\sigma}$.

Theorem 1 and Lemma 1 underly most of the results presented in this Note. Theorem 1 entails that for any $d \in \mathbb{R}^M$ and $\beta > 0$, the function $\mathcal{F}_d$ *does admit a global minimum.*

Corollary 1. *Let $\hat{u} \neq 0$ be a (local) minimizer of $\mathcal{F}_d$. Set $\hat{\sigma} = \sigma_{\hat{u}}$. Then $\hat{u}$ is given by*

$$\hat{u} = Z_{\hat{\sigma}}(\hat{u}_{\hat{\sigma}}) \quad \textit{where } \hat{u}_{\hat{\sigma}} \textit{ satisfies } A_{\hat{\sigma}}^T A_{\hat{\sigma}} \hat{u}_{\hat{\sigma}} = A_{\hat{\sigma}}^T d. \tag{7}$$

By Corollary 1, a minimizer $\hat{u}$ of $\mathcal{F}_d$ can perform denoising or sparse approximation following a pseudo-hard thresholding scheme. In denoising, the retained non-zero entries $\hat{u}[i]$, $\forall i \in \hat{\sigma}$, are noisy and their noise can be amplified if $A_{\hat{\sigma}}^T A_{\hat{\sigma}}$ is ill-conditioned. The quality of the result critically depends on the selected local minimizer.

3. The strict minimizers of the objective function

Lemma 2. *For any $d \in \mathbb{R}^M$ and $\beta > 0$, the objective $\mathcal{F}_d$ has a* strict *(local) minimum at $\hat{u} = 0$. If $\beta \geqslant \|d\|^2$, $\hat{u} = 0$ is a strict global minimizer of $\mathcal{F}_d$.*

We adopt the convention that $\operatorname{rank}(A_\varnothing) = 0$.

Theorem 2. *Let $\hat{u}$ be a (local) minimizer of $\mathcal{F}_d$. Set $\hat{\sigma} \stackrel{\text{def}}{=} \sigma_{\hat{u}}$. The minimizer $\hat{u}$ is* strict if and only if

$$\operatorname{rank} A_{\hat{\sigma}} = \sharp\hat{\sigma} \leqslant M. \tag{8}$$

If $\hat{u} \neq 0$ is a strict minimizer of $\mathcal{F}_d$, it reads $\hat{u} = Z_{\hat{\sigma}}(\hat{u}_{\hat{\sigma}})$ for $\hat{u}_{\hat{\sigma}} = (A_{\hat{\sigma}}^T A_{\hat{\sigma}})^{-1} A_{\hat{\sigma}}^T d$.

The next statement is constructive since it indicates *how to escape from a nonstrict local minimum.*

Proposition 1. *Given $d \in \mathbb{R}^M$ and $\beta > 0$, let $\bar{u} \neq 0$ be a local minimizer of $\mathcal{F}_d$ such that* $\operatorname{rank} A_{\bar{\sigma}} < \min\{M, \sharp\bar{\sigma}\}$ *where $\bar{\sigma} \stackrel{\text{def}}{=} \sigma_{\bar{u}}$. Then $\mathcal{F}_d$ has a* strict *minimizer $\hat{u}$ such that $\hat{\sigma} \stackrel{\text{def}}{=} \sigma_{\hat{u}} \subsetneqq \bar{\sigma}$, $A\hat{u} = A\bar{u}$ and $\mathcal{F}_d(\hat{u}) \leqslant \mathcal{F}_d(\bar{u}) - \beta$. More precisely, $\hat{u} \in \bar{u} - \mathrm{L}_{\bar{\sigma}}$ where* $\mathrm{L}_{\bar{\sigma}} = \{Z_{\bar{\sigma}}(v) \in \mathbb{R}^N\colon\ v \in \ker A_{\bar{\sigma}} \setminus \{0\}\}$.

Theorem 3. *For any $d \in \mathbb{R}^M$ and $\beta > 0$, each global minimizer $\hat{u}$ of $\mathcal{F}_d$* is strict, *and meets $\mathcal{F}_d(\hat{u}) \leqslant \beta M$ and $\|\hat{u}\|_0 \leqslant M$.*

We will focus on the strict (local) minimizers of $\mathcal{F}_d$. This motivates the definition given below.

