On the Determination of Lagrange Multipliers for a Weighted LASSO Problem Using Geometric and Convex Analysis Techniques

Gianluca Giacchi[1,2,3,4] · **Bastien Milani**[3] · **Benedetta Franceschiello**[2,4]

Abstract

Compressed Sensing (CS) encompasses a broad array of theoretical and applied techniques for recovering signals, given partial knowledge of their coefficients, cf. Candés (C. R. Acad. Sci. Paris, Ser. I **346**, 589–592 (2008)), Candés et al. (IEEE Trans. Inf. Theo (2006)), Donoho (IEEE Trans. Inf. Theo. **52**(4), (2006)), Donoho et al. (IEEE Trans. Inf. Theo. **52**(1), (2006)). Its applications span various fields, including mathematics, physics, engineering, and several medical sciences, cf. Adcock and Hansen (Compressive Imaging: Structure, Sampling, Learning, p. 2021), Berk et al. (2019 13th International conference on Sampling Theory and Applications (SampTA) pp. 1-5. IEEE (2019)), Brady et al. (Opt. Express **17**(15), 13040–13049 (2009)), Chan (Terahertz imaging with compressive sensing. Rice University, USA (2010)), Correa et al. (2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) pp. 7789–7793 (2014, May) IEEE), Gao et al. (Nature **516**(7529), 74–77 (2014)), Liu and Kang (Opt. Express **18**(21), 22010–22019 (2010)), McEwen and Wiaux (Mon. Notices Royal Astron. Soc. **413**(2), 1318–1332 (2011)), Marim et al. (Opt. Lett. **35**(6), 871–873 (2010)), Yu and Wang (Phys. Med. Biol. **54**(9), 2791

✉ Gianluca Giacchi
   gianluca.giacchi2@unibo.it

   Bastien Milani
   bastien.milani@chuv.ch

   Benedetta Franceschiello
   benedetta.franceschiello@hevs.ch

1   Università di Bologna, Dipartimento di Matematica, Piazza di Porta San Donato 5, 40126 Bologna, Italy

2   Institute of Systems Engineering, School of Engineering, HES-SO Valais-Wallis, Rue de l'Industrie 23, 1950 Sion, Switzerland

3   Lausanne University Hospital and University of Lausanne, Lausanne, Department of Diagnostic and Interventional Radiology, Rue du Bugnon 46, Lausanne 1011, Switzerland

4   The Sense Innovation and Research Center, Avenue de Provence 82 1007, Lausanne and Ch. de l'Agasse 5, 1950 Sion, Switzerland

 Springer

(2009)), Yu and Wang (Phys. Med. Biol. **54**(9), 2791 (2009)). Motivated by our interest in the mathematics behind Magnetic Resonance Imaging (MRI) and CS, we employ convex analysis techniques to analytically determine equivalents of Lagrange multipliers for optimization problems with inequality constraints, specifically a weighted LASSO with voxel-wise weighting. We investigate this problem under assumptions on the fidelity term $\|Ax - b\|_2^2$, either concerning the sign of its gradient or orthogonality-like conditions of its matrix. To be more precise, we either require the sign of each coordinate of $2(Ax - b)^T A$ to be fixed within a rectangular neighborhood of the origin, with the side lengths of the rectangle dependent on the constraints, or we assume $A^T A$ to be diagonal. The objective of this work is to explore the relationship between Lagrange multipliers and the constraints of a weighted variant of LASSO, specifically in the mentioned cases where this relationship can be computed explicitly. As they scale the regularization terms of the weighted LASSO, Lagrange multipliers serve as tuning parameters for the weighted LASSO, prompting the question of their potential effective use as tuning parameters in applications like MR image reconstruction and denoising. This work represents an initial step in this direction.

