The inverse problem in compressed sensing

Gianluca Giacchi

March 10, 2022

Abstract

Compressed Sensing (CS) is a well established theory involving many areas of modern research. It concerns a family of theoretical techniques and numerical algorithms aimed to recover sparse signals by a partial knowledge of their coefficients. This process is carried out through minimization problems involving the ℓ^1-norm, that has the property to be convex, while enforcing sparsity. This work aims to provide the fundamental theory of optimization that lies beneath the many CS applications, with particular regard to convex minimization and Lagrange theory.

Contents

1 Preliminaries

In this work, $\mathbb{C}$ denotes the complex field. If $z \in \mathbb{C}$, then $\Re(z)$ and $\Im(z)$ denote respectively the real part and the imaginary part of z. For all $n \geq 1$, $\mathbb{C}^n$ is a $2n$-dimensional real vector space with real inner product defined by

$$\Re\langle z, w\rangle := \Re(z^H w) = \Re\left(\sum_{j=1}^{n} \overline{z}_j w_j\right),$$

where $\overline{z}$ denotes the complex conjugate of $z = (z_1, \ldots, z_n)$ and z^H its conjugate-transpose. The norm induced by $\Re\langle\cdot,\cdot\rangle$ is the ℓ_n^2-norm on $\mathbb{C}^n$, i.e.

$$\|z\|_2 := (\Re\langle z, z\rangle)^{1/2},$$

which coincides with usual definition of ℓ_n^2-norm on $\mathbb{C}^n$ [1] [2].

Under these assumptions, all the results stated in this work in the complex framework also hold replacing $\mathbb{C}$ with $\mathbb{R}$. We denote the canonical inner product on $\mathbb{R}^n$ with

$$\langle x, y\rangle = x^T y := \sum_{j=1}^{n} x_j y_j,$$

which coincides with the restriction of $\Re\langle\cdot,\cdot\rangle$ to $\mathbb{R}^n$.

We consider constrained optimization problems in the form

$$\begin{cases} \min_{z\in\mathbb{C}^n} F_0(z), \\ Az = y, \\ F_l(z) \leq b_l \qquad l = 1, \ldots, M, \end{cases} \tag{1}$$

where $A \in \mathbb{C}^{m\times n}$, $y \in \mathbb{C}^n$ and $F_0, F_1, \ldots, F_M : \mathbb{C}^n \to (-\infty, +\infty]$. Throughout this work, we always assume $m \leq n$. A point $z \in \mathbb{C}^n$ is called **feasible** if it belongs to the constraint, that is if

$$z \in K := \left\{\zeta \in \mathbb{C}^n \;:\; A\zeta = y \textit{ and } F_l(\zeta) \leq b_l,\; l = 1, \ldots, M\right\}$$

and K is called the **set of feasible points**. To avoid triviality, we always assume $K \neq \varnothing$, in the which case the problem (1) is called **feasible**. In view of the definition of K, the problem (1) can be implicitly written as

$$\min_{z\in K} F_0(z).$$

[1]More precisely, if $\langle z, w\rangle = \sum_{j=1}^{n} \overline{z_j} \cdot w_j$ denotes the canonical inner product on $\mathbb{C}^n$, then $\langle z, z\rangle \geq 0$ for all $z \in \mathbb{C}^n$, so that $\Re\langle z, z\rangle = \langle z, z\rangle$

[2]Observe that $\Re\langle\cdot,\cdot\rangle$ coincides with the real part of $\langle\cdot,\cdot\rangle$.

Remark 1.1. In many of the applications, K is bounded and the functions $F_1, \dots, F_M$ are continuous, so that K is also a closed (and non-empty by our assumption) subset of $\mathbb{C}^n$ and, therefore, it is compact. If also F_0 is continuous and defined on $\mathbb{C}^n$, then the set $\{z \in \mathbb{C}^n \ : \ z \textit{ minimizes problem } (1)\}$ is non-empty by Weierstrass Theorem. In particular, $p^* := \inf_{z\in K} F_0(z)$ is attained.

In what follows, we will use the geometric version of Hahn-Banach Theorem, which states the fact that disjoint convex subset of $\mathbb{C}^n$ can always be separated by an hyperplane.

Theorem 1.2 (Cfr. [10] Theorem 3.4)**.** *Let $\mathcal{A}, \mathcal{B} \subset \mathbb{C}^n$ be two convex and disjoint subsets. Then, there exists $\xi \in \mathbb{C}^n$ and $\alpha \in \mathbb{R}$ such that*

$$\Re\langle \xi, x\rangle \le \alpha \le \Re\langle \xi, y\rangle$$

for all $x \in \mathcal{A}$ and all $y \in \mathcal{B}$.

2 Convex optimization and Lagrange duality

In what follows, we always assume that all the functions involved in problem (1) are continuous and that $K \neq \varnothing$ is bounded. As we observed in Remark 1.1, in this situation, the set of minimizers of (1) is non-empty and $\inf_{x\in K} F_0(x) = \min_{x\in K} F_0(x)$.

2.1 Lagrange duality

Let us consider the problem (1). The **Lagrange function** related to (1) is the function $L : \mathbb{C}^n \times \mathbb{C}^m \times [0, +\infty)^M \to (-\infty, +\infty]$ defined as

$$L(z, \xi, \nu) := F_0(z) + \Re\langle \xi, Az - y\rangle + \sum_{l=1}^{M} \nu_l (F_l(z) - b_l).$$

Let K be the set of feasible points and $z \in K$. Then, for all ξ, ν

$$L(z, \xi, \nu) = F_0(z) + \underbrace{\Re\langle \xi, Az - y\rangle}_{=0} + \sum_{l=1}^{M} \underbrace{\nu_l}_{\ge 0} \underbrace{(F_l(z) - b_l)}_{\le 0} \le F_0(z).$$

Therefore,

$$\inf_{z\in\mathbb{C}^n} L(z, \xi, \nu) \le \inf_{z\in K} L(z, \xi, \nu) \le \inf_{z\in K} F_0(z). \tag{2}$$

Definition 2.1. The function $H : \mathbb{C}^m \times [0, +\infty)^M \to [-\infty, +\infty]$ defined as

$$H(\xi, \nu) := \inf_{z\in\mathbb{C}^n} L(z, \xi, \nu)$$

is called **Lagrange dual function**.

Equation (2) reads as

$$H(\xi, \nu) \leq \inf_{x \in K} F_0(x)$$

for all $\xi \in \mathbb{R}^m$ and all $\nu \in [0, +\infty)^M$. In particular, the **weak duality inequality**

$$\sup_{\substack{\xi \in \mathbb{C}^m \\ \nu \in [0,+\infty)^M}} H(\xi, \nu) \leq \inf_{x \in K} F_0(x), \tag{W}$$

holds.