Definition 2. For any $r \in \mathbb{I}_M$, Ω_r is the set of all r-length supports relevant to full rank $M \times r$ submatrices of A:

$$\Omega_r = \big\{\omega \subset \mathbb{I}_N\colon\ \sharp\omega = r = \operatorname{rank} A_\omega,\ \omega[1] < \cdots < \omega[r]\big\}. \tag{9}$$

Set $\Omega_0 = \varnothing$. Define as well $\Omega \stackrel{\text{def}}{=} \bigcup_{t=0}^{M-1} \Omega_t$ and $\Omega_{\max} \stackrel{\text{def}}{=} \Omega \cup \Omega_M$.

The set $\Omega_{\max}$ is the complete list of the supports of all possible strict (local) minimizers of $\mathcal{F}_d$.

Corollary 2. *For $d \in \mathbb{R}^M$ and $\beta > 0$, let $\hat{u}$ be a (local) minimizer of $\mathcal{F}_d$ satisfying $\hat{\sigma} \stackrel{\text{def}}{=} \sigma_{\hat{u}} \in \Omega$. Define $\mathrm{N}_{\hat{\sigma}} \stackrel{\text{def}}{=} \ker A_{\hat{\sigma}}^T \subset \mathbb{R}^M$ for $\hat{\sigma} \neq \varnothing$ and $\mathrm{N}_\varnothing = \mathbb{R}^M$. Then $\dim \mathrm{N}_{\hat{\sigma}} = M - \sharp\hat{\sigma} \geqslant 1$ and for any $d' \in \mathrm{N}_{\hat{\sigma}}$, the relevant $\mathcal{F}_{d+d'}$ has a strict minimizer at the same $\hat{u}$.*

Even though $\mathcal{F}_d$ can admit numerous strict (local) minimizers, they are continuous in d.

Definition 3. Let $\mathcal{O}$ be an open domain in $\mathbb{R}^M$. We say that $\mathcal{U} : \mathcal{O} \to \mathbb{R}^N$ is a *local minimizer function* for the family of functions $\mathcal{F}_{\mathcal{O}} \stackrel{\text{def}}{=} \{\mathcal{F}_d\colon\ d \in \mathcal{O}\}$ if $\mathcal{F}_d$ has a *strict* (local) minimum at $\mathcal{U}(d)$ for any $d \in \mathcal{O}$.

Theorem 4. *Let $\omega \in \Omega_{\max}$ and $\beta > 0$.*

(i) *The family $\mathcal{F}_{\mathbb{R}^M}$ has a* linear (*local*) *minimizer function $\mathcal{U} : \mathbb{R}^M \to \mathbb{R}^N$. If $\omega \neq \varnothing$, it reads*

$$\mathcal{U}(d) = Z_\omega(U_\omega d), \quad \textit{where } U_\omega = \left(A_\omega^T A_\omega\right)^{-1} A_\omega^T \in \mathbb{R}^{\sharp\omega \times M}.$$

If $\omega = \varnothing$, then $\mathcal{U}(d) = 0$. For any $d \in \mathbb{R}^M$, $\hat{u} = \mathcal{U}(d)$ is a strict (local) minimizer of $\mathcal{F}_d$ satisfying $\sigma_{\hat{u}} \subseteq \omega$.

(ii) *There exists a closed subset $\mathrm{D}_\omega \subset \mathbb{R}^M$ with $\mathbb{L}^M(\mathrm{D}_\omega) = 0$ such that for any $d \in \mathbb{R}^M \setminus \mathrm{D}_\omega$ we have $\sigma_{\hat{u}} = \omega$ where $\hat{u} = \mathcal{U}(d)$. Moreover, $d \mapsto \mathcal{F}_d(\mathcal{U}(d))$ is $\mathcal{C}^\infty$ on $\mathbb{R}^M \setminus \mathrm{D}_\omega$.*

The statement below provides a necessary condition for a global minimizer of $\mathcal{F}_d$ which is independent of d.