Keywords Tuning parameters · LASSO · Convex optimization · Lagrange duality · MRI · Compressed Sensing

Mathematics Subject Classification 15A29 · 47N10 · 47A52 · 49K99 · 49N45 · 65F20 · 65F22 · 92C55

1 Introduction

Basis Pursuit is a well-known convex minimization problem that was first introduced by F. Santosa and W. W. Symes in 1986, cf. [50], in its simplest formulation:

$$\text{minimize} \quad \|x\|_1 + \lambda \|Ax - b\|_2^2, \qquad x \in \mathbb{R}^n, \tag{1}$$

where $A \in \mathbb{R}^{m \times n}$, the so-called *design matrix*, and $b \in \mathbb{R}^m$ are fixed. The same problem was later applied to signal processing by S. S. Chen and D. Donoho in 1994, cf. [15]. In 1996, R. Tibshirani re-introduced it as linear regression method, under the name of LASSO. Namely, in [52], they consider the constrained minimization problem:

$$\text{minimize} \quad \|Ax - b\|_2^2, \qquad x \in \mathbb{R}^n, \ \|x\|_1 \le \tau, \tag{2}$$

for $\tau > 0$ and

$$\|x\|_1 := \sum_{j=1}^{n} |x_j|$$

is the ℓ_1 norm. We will discuss the equivalence between (1) and (2) in the following.

Mathematical analysis approaches to study LASSO problems in all their facets are not new, and the literature is so vast that we can only limit ourselves to mention a few

examples, cf. [5, 16, 18, 38, 42, 43, 51]. For instance, in [53, 54], the authors study representation theorems for the solutions of general problems:

$$\arg \min_{x} E(b, \nu(x)) + \gamma(\|x\|),$$

in the framework of Banach space theory, where E is a loss functional, ν is a so-called *measurement mapping*, γ is a strictly increasing convex function and $\|\cdot\|$ is a Banach norm, we refer to [53, Theorem 2], [54, Theorem 2, Theorem 3] for more precise statements. In [3], the authors use convex analysis and variational calculus to study regularity properties of the set-valued mapping:

$$(b, \lambda) \in \mathbb{R}^m \times (0, +\infty) \mapsto \arg \min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - b\|_2^2 + \lambda \|x\|_1.$$

The main purpose of this work is to shed new light on the analytic dependence between the *Lagrange multipliers*, understood as explained below, and the constraints of a specific version of a constrained *generalized LASSO* problem. In this version, the relationship can be explicitly computed under further assumptions on target function. Let us provide a detailed explanation and motivation for this interest. As aforementioned, in its simplest definition, LASSO consists in the minimization of the function:

$$\|Ax - b\|_2^2 + \lambda \|x\|_1, \tag{3}$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ is a measurement vector. Clearly, problems (1) and (3) have the same minimizers and, therefore, for the purposes of this work, we will consider them as the same minimization problem. In short, (1) and (3) can be interpreted as regularization problems, where the aim is to minimize simultaneously the fidelity term $\|Ax - b\|_2^2$, that measures noise, and the regularization term $\|x\|_1$, that enforces sparsity. Recall that a vector $x = (x_1, \ldots, x_n)$ is $s-$**sparse** if card$\{j : x_j \neq 0\} \leq s$. When s is clear from the context or irrelevant, we drop s and say that x is *sparse*. In several applications, x is not sparse itself, but it is sparse with respect to a so-called *sparsity-promoting transform* $\Phi : \mathbb{R}^n \to \mathbb{R}^N$. Stated differently, when Φx is known to be sparse, problem (3) can be generalized to:

$$\text{minimize} \quad \|Ax - b\|_2^2 + \lambda \|\Phi x\|_1, \qquad x \in \mathbb{R}^n, \tag{4}$$