In Subsection 2.2, we will prove a condition under which $d^* := \sup_{\substack{\xi \in \mathbb{C}^m \\ \nu \in [0,+\infty)^M}} H(\xi, \nu)$ and $p^* := \inf_{z \in K} F_0(z)$ coincide. More precisely, if we consider the **dual problem** of (1), i.e. the optimization problem

$$\begin{cases} \max_{\substack{\xi \in \mathbb{C}^m \\ \nu \in \mathbb{R}^M}} H(\xi, \nu), & \\ \nu_l \geq 0 & l = 1, \dots, M, \end{cases} \tag{3}$$

we will establish conditions under which a solution of (3) is also a solution of (1). In this case, the identity $d^* = p^*$, which is called the **strong dual equality**, holds.

We end this paragraph proving a more symmetric expression of the weak dual inequality and of the strong dual equality. In what follows, we omit the set over which ξ and ν vary, as they are clear by the context. First, we prove the following lemma:

Lemma 2.2. *One has*

$$\sup_{\xi, \nu} L(z, \xi, \nu) = \begin{cases} F_0(z) & \text{if } z \in K, \\ +\infty & \text{otherwise.} \end{cases}$$

Proof. Let $z \notin K$ and suppose that $(Az - y)_J \neq 0$ for some $1 \leq J \leq m$ or $F_L(z) > b_L$ for some $1 \leq L \leq M$. Let $\{\nu^{(k)}\}_{k=1}^{\infty} \subset \mathbb{R}^M$ be a sequence such that $\nu_l^{(k)} = k \cdot \delta_{l,L}$ and take $\{\xi^{(r)}\}_j \subseteq \mathbb{R}^m$ as $\xi_j^{(r)} = r \cdot \operatorname{sgn}((Az - y)_J)\delta_{j,J}$ [3]. Then,

$$L(z, \xi^{(r)}, \nu^{(k)}) = F_0(z) + r\Re((Az - y)_J) + k(F_L(z) - b_L) \xrightarrow[k,r \to +\infty]{} +\infty.$$

If $z \in K$, then for all $\xi \in \mathbb{C}^m$ and $\nu \in \mathbb{R}^M$,

$$L(z, \xi, \nu) \leq F_0(z)$$

and the equality holds for $\xi = 0$ and $\nu = 0$. □

[3] Suppose that only the condition $(Az - y)_J \neq 0$ for some $1 \leq J \leq m$ is satisfied, in this case there is no $1 \leq L \leq M$ such that $F_L(z) > b_L$. Hence, $\delta_{l,L} = 0$ for all $1 \leq l \leq M$ and the following argument still holds. An analogous argument applies if $(Az - y)_J = 0$ for all $1 \leq J \leq m$.

As an immediate consequence,

$$p^* = \inf_z \sup_{\xi,\nu} L(z,\xi,\nu).$$

Since also $d^* := \sup_{\xi,\nu} H(\xi,\nu) = \sup_{\xi,\nu} \inf_z L(z,\xi,\nu)$, (W) can be written as

$$\sup_{\xi,\nu} \inf_z L(z,\xi,\nu) \le \inf_z \sup_{\xi,\nu} L(z,\xi,\nu), \tag{4}$$

while the strong duality equality reads as

$$\sup_{\xi,\nu} \inf_z L(z,\xi,\nu) = \inf_z \sup_{\xi,\nu} L(z,\xi,\nu). \tag{S}$$

2.2 Convex optimization

In this section, we approach the problem (1) in the in the case in which $F_1,\dots,F_M$ are convex functions. For the sake of clarity, we recall the definitions of convex functions and convex subsets.

Definition 2.3 (Convex functions and subsets)**.** Let $X \subseteq \mathbb{C}^n$. A function $F : X \to (-\infty,+\infty]$ is:

(a) **convex** if for all $x,y \in X$ and all $t \in [0,1]$, $F(tx+(1-t)y) \le tF(x)+(1-t)F(y)$;

(b) **strictly convex** if for all $x,y \in X$, $x \ne y$, and all $t \in (0,1)$, $F(tx+(1-t)y) < tF(x)+(1-t)F(y)$;

(c) **concave** (resp. **strictly concave**) if $-F$ is convex (resp. strictly convex).

A **convex subset** of $\mathbb{C}^n$ is a subset $X \subseteq \mathbb{C}^n$ such that either $X = \varnothing$ or X is closed with respect to convex combinations, i.e. $tx+(1-t)y \in X$ for all $x,y \in X$ and all $t \in [0,1]$.

Remark 2.4. (a) If $h : \mathbb{C}^n \to (-\infty,+\infty]$ is a linear functional, then h is both convex and concave. This follows trivially by Definition 2.3. In particular, if $A \in \mathbb{C}^{m\times n}$ and $h_j(z) = (Az)_j$, then h_j is both convex and concave.

(b) Sublevels of convex functions are convex subsets. More precisely, if F is convex, then for all $\alpha \in (-\infty,+\infty]$, $Y_\alpha := F^{-1}((-\infty,\alpha])$ is convex. We prove it for the sake of completeness: if $Y_\alpha = \varnothing$ there is nothing to prove. Suppose $Y_\alpha \ne \varnothing$ and take $z,w \in Y_\alpha$ and $t \in [0,1]$, then:

$$F(tz+(1-t)w) \le tF(z)+(1-t)F(w) \le t\alpha+(1-t)\alpha = \alpha \quad \Longrightarrow \quad tz+(1-t)w \in Y_\alpha.$$

(c) Trivially, if $C_1, C_2 \subseteq \mathbb{C}^n$ are convex, then $C_1 \cap C_2$ is convex.

(d) If F defined on $\mathbb{C}^n$ is convex, then F is continuous. As a consequence, the sublevels of such a function F are all closed and convex.

Under the previously stated convexity assumptions, the set K of the feasible points is closed and convex. In fact, using the notation introduced in Remark 1.1, the convexity of K follows by Remark 2.4 (a)-(c), writing K as

$$K = \bigcap_{l=1}^{M} F_j^{-1}((-\infty, b_l]) \cap h^{-1}(\{y\}),$$

where $h(z) := Az$, while the fact that it is closed follows by Remark 1.1.

Definition 2.5 (Local and global minimizers)**.** Let $F : Y \subseteq \mathbb{C}^n \to (-\infty, +\infty]$. A point $z \in Y$ is:

(a) a **local minimizer** of F if there exists $\varepsilon > 0$ such that

$$w \in Y,\ \|z - w\|_2 \leq \varepsilon \quad \Longrightarrow \quad F(z) \leq F(w).$$

(b) a **global minimizer** of F if $F(z) \leq F(w)$ for all $w \in Y$.