Theorem 5. *For $d \in \mathbb{R}^M$ and $\beta > 0$, let $\mathcal{F}_d$ in (1) reach a global minimum at $\hat{u}$. Then*

$$\textit{either } \hat{u}[i] = 0 \quad \textit{or} \quad \left|\hat{u}[i]\right| \geqslant \frac{\sqrt{\beta}}{\|a_i\|}, \quad \forall i \in \mathbb{I}_N. \tag{10}$$

4. The exact solution at a strict (local) minimum of $\mathcal{F}_d$

We have found [11] that if data read $d = A\ddot{u}$ for a $\ddot{u}$ with $\sigma_{\ddot{u}} \in \Omega_M$, then $\mathcal{F}_d$ *generically* has $\sharp\Omega_M$ strict minimizers satisfying $\mathcal{F}_d(\hat{u}) = \beta M$ and $A\hat{u} = d$. It is impossible to recover exactly a $\ddot{u}$ such that $\sharp\sigma_{\ddot{u}} \geqslant M$ as a minimizer of $\mathcal{F}_d$.

Corollary 3. *Let an original $\ddot{u} \neq 0$ with $\ddot{\varpi} \stackrel{\text{def}}{=} \sigma_{\ddot{u}}$ satisfy $\ddot{\varpi} \in \Omega_{\max}$ and there exists $i \in \ddot{\varpi}$ such that $|\ddot{u}[i]| < \beta/\|a_i\|$. Consider that $d = A\ddot{u}$. Then the exact solution $\hat{u} = \ddot{u}$ is a strict local minimizer of $\mathcal{F}_d$ such that $\mathcal{F}_d(\hat{u}) > \min_{v \in \mathbb{R}^N} \mathcal{F}_d(v)$.*

Next we focus on original vectors $\ddot{u}$ satisfying $\|\ddot{u}\|_0 \leqslant M - 1$ and data $d = A\ddot{u}$.

Proposition 2. *Consider an original $\ddot{u} \in \mathbb{R}^N$ such that $\ddot{\varpi} \stackrel{\text{def}}{=} \sigma_{\ddot{u}} \in \Omega$, where Ω is given in Definition 2. Let $d = A\ddot{u}$. Then for any $\beta > 0$, a strict (local) minimizer $\hat{u}$ of $\mathcal{F}_d$ satisfies $\hat{u} = \ddot{u}$ if and only if*

$$\omega \in \{\varpi \in \Omega_{\max}\colon\ \ddot{\varpi} \subseteq \varpi\} \quad \text{and} \quad \hat{u} \textit{ solves } (\mathcal{P}_\omega)\,. \tag{11}$$

Proposition 2 and Corollary 3 show that striving after global minimization of $\mathcal{F}_d$ can prevent exact recovery.

Any original $\ddot{u}$ with $\ddot{\varpi} \stackrel{\text{def}}{=} \sigma_{\ddot{u}} \in \Omega$ can be recovered *exactly* from data $d = A\ddot{u}$ by solving (11) *if we know in advance an* $\omega \in \Omega_{\max}$ such that $\ddot{\varpi} \subseteq \omega$. If such an ω is unavailable, and if $\mathcal{F}_d$ has other minimizers $\bar{u}$ meeting $\ddot{\varpi} \nsubseteq \sigma_{\bar{u}} \in \Omega$ and $\|A\bar{u} - d\|^2 = 0$, we cannot distinguish the exact solution. We adopt a weak assumption on A to remove such ambiguities.

H1. *For any $t \in \mathbb{I}_{M-1}$ and $r \in \mathbb{I}_t$ the matrix $A \in \mathbb{R}^{M \times N}$ in (1) satisfies*

$$(\varpi, \omega) \in (\Omega_r \times \Omega_t), \quad \varpi \nsubseteq \omega \quad \Rightarrow \quad A_\varpi^T A_\varpi \neq A_\varpi^T A_\omega \left(A_\omega^T A_\omega\right)^{-1} A_\omega^T A_\varpi.$$

Proposition 3. *For $t \in \mathbb{I}_{M-1}$ and $r \in \mathbb{I}_t$, define the following subset of matrices*:

$$\mathcal{A}_r(t) = \bigcup_{\omega \in \Omega_t} \left\{A_\varpi\colon\ \varpi \in \Omega_r, \varpi \nsubseteq \omega \text{ and } A_\varpi^T A_\varpi = A_\varpi^T A_\omega \left(A_\omega^T A_\omega\right)^{-1} A_\omega^T A_\varpi\right\}. \tag{12}$$

Then $\mathcal{A}_r(t)$ belongs to a closed subset in the space of all $M \times r$ matrices whose Lebesgue measure in this space is null.