i.e. the regularization term $\|x\|_1$ in (3) is replaced by $\|\Phi x\|_1$. The parameter $\lambda > 0$ in (3) acts as a tuning parameter that balances the contributions of the fidelity term $\|Ax - b\|_2^2$ and regularization addendum $\|\Phi x\|_1$: small values of λ lower the contribution of the regularization, strengthening the effect of the fidelity term; vice-versa, large values of λ make $\|Ax - b\|_2^2$ negligible and force $\|\Phi x\|_1$ to be small in order for the overall sum to be small. Consequently, solutions corresponding to $\lambda \ll 1$ will be noisy, being close to the set $A^{-1}b$, while solutions $x^{\#}$ corresponding to $\lambda \gg 1$ have $\Phi x^{\#}$ more sparse. From this perspective, estimates of tuning parameters for inverse problems can be performed pursuing different approaches. *A posteriori rules* can be used when some

a-priori knowledge on the amplitude of noise $e \in \mathbb{R}^m$ is available, say $\|e\|_2 \leq \varepsilon$. For instance, using Morozov's discrepancy principle, λ can be chosen so that a solution x_λ of (3) satisfies $\|Ax_\lambda - b\|_2 \leq \varepsilon$, cf. [9, 29, 33]. *A priori rules* require knowledge of noise level, as before, but also a-priori information on the regularity of the solution. For this reason, a-priori approaches are usually bad suited for applications, cf. [2]. Heuristic methods, such as the L-curve are also available, cf. [11, 28, 34]. The L-curve method consists of choosing the optimal tuning parameter empirically by tracing a trade-off curve (the L-curve), whereas the generalized cross-validation (GCV) is a well-performing method that requires high-dimensional matrix calculus, cf. [26, 31, 55]. Other non-standard methods can be found in [32, 45], where the parameter is chosen so that statistical properties of noise, such as whiteness, are optimized; an implementation that avoids the computation of matrix inverses can be found in [6]. CNN and other learning methods were deployed in [30, 41], while a more statistical point of view was adopted in [10].

However, the very reason why λ is interpreted as a trade-off between noise and sparsity, in (1) and (3), is that it depends on estimates that are usually unavailable, such as a-priori upper bounds for the ℓ_1 norm of the unknown vector, i.e. a-priori information on the sparsity of the solution, or upper bounds for the noise, cf. [42]. For $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $\eta \geq 0$, the function:

$$L(x, \lambda) = \|x\|_1 + \lambda(\|Ax - b\|_2^2 - \eta^2) \tag{5}$$

is the Lagrangian associated to the constrained minimization problem:

$$\text{minimize} \quad \|x\|_1, \qquad x \in \mathbb{R}^n, \|Ax - b\|_2^2 \leq \eta^2, \tag{6}$$

cf. [7]. Roughly speaking, this entails that (1) and (6) are equivalent, up to choosing:

$$\lambda = \lambda(\eta) \tag{7}$$

or, equivalently, $\eta = \eta(\lambda)$, properly. Please note that η may not be uniquely determined. We refer to [23, Proposition 3.2] for a more precise statement of this fact. Throughout this work, we call the parameter λ in (7) a *Lagrange multiplier* associated to (6), since it plays the same role of Lagrange multipliers in optimization problems with equality constraints. We will use this terminology in a more general setting, see Definition 2.7 below. Since a slightly modified proof of [23, Theorem 3.1] shows that a solution of (6), if unique, must be m-sparse, the ℓ_1 norm is said to *enforce sparsity*. For this reason, the Lagrange multipliers in (5) could be used, in principle, in an equivalent manner as *tuning parameters* for (1) to recover sparse vectors.

In the same way,

$$L(x, \lambda) = \|Ax - b\|_2^2 + \lambda(\|\Phi x\|_1 - \tau) \tag{8}$$

is the Lagrange function of the constrained problem:

$$\text{minimize} \quad \|Ax - b\|_2^2, \qquad x \in \mathbb{R}^n, \|\Phi x\|_1 \leq \tau. \tag{9}$$