Proposition 2.6. *Let $K \neq \varnothing$ be convex and $F_0 : K \to (-\infty, +\infty]$ be a convex function. Then,*

(i) any local minimizer of F_0 is a global minimizer;

(ii) the set of the minimizers of F_0 is convex.

(iii) If F_0 is strictly convex, the minimizer of F_0 is unique.

Proof. (i) Let $\varepsilon > 0$ such that $\|z - w\|_2 \leq \varepsilon \Longrightarrow F_0(z) \leq F_0(w)$. Let $\zeta \in K$ and $t \in (0, 1)$ be such that $\|z - w\|_2 \leq \varepsilon$, where $w := tz + (1 - t)\zeta$. Then,

$$F_0(z) \leq F_0(w) \leq tF_0(z) + (1 - t)F_0(\zeta) \quad \Longrightarrow \quad (1 - t)F_0(z) \leq (1 - t)F_0(\zeta).$$

Since $1 - t \in (0, 1)$, it follows that for all $\zeta \in Y$ one has $F_0(z) \leq F_0(\zeta)$. Thus, z is a global minimizer.

(ii) Let $z, w \in K$ be such that $F_0(z) = F_0(w) = \inf_{\zeta \in K} F_0(\zeta)$. Then, for $t \in [0, 1]$

$$F_0(tz + (1 - t)w) \leq tF_0(z) + (1 - t)F_0(w) = \inf_{\zeta \in K} F_0(\zeta).$$

Therefore, $tz + (1 - t)w$ is a minimizer.

(iii) If $z \neq w$ are both minimizers of F_0 and $t \in (0, 1)$, by strict convexity:

$$F_0(tz + (1 - t)w) < tF_0(z) + (1 - t)F_0(w) = \inf_{\zeta \in K} F_0(\zeta).$$

This contradicts the definition of infimum.

□

Remark 2.7. In several applications, the following hypothesis are assumed: let $F_0, F_1, \ldots, F_M$ be as in (1),

(a) $F_0 : \mathbb{C}^n \to \mathbb{R}$ is convex;

(b) $F_1, \ldots, F_M : \mathbb{C}^n \to \mathbb{R}$ are convex;

(c) K is non-empty and bounded.

Since convex functions defined on $\mathbb{C}^n$ are continuous, (b) and (c) imply that, in this framework, the set of the feasible points K is compact, convex and non-empty. Since, by (a) F_0 is continuous, the infimum $\inf_K F_0$ is attained. In particular, under assumptions (a)-(c), problem (1) has, at least, one solution and the conclusions of Proposition 2.6 hold.

In the convex framework, if the constraints $F_l(z) \le b_l$ do not reduce to $F_l(z) = b_l$, more precisely if for all $l = 1, \ldots, M$ the inequality $F_l(z) < b_l$ holds for some $z \in \mathbb{C}^n$, then the strong duality holds.

Theorem 2.8 (Cfr. [1], Section 5.3.2)**.** *Assume that $F_0, F_1, \ldots, F_M$ are convex functions defined on $\mathbb{C}^n$ and $F_0 : \mathbb{C}^n \to (-\infty, +\infty]$. Let $z^\#$ and $(\xi^\#, \nu^\#)$ be optimizers for (1) and (3) respectively.*

(i) If there exists $\tilde{z} \in K$ such that $A\tilde{z} = y$ and $F_l(\tilde{z}) < b_l$ for all $l = 1, \ldots, M$, then $H(\xi^\#, \nu^\#) = F_0(z^\#)$.

(ii) In absence of inequality constraints, if $K \neq \varnothing$ (i.e. if there exists $\tilde{z} \in \mathbb{C}^n$ such that $A\tilde{z} = y$), then $H(\xi^\#, \nu^\#) = F_0(z^\#)$.

Before proving Theorem 2.8, we need some prior concepts.

Definition 2.9 (Hyperplanes and supporting hyperplanes)**.** (a) A **hyperplane** is any subset of $\mathbb{C}^n$ in the form

$$\Gamma = \{z \in \mathbb{C}^n \ : \ \Re\langle \xi, z\rangle = \alpha\},$$

for some $\xi \in \mathbb{C}^n$ and $\alpha \in \mathbb{R}$.

(b) Let Y be a subset of $\mathbb{C}^n$. A **supporting hyperplane** to Y is an hyperplane $\Gamma = \{z \in \mathbb{C}^n \ : \ \Re\langle \xi, z\rangle = \alpha\}$, with $\alpha \in \mathbb{R}$, such that $Y \subseteq \{z \in \mathbb{C}^n \ : \ \Re\langle z, \xi\rangle \ge \alpha\}$ and $Y \cap \Gamma \neq \varnothing$.

Suppose that F_0 is defined on $\mathbb{C}^n$, as in the assumptions of Theorem 2.8, and consider the image of the function

$$z \in \mathbb{C}^n \mapsto ((F_1(z) - b_1, \ldots, F_M(z) - b_M, Az - y, F_0(z)) \in \mathbb{R}^M \times \mathbb{C}^m \times \mathbb{R},$$

that is the set

$$\mathcal{G} := \Big\{ (F_1(z) - b_1, \ldots, F_M(z) - b_M, Az - y, F_0(z)) \in \mathbb{R}^M \times \mathbb{C}^m \times \mathbb{R} \ : \ z \in \mathbb{C}^n \Big\}.$$

Clearly, the minimum in problem (1) is

$$p^* = \inf\Big\{t \in \mathbb{R} \ : \ (u, v, t) \in \mathcal{G}, \ u_l \le 0 \ \ \forall l = 1, \dots, M, \ v = 0\Big\}. \tag{5}$$

The Lagrange function associated to (1) evaluated in $(u, v, t) \in \mathcal{G}$ can be written as

$$t + \Re\langle \xi, v\rangle + \sum_{l=1}^{M} \nu_l \cdot u_l = \Lambda(u, v, t),$$

where Λ is the linear functional defined by the vector $(\nu, \xi, 1)^T \in \mathbb{R}^M \times \mathbb{C}^m \times \mathbb{R}$, that is

$$\Lambda(u, v, t) = \Re\langle(\nu, \xi, 1), (u, v, t)\rangle.$$

For this reason, $H(\xi, \nu)$ is obtained by minimizing Λ over $\mathcal{G}$, i.e.

$$H(\xi, \nu) = \inf\Big\{\Lambda(u, v, t) \ : \ (u, v, t) \in \mathcal{G}\Big\}.$$

If we suppose that $H(\xi, \nu) > -\infty$, then the inequality

$$\Lambda(u, v, t) \ge H(\xi, \nu) \quad \forall (u, v, t) \in \mathcal{G}$$

tells that $\Lambda(u, v, t) = H(\xi, \nu)$ defines a supporting hyperplane to $\mathcal{G}$ in the point (u, v, t).