From Proposition 3, assumption H1 *generically* holds true for all $M \times N$ real matrices A, see [11].

Let I_M denote the $M \times M$ identity matrix. For $r \in \mathbb{I}_{M-1}$ define

$$\Theta_r = \bigcup_{\varpi \in \Omega_r} \bigcup_{t=1}^{M-1} \bigcup_{\omega \in \Omega_t} \left\{v \in \mathbb{R}^r\colon\ v^T A_\varpi^T \left(I_M - A_\omega \left(A_\omega^T A_\omega\right)^{-1} A_\omega^T\right) A_\varpi v = 0, (\varpi, \omega) \in (\Omega_r \times \Omega_t) \text{ and } \varpi \nsubseteq \omega\right\}. \tag{13}$$

For the original vectors $\ddot{u}$ with $\ddot{u}_{\sigma_{\ddot{u}}} \in \Theta_r$, $r \in \mathbb{I}_{M-1}$ we cannot catch the exact solution at a minimum of $\mathcal{F}_d$.

Lemma 3. *Let* H1 *hold. For any* $r \in \mathbb{I}_{M-1}$*, the set* Θ_r *in* (13) *is closed in* $\mathbb{R}^r$ *and meets* $\mathbb{L}^r(\Theta_r) = 0$.

Since Θ_r in (13) is *negligible*, *there is no chance* that an original $\ddot{u}$ with $\sigma_{\ddot{u}} \in \Omega_r$ meets $\ddot{u}_{\sigma_{\ddot{u}}} \in \Theta_r$.

Proposition 4. *Let A satisfy* H1. *For* $\ddot{u} \in \mathbb{R}^N$*, denote* $\ddot{\varpi} \stackrel{\text{def}}{=} \sigma_{\ddot{u}}$ *and* $r \stackrel{\text{def}}{=} \sharp\ddot{\varpi}$*. Assume that* $\ddot{\varpi} \in \Omega$*,* $r \geqslant 1$ *and* $\ddot{u}_{\ddot{\varpi}} \in \mathbb{R}^r \setminus \Theta_r$*. Let* $d = A\ddot{u}$*. If* $\omega \in \{\varpi \in \Omega\colon\ \ddot{\varpi} \nsubseteq \varpi\}$ *and* $\hat{u}$ *solves* $(\mathcal{P}_\omega)$*, then* $\|A\hat{u} - d\|^2 > 0$.

Next we show how to recognize the exact solution of our linear system *as a strict* (*local*) *minimizer of* $\mathcal{F}_d$.

Theorem 6. *Let* H1 *hold. For* $\ddot{u} \in \mathbb{R}^N$*, put* $\ddot{\varpi} \stackrel{\text{def}}{=} \sigma_{\ddot{u}}$ *and* $r \stackrel{\text{def}}{=} \sharp\ddot{\varpi}$*. Assume that* $\ddot{\varpi} \in \Omega$ *and* $\ddot{u}_{\ddot{\varpi}} \in \mathbb{R}^r \setminus \Theta_r$ *if* $r \geqslant 1$*. Let* $d = A\ddot{u}$*. Then* $\forall \beta > 0$*, a strict* (*local*) *minimizer* $\hat{u}$ *of* $\mathcal{F}_d$ *meets* $\hat{u} = \ddot{u}$ if and only if $\hat{u}$ *solves the* (*nonlinear*) *problem*

$$\min_{\omega \in \Omega}\left\{\|A\bar{u} - d\|^2\colon\ \omega \in \Omega \text{ and } \bar{u} \text{ solves } (\mathcal{P}_\omega)\right\}. \tag{14}$$

Among all strict (local) minimizers $\bar{u}$ of $\mathcal{F}_d$ with support $\sigma_{\bar{u}}$ no longer than $M-1$, we have $\|A\bar{u} - d\|^2 = 0$ only for $\hat{u} = \ddot{u}$. This result provides a safe criterion enabling to know whether or not an algorithm finds the exact solution.