A first question that may be addressed is whether the corresponding Lagrange multiplier λ of (9) could still be used as a tuning parameter in (4). If so, the relationship between the Lagrange multipliers and the constraints of the corresponding constrained problems could be useful in concrete applications, such as static and dynamic MRI, cf. [19, 24]. In MRI, indeed, vectors of interest are MR images, which tend to be approximately sparse with respect to the discrete Fourier transform (DFT), the discrete cosine transform (DCT) or the discrete wavelet transform (DWT), cf. [36]. This means that a solution of the generalized LASSO problem (4), where the design matrix A is a proxy of the acquisition methods properties (coil sensitivity, undersampling schemes and DFT), and b is an underdetermined, noisy measurement, will have a sparse regularization term, i.e. sparse Φx. We stress that (4) is known to admit a, in general not unique, solution for any choice of A, b, λ and Φ, and for the sake of completeness we report a proof, that uses only linear algebra, in the appendix. In order to exploit more a-priori knowledge on the structure of MRI data, (4) can be generalized further to consider target functions that are sum of more regularizing terms, cf. [22, 27, 44].

Let us note that sparsity is not always the correct assumption in MRI. For instance, dynamic MR images (e.g. a sequence of images of a moving organ, cf. [19, 24]) are highly compressible, rather than sparse, cf. [37]. This means that most of their coefficients with respect to some sparsity-promoting transform do not vanish, yet are small or negligible.

Surprisingly, it is easier to identify an equivalent of the relationship (7), between the parameter λ and the upper bound for the constraint, say η, when another weighted version of LASSO is considered. Namely, we aim to utilize convex analysis to compute the Lagrange multipliers for the constrained optimization problem:

$$\text{minimize} \quad \|Ax - b\|_2^2, \qquad x \in \mathbb{R}^n, \quad |x_j| \le \tau_j, \quad j = 1, \ldots, n. \qquad (10)$$

For given $\tau_1, \ldots, \tau_n > 0$ and a given minimizer $x^{\#}$ of (10) there exist $\lambda_1, \ldots, \lambda_n \ge 0$ such that $x^{\#}$ is also a minimizer of:

$$\text{minimize} \quad \|Ax - b\|_2^2 + \sum_{j=1}^{n} \lambda_j |x_j|, \qquad (11)$$

see [7, Section 5.3.2] or Theorem 2.4 below for a complete statement. Other weighted versions of this problem have been considered in the literature. For instance, in [41], the authors present a total variation (TV) regularization-based weighted LASSO for image denoising. Other references include [9], where the authors consider *space-variant* problems, such as:

$$\text{minimize} \quad \frac{1}{2} \|Ax - b\|_2^2 + \sum_{j=1}^{k} \lambda_k \left\| (Dx)_j \right\|_p,$$

where $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$, D is the discrete gradient, $p \in \{1, 2\}$, and $\lambda_1, \ldots, \lambda_k > 0$. In a certain sense, problem (11) can be considered as a space-variant problem, where

every component of the unknown vector is weighted by a different parameter. In [46], the author discusses the importance of space-variance in TV regularization, as a mathematical modeling which has the advantage of recovering a description of local features, which is lost by classical TV regularization, i.e. (4) with $\Phi = D$.

As we shall see, the relationship between these parameters is non-trivial if A is non orthogonal, due to the complicated geometry of (10). Loosely speaking, this is due to the fact that if $A^T A$ is non-diagonal, A shuffles the coordinates of x in such a way that each pair of sets $M_j := \{x \in \mathbb{R}^n : x_j = -\tau_j\}$ and $N_j := \{x \in \mathbb{R}^n : \frac{\partial}{\partial x_j}(\|Ax - b\|_2^2) = 0\}$ are no longer parallel.

To summarize, the results contained in this work serve as first steps towards the understanding of the analytical relationship between Lagrange multipliers, as defined in Definition 2.7 below, and tuning parameters for LASSO problems. This relationship is non-trivial, since it involves a-priori estimates, such as estimates of the ℓ_2-norm of the noise. Our interest is justified by the fact that, plugging the identity as regularization term in (4), the results mentioned above tell that the corresponding Lagrange multipliers yield to the recovery of sparse vectors. Since MR images are known to be sparse or compressible with respect to several sparsify-promoting transforms, the question that arises is whether it is possible to use Lagrange multipliers associated to (9) as tuning parameters to retrieve good-quality anatomical MR images via LASSO problems.