Remark 2.10. Since the last coordinate of the vector $(\nu, \xi, 1)^T$ is non-zero and $(\nu, \xi, 1)$ defines the normal vector to the supporting hyperplane, it follows that the supporting hyperplane is never parallel to the t-axis. In particular, it intersects the axis $u = v = 0$. More precisely, the intersection between the supporting hyperplane and the line $u = v = 0$ is the point $(0, 0, H(\xi, \nu))$. By weak duality, we also have $H(\xi, \nu) \le p^*$, so that this intersection always lies on the closed half-line $\{(0, 0, s) \ : \ s \le p^*\}$.

Consider the epigraph

$$\begin{aligned} \mathcal{A} &:= \mathcal{G} + ((\mathbb{R}_{\ge 0})^M \times \mathbb{C}^m \times \mathbb{R}_{\ge 0}) = \\ &= \Big\{(u, v, t) \in \mathbb{R}^M \times \mathbb{C}^m \times \mathbb{R} \ : \ u_l \ge F_l(z) - b_l \ (l = 1, \dots, M), \\ &\quad v = Az - y, \ t \ge F_0(z) \ for\ some\ z \in \mathbb{C}^n\Big\}. \end{aligned} \tag{6}$$

Lemma 2.11. *If $F_0, F_1, \dots, F_M$ are convex, then $\mathcal{A}$ is convex.*

Proof. Let $(u, v, t), (U, V, T) \in \mathcal{A}$ be such that

$$\begin{cases} u_l \ge F_l(z) - b_l & l = 1, \dots, M, \\ v = Az - y & j = 1, \dots, m, \\ t \ge F_0(z) \end{cases} \quad \text{and} \quad \begin{cases} U_l \ge F_l(Z) - b_l & l = 1, \dots, M, \\ V = AZ - Y & j = 1, \dots, m, \\ T \ge F_0(Z) \end{cases}$$

for some $z, Z \in \mathbb{C}^n$. Let $\tau \in [0,1]$ be fixed. We have to prove that $\tau(u,v,t) + (1-\tau)(U,V,T) \in \mathcal{A}$. For all $l = 1,\dots,M$, by the convexity of F_l, we have

$$\begin{aligned}\tau u_l + (1-\tau)U_l &\geq \tau(F_l(z) - b_l) + (1-\tau)(F_l(Z) - b_l) = \\ &= \tau F_l(z) + (1-\tau)F_l(Z) - b_l \geq F_l(\tau z + (1-\tau)Z) - b_l.\end{aligned} \tag{7}$$

Analogously, for all $j = 1,\dots,m$, by linearity,

$$\tau v_j + (1-\tau)V_j = A(\tau z + (1-\tau)Z) - y \tag{8}$$

and, by the convexity of F_0,

$$\tau t + (1-\tau)T \geq F_0(\tau z + (1-\tau)Z). \tag{9}$$

Equations (7), (8) and (9) together give the assertion. $\square$

Proof of Theorem 2.8. Without loss of generality, we may assume $K \neq \varnothing$ and $rk(A) = m$ [4]. If the set K is non-empty, $p^* = \inf_{z\in\mathbb{C}^n} F_0(z)$ [5] is either finite or $-\infty$. In the latter case, weak duality implies that also $d^* > -\infty$ and the assertion follows. Thus, we suppose that $p^* \neq -\infty$.

Let $\mathcal{A}$ be defined as in (6) and set

$$\mathcal{B} := \Big\{(0,0,s) \in \mathbb{R}^M \times \mathbb{C}^m \times \mathbb{R} \ : \ s < p^*\Big\}.$$

$\mathcal{A}$ and $\mathcal{B}$ are disjoint. In fact, if $(u,v,s) \in \mathcal{A}\cap\mathcal{B}$, then $s < p^*$ and also $t \geq F_0(z)$ for some $z \in \mathbb{C}^n$. The fact that $F_0(z) \geq p^*$ by definition leads to a contradiction.

By Lemma 2.11, $\mathcal{A}$ is convex, while $\mathcal{B}$ (which is an open half-line) is trivially convex. Therefore, the assumptions of Theorem 1.2 are satisfied and we conclude that there exists a triple of parameters $(\tilde{\nu}, \tilde{\xi}, \mu) \neq 0$ and $\alpha \in \mathbb{R}$ such that

$$(u,v,t) \in \mathcal{A} \implies \tilde{\nu}^T u + \Re\langle\tilde{\xi}, v\rangle + \mu t \geq \alpha, \tag{10}$$

$$(u,v,t) \in \mathcal{B} \implies \tilde{\nu}^T u + \Re\langle\tilde{\xi}, v\rangle + \mu t \leq \alpha. \tag{11}$$

By (10), we conclude that $\tilde{\nu}_l \geq 0$ for all $l = 1,\dots,M$ and $\mu \geq 0$, otherwise $\tilde{\nu}^T u + \mu t$ would be unbounded from below in $\mathcal{A}$ [6]. Also, applying the definition of $\mathcal{B}$ to (11), we find $\mu t \leq \alpha$ for all $t < p^*$, which implies that $\mu p^* \leq \alpha$. It follows that:

$$\begin{cases} \tilde{\nu}_l \geq 0 & \text{for all } l = 1,\dots,M, \\ 0 \leq \mu p^* \leq \alpha. \end{cases}$$

[4] The condition $rk(A) = m' < m$ means that $m - m'$ equations of the linear system $Az = y$ are redundant. Hence, they can be dropped. This procedure, replaces A with the full row rank matrix obtained by suppressing $m - m'$ of its rows properly, whose rank is maximum.

[5] $p^* = \inf\{(0,0,t) \ : \ t < p^*\}$ also.

[6] For instance, if $\nu_l < 0$, since in $\mathcal{A}$ you have $u_l = F_l(z) - b_l + U_l^2$ for some $U_l \in \mathbb{R}$, you may take U_l^2 small enough to contradict (10).

Therefore, for all $z \in \mathbb{C}^n$ (since $(F_1(z) - b_1, \dots, F_M(z) - b_M, Az - y, F_0(z)) \in \mathcal{G} \subseteq \mathcal{A}$),

$$\sum_{l=1}^{M} \tilde{\nu}_l (F_l(z) - b_l) + \Re\langle \tilde{\xi}, Az - y \rangle + \mu F_0(z) \geq \alpha \geq \mu p^*. \tag{12}$$

If $\mu > 0$, then (12) gives that $L(z, \tilde{\xi}/\mu, \tilde{\nu}/\mu) \geq p^*$ for all $z \in \mathbb{C}^n$, which implies that $H(\tilde{\xi}/\mu, \tilde{\nu}/\mu) \geq p^*$. Since the other inequality is implicit in weak duality inequality, we conclude that $H(\tilde{\xi}/\mu, \tilde{\nu}/\mu) = p^*$.