5. Numerical toy-example

A is an $M \times N$ matrix for $M = 5$ and $N = 10$. We consider $\mathcal{F}_d$ in (1) for

$$A = \begin{bmatrix} 3 & 5 & 7 & 9 & 8 & 4 & 4 & 3 & 1 & 2 \\ 7 & 9 & 3 & 5 & 3 & 2 & 8 & 7 & 1 & 6 \\ 6 & 4 & 5 & 2 & 8 & 3 & 6 & 7 & 5 & 5 \\ 2 & 6 & 7 & 2 & 3 & 6 & 5 & 4 & 8 & 1 \\ 2 & 3 & 9 & 3 & 9 & 5 & 9 & 6 & 9 & 4 \end{bmatrix} \quad d = A\ddot{u}, \quad \text{where } \ddot{u} = [0 \ \ 0 \ \ 0 \ \ 0 \ \ 9 \ \ 8 \ \ 1 \ \ 1 \ \ 0 \ \ 0]^T. \tag{15}$$

Using exhaustive combinatorial search, we check that the arbitrary matrix A satisfies H1 and compute the minimizers of $\mathcal{F}_d$. The global minimizer is $\hat{u} = \ddot{u}$ for $\beta = 0.01$ and it reads $\hat{u} = 0$ for $\beta = 10^5 \geqslant \|d\|^2$ (see Lemma 2). In all cases, $\hat{u} = \ddot{u}$ is *the unique strict* (*local*) *minimizer of* $\mathcal{F}_d$ *with* $\|\hat{u}\|_0 \leqslant M - 1$ *yielding* $\|A\hat{u} - d\|^2 = 0$. This confirms Theorem 6.

6. Some conclusions and perspectives

The normal equation relevant to any M-row submatrix of A yields a (local) minimizer of $\mathcal{F}_d$. We learned how to recognize a *strict* (local) minimizer of $\mathcal{F}_d$ and that its global minimizers are strict. Strict local minimizers are linear in data. Theorem 6 uses a *generically true* assumption on A. It provides a safe condition for perfect identification of the exact $(M-1)$-sparse solution at a strict (local) minimum of $\mathcal{F}_d$ from noise-free data. This can help the conception of new algorithms for exact recovery.

Under the conditions of Theorem 6, an original $(M-1)$-sparse vector $\ddot{u}$ is also the unique (global) solution to the problem $\min_{u \in \mathbb{R}^N} \|u\|_0$ subject to $Au = d$ (see [11]). This question deserves a deeper exploration.

The behavior of the minimizers of $\mathcal{F}_d$ for noisy data or in sparse approximation hides open questions.

Acknowledgements

The author would like to thank the anonymous referee for her/his constructive remarks.

References

[1] T. Blumensath, M. Davies, Iterative thresholding for sparse approximations, Journal of Fourier Analysis and Applications 14 (4–6) (2008) 629–654.
[2] A.M. Bruckstein, D.L. Donoho, M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, SIAM Review 51 (1) (2009) 34–81.
[3] G. Davis, S. Mallat, M. Avellaneda, Adaptive greedy approximations, Constructive Approximation 13 (1) (1997) 57–98.
[4] D.L. Donoho, I.M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika 81 (3) (1994) 425–455.
[5] G. Gasso, A. Rakotomamonjy, S. Canu, Recovering sparse signals with a certain family of non-convex penalties and DC programming, IEEE Transactions on Signal Processing 57 (12) (2009) 4686–4698.
[6] J. Haupt, R. Nowak, Signal reconstruction from noisy random projections, IEEE Transactions on Information Theory 52 (9) (2006) 4036–4048.
[7] Y. Liu, Y. Wu, Variable selection via a combination of the ℓ_0 and ℓ_1 penalties, Journal of Computational and Graphical Statistics 16 (4) (2007) 782–798.
[8] J. Lv, Y. Fan, A unified approach to model selection and sparse recovery using regularized least squares, The Annals of Statistics 37 (6A) (2009).
[9] S. Mallat, A Wavelet Tour of Signal Processing (The Sparse Way), 3rd ed., Academic Press, London, 2008.
[10] A.J. Miller, Subset Selection in Regression, Chapman and Hall, London, UK, 2002.
[11] M. Nikolova, On the minimizers of least squares regularized with ℓ_0 norm, Technical report, 2011.
[12] J. Neumann, C. Schörr, G. Steidl, Combined SVM-based feature selection and classification, Machine Learning 61 (2005) 129–150.
[13] J.A. Tropp, Just relax: convex programming methods for identifying sparse signals in noise, IEEE Transactions on Information Theory 52 (3) (2006) 1030–1051.