Overview. In Sect. 2, we establish preliminaries and notation. In Sect. 3 we compute the deterministic relationships between the parameters λ_j's and the τ_j's in order for problems (10) and (11) to be equivalent, under the following specific assumptions: given A such that $A^T A$ is diagonal, for instance when A is either a subsampling matrix, the Fourier transform matrix or the identity matrix, the Lagrange multipliers are explicitly given by:

$$\lambda_j^{\#} = 2 \|a_{*,j}\|_2^2 \left(\frac{|\langle b, a_{*,j}\rangle|}{\|a_{*,j}\|_2^2} - \tau_j \right) \chi_{\left[0, \frac{|\langle b,a_{*,j}\rangle|}{\|a_{*,j}\|_2^2}\right]}(\tau_j), \qquad (12)$$

where $a_{*,j}$ denotes the j-th column of A and $\chi_{\left[0, \frac{|\langle b,a_{*,j}\rangle|}{\|a_{*,j}\|_2^2}\right]}$ is the characteristic function

on $\left[0, \frac{|\langle b,a_{*,j}\rangle|}{\|a_{*,j}\|_2^2}\right]$, $j = 1, ..., n$. We also provide deterministic results for those cases where there is a control on the sign of the gradient of $\|Ax - b\|_2^2$, providing the explicit expression of the Lagrange multipliers under the assumption $\frac{\partial}{\partial x_j}(\|Ax - b\|_2^2) \leq 0$ for every $j = 1, \ldots, n$ in a properly defined hypercube. The conclusions are reported in Sect. 4.

2 Preliminaries and notation

Notation. For the theory of this section, we refer to [7, 23, 49] as reference therein. We denote by $\mathbb{R}^n$ the n-dimensional vector space of real column vectors, whereas

$\mathbb{R}^{m \times n}$ denotes the space of real $m \times n$ matrices. To ease the notation, if $x \in \mathbb{R}^n$, the notation $x = (x_1, ..., x_n)$ means that x is the column vector with coordinates $x_1, ..., x_n$. If $A \in \mathbb{R}^{m \times n}$, A^T denotes the transpose of A.

If $A \in \mathbb{R}^{m \times n}$, $\ker(A)$ and $\mathrm{Im}(A)$ denote the kernel and the image of A, respectively. $\mathcal{M}_n$ denotes the set of $n \times n$ signature matrices and, for $x \in \mathbb{R}^n$, $\mathrm{sgn}(x)$ denotes the set of all the possible signatures of x, see Sect. 3.2 below.

For $1 \le p < \infty$, the ℓ_p-norm on $\mathbb{R}^n$ is defined as:

$$\|x\|_p := \left(\sum_{j=1}^{n} |x_j|^p \right)^{1/p}, \qquad x \in \mathbb{R}^n,$$

whereas $\|x\|_\infty := \max_{j=1,...,n} |x_j|$. We denote by $\langle \cdot, \cdot \rangle$ the canonical inner product of $\mathbb{R}^n$, i.e.

$$\langle x, y \rangle = x^T y = \sum_{j=1}^{n} x_j y_j, \qquad x, y \in \mathbb{R}^n.$$

If $x \in \mathbb{R}^n$, x^+ is its positive part, i.e. $x^+ \in \mathbb{R}^n$ has coordinates $(x^+)_j = \max\{x_j, 0\}$ ($j = 1, ..., n$). If $\Omega \subseteq \mathbb{R}^n$, $\Omega^\perp$ denotes its orthogonal complement. For vectors $x, y \in \mathbb{R}^n, x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n)$, the notation $x \preceq y$ means that $x_j \le y_j$ for every $j = 1, \ldots, n$. Analogously, $x \prec y$ if $x_j < y_j$ for every $j = 1, \ldots, n$. The relationships $x \succeq y$ and $x \succ y$ are defined similarly.