We prove that it must be $\mu > 0$. For, suppose that $\mu = 0$ and observe that in this case (12) reads as

$$\sum_{l=1}^{M} \tilde{\nu}_l (F_l(z) - b_l) + \Re\langle \tilde{\xi}, Az - y \rangle \geq 0. \tag{13}$$

Let $\tilde{z}$ be the point such that $A\tilde{z} = y$ and $F_l(\tilde{z}) < b_l$ $(l = 1, \dots, M)$, whose existence is assumed by hypothesis. Then,

$$0 \leq \sum_{l=1}^{M} \underbrace{\tilde{\nu}_l}_{\geq 0} \underbrace{(F(\tilde{z}) - b_l)}_{< 0} + \underbrace{\Re\langle \tilde{\xi}, A\tilde{z} - y \rangle}_{= 0} \leq 0,$$

which implies that $\tilde{\nu}_l = 0$ for all $l = 1, \dots, M$. In particular, since the triple $(\tilde{\nu}, \tilde{\xi}, \mu)$ is non-zero, we deduce that $\tilde{\xi} \neq 0$. So, (13) reduces to

$$\Re\langle \tilde{\xi}, Az - y \rangle \geq 0 \tag{14}$$

for all $z \in \mathbb{C}^n$, with $\Re\langle \tilde{\xi}, A\tilde{z} - y \rangle = 0$. But $\{z \in \mathbb{C}^n : \Re\langle \tilde{\xi}, Az - y \rangle = 0\}$ defines an hyperplane of $\mathbb{C}^n$, so that $\Re\langle \tilde{\xi}, Az - y \rangle < 0$ for some for some $z \in \mathbb{C}^n$ or it vanishes identically. However, neither of the two options is possible, since (14) holds and, on the other hand, $\Re\langle \tilde{\xi}, Az - y \rangle = 0$ for all $z \in \mathbb{C}^n$ contradicts both $\tilde{\xi} \neq 0$ and the assumption $rk(A) = m$. □

Remark 2.12. In CS, the uniqueness of the solution is granted by further properties imposed on the sensing matrix A, e.g. the restricted isometry property (RIP). The problem of giving conditions under which the solutions to these minimization problems are unique is beyond the purpose of this section (see [2]). The interested reader may consult [6], Chapters 4, 5 and 6.

2.3 Sparsity recovering

In this section, we study the equivalence that holds in several situations between (1) and other minimization problems. The inverse problem in CS consists in solving the following version of (1):

$$\min_{z \in \mathbb{C}^n} \|Az - y\|_2 \quad \textit{subject to} \quad \Psi(z) \leq r \tag{15}$$

for some $r > 0$, where $y \in \mathbb{C}^m$ is an (undersampled) acquired data, A is a properly chosen sensing matrix and $\Psi(z)$ is a regularizing term, which is typically a convex function of z. It is well known (and we will prove it) that under certain hypothesis on Ψ, (15) is equivalent to the unconstrained minimization problem

$$\min_{z\in\mathbb{C}^n} \|Az-y\|_2^2 + \lambda\Psi(z), \tag{16}$$

where $\lambda > 0$ is called *tuning parameter.*

The first result in this direction is a consequence of Theorem 2.8.

Corollary 2.13. *Let $F_0 : \mathbb{C}^n \to [0,+\infty)$ and $\phi : [0,+\infty) \to \mathbb{R}$ be such that ϕ is monotonically increasing and $\phi \circ F_0$ is convex. Let $\tau > 0$ and $\Psi : \mathbb{C}^n \to \mathbb{R}$ be a convex function such that $\Psi^{-1}([0,\tau)) \neq \varnothing$. Let $z^\#$ is a minimizer of the problem*

$$\min_{z\in\mathbb{C}^n} F_0(z) \quad \textit{subject to } \Psi(z) \leq \tau. \tag{17}$$

Then, there exists $\lambda \geq 0$ such that $z^\#$ is a minimizer of

$$\min_{z\in\mathbb{C}^n} \phi(F_0(z)) + \lambda\Psi(z). \tag{18}$$

Proof. Since ϕ is monotonically increasing, (17) is obviously equivalent to

$$\min_{z\in\mathbb{C}^n} \phi(F_0(z)) \quad \textit{subject to } \Psi(z) \leq \tau, \tag{19}$$

whose Lagrangian is given by

$$L(z,\nu) = \phi(F_0(z)) + \nu(\Psi(z) - \tau). \tag{20}$$

By the assumption, $\phi \circ F_0$ and Ψ are convex and the inequality $\Psi(\tilde{z}) < \tau$ is satisfied by some $\tilde{z} \in \mathbb{C}^n$ (observe that here we need $\tau > 0$), so we can apply Theorem 2.8 to get $H(z^\#, \nu^\#) = \phi(F_0(z^\#))$ for some $\nu^\# \geq 0$. By (4), for all $z \in \mathbb{C}^n$ and all $\nu \in \mathbb{R}$,

$$L(z^\#, \nu^\#) \leq L(z, \nu^\#),$$

so that $z^\#$ is also a minimizer of the function $z \in \mathbb{C}^n \mapsto L(z, \nu^\#)$. Since the constant term $-\nu\tau$ in (20) does not affect the minimum, we have that $z^\#$ is a minimizer of

$$\min_{z\in\mathbb{C}^n} \phi(F_0(z)) + \nu^\#(\Psi(z) - \tau).$$

$\square$

Example 2.14. Clearly, the assumptions of Corollary 2.13 apply to problem (15), where z is the approximated solution, $\|Az-y\|_2$ is a measure of the approximation error and $\Psi(z)$ is usually taken as $\|z\|_1$, $\|z\|_{TV}$ or $\|\mathcal{F}z\|_1$, where $\mathcal{F}$ is the discrete Fourier transform of z. In particular, CS inverse problem is equivalent (for an appropriate tuning parameter $\lambda \geq 0$) to (16). The tuning parameter which provides the exact solution is $\lambda = \tilde{\nu}/\mu$, where $\tilde{\nu}$ and μ are given as in the proof of Theorem 2.8, i.e. they are the coefficients defining a separating hyperplane between $\mathcal{A}$ and $\mathcal{B}$ in the proof of Theorem 2.8.

Another point of view one may assume in CS is the following: consider the problem of recovering a signal $z \in \mathbb{C}^n$ which is contaminated by noise. This means that the received signal y has the form $y = Az + e$, where A is the sensing matrix and $e \in \mathbb{C}^m$ is a noise-like signal. Suppose to know that $\|e\|_2 \leq \eta$ for some $\eta > 0$, i.e. suppose that the amplitude of noise can be estimated a priori. Then, since $e = y - Az$, the condition $\|e\|_2 \leq \eta$ takes the form of a constraint: $\|Az - y\|_2 \leq \eta$. In CS, the model of vectors to recover is provided by a **sparsity condition**, this means that we know in advance that many of the coordinates of the acquired signal y vanish. More precisely:

Definition 2.15 (Sparse vectors)**.** Let $0 \leq k \leq n \leq N$ be integers. A vector $y \in \mathbb{C}^n$ is k**-sparse** with respect to a frame [7] $\{\Phi_1, \dots, \Phi_N\}$ if $\|\Phi y\|_0 := card(supp(\Phi y)) \leq k$, where $\Phi \in \mathbb{R}^{N\times n}$ is the matrix given by $\Phi = (\Phi_1|\Phi_2|\dots|\Phi_N)$.

For the sake of simplicity, in the following, we always assume that $\Phi = \mathbb{1}_n \in \mathbb{R}^{n\times n}$, i.e. the frame $\{\Phi_1, \dots, \Phi_N\}$ consists of the canonical basis of $\mathbb{R}^n$.

Remark 2.16. It is worthy to observe that a vector $y \in \mathbb{C}^n$ is k-sparse if and only if $y \in \Sigma_k$, where Σ_k is the union of all the subspaces of $\mathbb{C}^n$ in the form

$$\Big\{z \in \mathbb{C}^n \ : \ z_{j_1} = \dots = z_{j_k} = 0\Big\}, \qquad 1 \leq j_1 < \dots < j_k \leq n.$$

This allows the following generalization of Definition 2.15, that can be used in the context of infinite dimensional vector spaces: *let X be a vector space, a vector $y \in X$ is called **sparse** if $y \in \Sigma$, where Σ is a finite union of subspaces of X*; and, further, using the notion of variety.

Therefore, the inverse problem for sparse noise-contaminated signals becomes:

$$\min_{z\in\mathbb{C}^n} \|z\|_0 \quad \textit{subject to } \|Az - y\|_2 \leq \eta. \tag{21}$$

However, $\|\cdot\|_0$ is highly non-convex so that, not only theorems of Subsection 2.2 cannot be applied in this context, but also the numerical algorithms that lead to a minimizer of (21) may turn to be bad-conditioned. For this reason, one replaces $\|\cdot\|_0$ with a convex norm of $\mathbb{C}^n$, for instance the ℓ^1_n-norm, which is convex:

$$\min_{z\in\mathbb{C}^n} \|z\|_1 \quad \textit{subject to } \|Az - y\|_2 \leq \eta. \tag{22}$$

In view of the following result, it turns out that $\|\cdot\|_1$ enforces sparsity, i.e. a minimizer of (22) is sparse, at least in the real setting:

Proposition 2.17 (Cfr. [6] Theorem 3.1, Exercise 3.3)**.** *Let $A \in \mathbb{R}^{m\times n}$, $y \in \mathbb{R}^m$. Assume the uniqueness of a minimizer $x^{\#}$ of*

$$\min_{x\in\mathbb{R}^n} \|x\|_1 \quad \textit{subject to } \|Ax - y\| \leq \eta, \tag{23}$$

where $\|\cdot\|$ is any (quasi-)norm [8] on $\mathbb{R}^m$ and $\eta \geq 0$, then:

[7] Recall that a set $\{\Phi_j\}_j \subset X$ is a **frame** for the inner-product space $(X, \langle\cdot,\cdot\rangle)$ if there exist $A, B > 0$ such that for all $x \in X$ one has $A\|x\|^2 \leq \sum_j \langle x, \Phi_j\rangle \leq B\|x\|^2$.

[8] Actually, $\|Ax - y\|$ can be any constraint function of Ax. For instance, $\|Ax - y\| = \log(1 + \|Ax - y\|_p)$ for any $p \geq 0$.

(i) the columns $\{a_j \; : \; j \in supp(x^{\#})\}$ of A are linearly independent;

(ii) $x^{\#}$ is m-sparse.

Proof. (i) By the way of contradiction, assume that the system $\{a_j \; : \; j \in S\}$ is linearly dependent, where $S = supp(x^{\#})$. Then, there exists $v \in \mathbb{R}^n$ such that $supp(v) = S$ and $Av = 0$. Since for all $t \in \mathbb{R}$, $\left\|A(x^{\#} + tv) - y\right\| = \left\|Ax^{\#} - y\right\| \leq \eta$ [9] and $x^{\#}$ is the unique minimizer of (23), we have

$$\left\|x^{\#}\right\|_1 < \left\|x^{\#} + tv\right\|_1 = \sum_{j \in S} |x_j^{\#} + tv_j| = \sum_{j \in S} \operatorname{sgn}(x_j^{\#} + tv_j)(x_j^{\#} + tv_j)$$

for all $t \neq 0$. If $|t| < \min_{j \in S} \frac{|x_j^{\#}|}{\|v\|_\infty}$, $\operatorname{sgn}(x_j^{\#} + tv_j) = \operatorname{sgn}(x_j^{\#})$ for all $j \in S$. For such values of t,

$$\begin{aligned}\left\|x^{\#}\right\|_1 &< \sum_{j \in S} \operatorname{sgn}(x_j^{\#})(x_j^{\#} + tv_j) = \sum_{j \in S} |x_j^{\#}| + t \sum_{j \in S} \operatorname{sgn}(x_j^{\#}) v_j = \\ &= \left\|x^{\#}\right\|_1 + t \underbrace{\sum_{j \in S} \operatorname{sgn}(x_j^{\#}) v_j}_{=:\, \alpha} .\end{aligned}$$

Hence, for all $t \neq 0$, $t\alpha > 0$. However, α is a fixed real number, in particular, we can always choose $t \neq 0$ small enough such that $t\alpha \leq 0$. This is a contradiction.

(ii) By (i), if $m < \|x\|_0 \leq n$, A would have $\|x\|_0 > m$ linearly independent columns, but this is impossible, since $rk(A) \leq m$, since $A \in \mathbb{R}^{m \times n}$.

□

Remark 2.18. In the complex setting, if $\mathbb{C}$ must be thought as a real vector space in order for Proposition 2.17 to hold. In fact, $z = (1, e^{2\pi i/3}, e^{4\pi i/3})^T$ is the unique solution to (23)

$$A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix},$$

$y = (3/2 + \sqrt{3}i/2, \sqrt{3}i)^T$ and $\eta = 0$ (i.e. $Ax = y$), but it is not 2-sparse.

We presented two different point of views for the inverse problem in CS:

$$\min_{z \in \mathbb{C}^n} \|Az - y\|_2 \quad \textit{subject to} \quad \|z\|_1 \leq \tau, \tag{24}$$

$$\min_{z \in \mathbb{C}^n} \|z\|_1 \quad \textit{subject to} \quad \|Az - y\|_2 \leq \eta. \tag{25}$$

Approaching (24), we also came up with the problem

$$\min_{z \in \mathbb{C}^n} \|Az - y\|_2^2 + \lambda \|z\|_1 . \tag{26}$$

[9]This is the only point where we use the form assumed by the constraint. We stress that it is only important that the constraint is a function of Ax, in order for the equality $A(x^{\#} + tv) = Ax^{\#}$ to hold.

Theorem 2.19. *Let $F_0 : \mathbb{C}^n \to [0,+\infty)$ and $\phi : [0,+\infty) \to \mathbb{R}$ be such that ϕ is monotonically increasing and $\phi \circ F_0$ is convex* [10]. *Let $\Psi : \mathbb{C}^n \to [0,+\infty)$ be a convex function such that $\Psi(0) = 0$* [11] [12] [13]. *Consider the problems:*

$$\min_{z\in\mathbb{C}^n} \Psi(z) \quad \textit{subject to } F_0(z) \leq \eta, \tag{27}$$

$$\min_{z\in\mathbb{C}^n} \phi(F_0(z)) + \lambda\Psi(z), \tag{28}$$

$$\min_{z\in\mathbb{C}^n} F_0(z) \quad \textit{subject to} \quad \Psi(z) \leq \tau. \tag{29}$$

(i) If $z^\#$ is a minimizer of (28) with $\lambda > 0$, then there exists $\eta = \eta_{z^\#} \geq 0$ such that $z^\#$ is a minimizer of (27).

(ii) If $z^\#$ is the unique minimizer of (27), then there exists $\tau = \tau_{z^\#} \geq 0$ such that $z^\#$ is the unique minimizer of (29).

(iii) If $z^\#$ is a minimizer of (29) with $\tau > 0$, then there exists $\lambda = \lambda_{z^\#} \geq 0$ such that $z^\#$ is a minimizer of (28).

Proof. Item (iii) is the content of Corollary 2.13. We prove (i) and (ii).

(i) Let $z^\#$ be the minimizer of (28) and set $\eta := F_0(z^\#)$. $\eta \geq 0$ by the assumptions on F_0. Let $\zeta \in \mathbb{C}^n$ satisfy $F_0(\zeta) \leq \eta$, then using the fact that $z^\#$ is the minimizer and the monotony of ϕ:

$$\lambda\Psi(z^\#)+\phi(F_0(z^\#)) \leq \lambda\Psi(\zeta)+\phi(F_0(\zeta)) \leq \lambda\Psi(\zeta) \leq \lambda\Psi(\zeta)+\phi(\eta) = \lambda\Psi(\zeta)+\phi(F_0(z^\#)).$$

Since $\lambda > 0$, the assertion follows.

(ii) Let $z^\#$ be the unique minimizer of (27). Let $\tau := \Psi(z^\#)$, which is non-negative by the assumption, and consider $\zeta \in \mathbb{C}^n$ satisfying both $z^\# \neq \zeta$ and $\Psi(\zeta) \leq \tau$. Since $z^\#$ is the unique minimizer of (27) and $\tau = \Psi(z^\#)$, ζ must satisfy $F_0(\zeta) > \eta$, otherwise it would be another minimizer of (27). Therefore,

$$F_0(\zeta) > \eta \geq F_0(z^\#).$$

We proved that for all the feasible points $\zeta \neq z^\#$, $F_0(z^\#) < F_0(\zeta)$. This implies that $z^\#$ is the unique solution of (29).

□

Remark 2.20. (a) Observe that the non-negativity of Ψ is only needed in the proof of Theorem 2.19 (ii).

(b) In Theorem 2.19 (i) and (ii), Ψ does not need to be convex.

[10] The convexity of $\phi \circ F_0$ is only required in the proof of (iii).

[11] This condition is stronger than the one encountered in Corollary 2.13, which required $\Psi^{-1}([0,\tau)) \neq \varnothing$ for some $\tau > 0$. We use this formulation, here, because it is satisfied by all the functions used in this context and to consider Ψ as independent of τ.

[12] Beside the condition $\Psi(0) = 0$, the condition $\Psi(z) \geq 0$ for all $z \in \mathbb{C}^n$ is a further condition that we have to impose to prove (ii).

[13] The convexity condition is required only to prove (iii).

3 Bi-criterion regularized approximation

In view of Corollary 2.13, the inverse problem in CS, namely

$$\min_{z\in\mathbb{C}^n} \|z\|_1 \quad \textit{subject to} \quad \|Az-y\|_2 \le \eta, \tag{30}$$

is, prior a correct choice of the parameter λ, *equivalent* to

$$\min_{z\in\mathbb{C}^n} \|Az-y\|_2^2 + \lambda \|z\|_1 . \tag{31}$$

(31), read as it is, provides an example of a regularization technique used to approach the so-called *bi-criterion problems*, in which the purpose is the simultaneous minimization of the residual error, expressed as $\|Az-y\|$, and the size of z, expressed as $|||z|||$, for given norms $\|\cdot\|$ and $|||\cdot|||$. More precisely, a bi-criterion problem has the form:

$$\min_{z\in\mathbb{C}^n} \{\|Az-y\|, |||z|||\}, \tag{32}$$

where the term $|||z|||$ can be interpreted as the the priori information about the size of z.

Differently from the point of view previously adopted, all the results of this section are stated in the real setting. Actually, we always regarded $\mathbb{C}^n$ as a real Euclidean space and all the results in the following paragraphs can be transposed to the complex setting via any isomorphism between $\mathbb{R}^{2n}$ and $\mathbb{C}^n$.

3.1 Regularization

Regularization consists in approaching (32) assigning a weighting to one of the two terms, solving the minimization problem

$$\min_{x\in\mathbb{R}^n} \|Ax-y\| + \gamma|||x||| \tag{33}$$

for all $\gamma > 0$ and comparing the obtained solutions. The function $\|Az-y\| + \gamma|||z|||$ in (33) is called **objective function**.

Remark 3.1. When the euclidean norms are involved, a common procedure is to replace $\|\cdot\|$ and $|||\cdot|||$ with their squares in (33).

We point out that, with respect to the approach illustrated in the previous sections, the point of view adopted here is significantly different, while the minimizing equation being the same: in (31), the parameter λ has a precise definition, depends of the noise amplitude η in (30) and of the input y; in turn, the parameter $\gamma > 0$ in (32) is a trade-off between the residual error and the a priori information about the size of z. Different values of γ lead to different solutions of (32), which trace out the *optimal trade-off curve* as γ runs over $(0, +\infty)$. To decide which parameter $\gamma > 0$ is *optimal*, a precise definition of *optimality* would be mandatory. Once this definition is established, there is still no reason for the optimal parameter γ to coincide with λ, which in turn is mathematically established in terms of η and y. In MRI, the optimality of γ is decided a posteriori by comparing the images reconstructed by (33), usually asking the opinion of some experts. Anyway, there exist well established procedures that may help in the localization the optimal parameter, for instance the L-curve criterion (see Subsection 3.3).

There are several situations in which the solution to (33) is explicit in terms of A, y and γ. In the Tikhonov regularization problem, where the ℓ^2-norms are involved, the explicit solution is derived differentiating the objective function:

Theorem 3.2 (Cfr. [1], Section 6.3.2)**.** *Let $A \in \mathbb{R}^{m\times n}$ and $\gamma > 0$. Then, the minimizer of*

$$\min_{x\in\mathbb{R}^n} \|Ax - y\|_2^2 + \gamma||x||_2^2 \tag{34}$$

is given by

$$x^{\#} = (A^T A + \gamma \mathbb{1}_n)^{-1} A^T y.$$

We point out that $(A^T A + \gamma \mathbb{1}_n)^{-1}$ exists because $\gamma > 0$ and, since A is real, $A^T A$ is positive semidefinite.

Proof of Theorem 3.2. Explicitly, the objective function is

$$\|Ax - y\|_2^2 + \gamma||x||_2^2 = (Ax - y)^T (Ax - y) + \gamma x^T x.$$

Hence, for all $l = 1, \dots, n$,

$$\begin{aligned}\frac{\partial}{\partial x_l}(\|Ax - y\|_2^2 + \gamma||x||_2^2) &= 2\left(\frac{\partial}{\partial x_l}(Ax - y)\right)^T (Ax - y) + 2\gamma x^T \frac{\partial x}{\partial x_l} = \\ &= 2(A^{(l)})^T (Ax - y) + 2\gamma x_l = \\ &= 2(A^l)^T Ax - 2(A^{(l)})^T y + 2\gamma x_l,\end{aligned}$$

where $A^{(l)}$ is the l-th column of A. Therefore,

$$\nabla(\|Ax - y\|_2^2 + \gamma||x||_2^2) = 2A^T Ax - 2A^T y + 2\gamma x = 0$$

if and only if

$$A^T Ax + \gamma x = A^T y,$$

that is $(A^T A + \gamma \mathbb{1}_n)x = A^T y$. □

Different formulations of (34) can be treated analogously. For instance, the explicit minimizer to the problem

$$\min_{x\in\mathbb{R}^n} \|Ax - y\|_2^2 + \gamma||\Phi x||_2^2,$$

for a given sparsifying transform $\Phi \in \mathbb{R}^{p\times n}$, is provided by the linear estimator

$$x^{\#} = (A^T A + \gamma \Phi^T \Phi)^{-1} A^T y$$

(cfr. [9] Theorem 13.1).

3.2 De-noising

Bi-criterion regularization is used in reconstruction problems, when the acquired data is corrupted by noise, to recover the noiseless signal. Here, as *noise* we intend a vector $e \in \mathbb{R}^m$ such that the acquired signal $y \in \mathbb{R}^m$ is the superposition of the exact signal $x \in \mathbb{R}^m$ and e, i.e. $y = x + e$. Typically, $\|e\|_2$ is assumed to be small and the process modeled by the bi-criterion problem

$$\min_{x \in \mathbb{R}^m} \{\|x - y\|_2 \,, \phi(x)\}, \tag{35}$$

for a given convex *regularizing function* $\phi : \mathbb{R}^m \to \mathbb{R}$, is called *de-noising*.

Example 3.3. (a) When the original signal is known to be smooth and the noise is known to be rapidly varying, *quadratic smoothing* produces a reliable de-noised signal. It consists in solving (35) with

$$\phi(x) = \sum_{j=1}^{m-1} (x_{j+1} - x_j)^2 = \|Dx\|_2^2 \,,$$

where $D \in \mathbb{R}^{(m-1)\times m}$ is the bidiagonal matrix

$$D = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & -1 & 1 \end{pmatrix}.$$

For, one writes the corresponding regularization equation, i.e.

$$\|x - y\|_2^2 + \gamma \|Dx\|_2^2 \,,$$

which produces the explicit solution $x^{\#} = (\mathbb{1}_m + \gamma D^T D)^{-1} y$.

(b) If the original signal features rapid variations, the quadratic smoothing method is not well suited to its reconstruction. In this situations, it is preferable to use the regularizing term

$$\phi_{TV}(x) := \sum_{j=1}^{m-1} |x_{j+1} - x_j| = \|Dx\|_1 \,,$$

which is called *total variation* (TV) of x. Enforcing the sparsity of Dx, $\phi_{TV}(x)$ is a regularization term well-suited to recover signals having few rapid variations.

3.3 L-curve method for the regularization parameter selection

It has become clear that a correct choice of the optimal tuning parameter γ for problem (33) is fundamental for regularization methods. One possibility is to find the minimizer x_γ to (33) for different values of γ, and trace the curve $(\log \|Ax_\gamma - y\|, \log |||x_\gamma|||)$, whose shape is reminiscent of the letter L. Heuristically, the optimal tuning parameter, γ_{opt} is the one corresponding to the point $(\log \|Ax_{\gamma_{opt}} - y\|, \log |||x_{\gamma_{opt}}|||)$ closest to the corner of the graph.

References

[1] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[2] E. J. Candés, *The Restricted Isometry Property and its Implications for Compressed Sensing*, C. R. Acad. Sci. Paris, Ser. I 346 (2008) 589-592.

[3] E. J. Candés, J. Romberg, T. Tao, *Robust Uncertainty Principles: Exact Signal Reconstruction From Highly Incomplete Frequency Information*, Ieee Transactions on Information Theory, Vol. 52, No. 2, February 2006.

[4] D. L. Donoho, *Compressed Sensing*, Ieee Transactions on Information Theory, Vol. 52, No. 4, April 2006.

[5] D. L. Donoho, M. Elad, V. N. Temlyakov, *Stable Recovery of Sparse Overcomplete Representations in the Presence of Noise*, Ieee Transactions on Information Theory, Vol. 52, No. 1, January 2006.

[6] S. Foucart, H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Birkhäuser, 2013.

[7] P. C. Hasen, *The L-curve and its use in the numerical treatment of inverse problems*, `https://www.sintef.no/globalassets/project/evitameeting/2005/lcurve.pdf`.

[8] M. Lustig, D. Donoho, J. M. Pauly, *Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging*, Magn. Res. Med. 58:1182-1195 (2007).

[9] S. Mallat, *A Wavelet Tour in Signal Processing* Third edition, Academic Press (2009).

[10] W. Rudin, *Functional Analysis* 2nd edition, McGrow Hill, 1990.

[11] G. Varela-Mattatal, C. A. Baron, R. S. Menon, *Automatic determination of the regularization weighting for wavelet-based compressed sensing MRI reconstructions*, Magn. Reson. Med. 2021;86:1403-1419.