We always consider $\mathbb{R}^n$ endowed with the Euclidean topology. If $\Omega \subseteq \mathbb{R}^n$, $\mathring{\Omega}$ denotes the interior of Ω and $\partial \Omega$ denotes the boundary of Ω. If g is a real-valued function defined on an open neighbourhood of $x_0 \in \mathbb{R}^n$, $\partial g(x_0)$ denotes the subdifferential of g at x_0, see Sect. 2.2 below for the definition of subdifferential. Using the same notation to denote both the boundary of a set and the subdifferential of a function shall not cause confusion. If $\Omega \subseteq \mathbb{R}^n$, a-int$(\Omega)$ denotes the algebraic interior of Ω, see Definition 3.8 below. If g is a function and Ω is a subset of its domain, $g|_\Omega$ denotes the restriction of g to Ω. Finally, if $\Omega \subseteq \mathbb{R}^n$, χ_Ω denotes the characteristic function of Ω.

2.1 Lagrange Duality

Consider a constrained optimization problem in the form:

$$\text{minimize } F_0(x), \qquad \Psi x = y, \ F_l(x) \le b_l, \ l = 1, \ldots, M, \tag{13}$$

where $\Psi \in \mathbb{R}^{m \times n}$, $y \in \mathbb{R}^m$ and $F_0, F_1, \ldots, F_M : \mathbb{R}^n \to (-\infty, +\infty]$ are convex. We always assume that a minimizer of (13) exists.

A point $x \in \mathbb{R}^n$ is called **feasible** if it belongs to the constraints, that is if:

$$x \in K := \left\{ \zeta \in \mathbb{R}^n \ : \ \Psi \zeta = y \text{ and } F_l(\zeta) \le b_l, \ l = 1, \ldots, M \right\} \tag{14}$$

and K is called the **set of feasible points**. To avoid triviality, we always assume $K \neq \varnothing$, in which case problem (13) is called **feasible**. In view of the definition of K, problem (13) can be implicitly written as:

$$\text{minimize} \quad F_0(x), \qquad x \in K.$$

Convex problems such as (6) and (10) can be approached by considering their Lagrange formulation, see Sect. 2.3 below. The **Lagrange function** related to (13) is the function $L : \mathbb{R}^n \times \mathbb{R}^m \times [0, +\infty)^M \to (-\infty, +\infty]$ defined as:

$$L(x, \xi, \lambda) := F_0(x) + \langle \xi, \Psi x - y \rangle + \sum_{l=1}^{M} \lambda_l (F_l(x) - b_l).$$

Observe that for all ξ, λ and $x \in K$:

$$L(x, \xi, \lambda) = F_0(x) + \underbrace{\langle \xi, \Psi x - y \rangle}_{=0} + \sum_{l=1}^{M} \underbrace{\lambda_l}_{\geq 0} \underbrace{(F_l(x) - b_l)}_{\leq 0} \leq F_0(x),$$

so that:

$$\inf_{x \in \mathbb{R}^n} L(x, \xi, \lambda) \leq \inf_{x \in K} L(x, \xi, \lambda) \leq \inf_{x \in K} F_0(x). \tag{15}$$

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

$$H(\xi, \lambda) := \inf_{x \in \mathbb{R}^n} L(x, \xi, \lambda)$$

is called **Lagrange dual function**.

Inequalities (15) read as:

$$H(\xi, \lambda) \leq \inf_{x \in K} F_0(x) \tag{16}$$

for all $\xi \in \mathbb{R}^m$ and all $\lambda \in [0, +\infty)^M$. Stating (16) differently, we have the **weak duality inequality**:

$$\sup_{\substack{\xi \in \mathbb{R}^m \\ \lambda \geq 0}} H(\xi, \lambda) \leq \inf_{x \in K} F_0(x). \tag{W}$$

We point out that (W) is equivalent to:

$$\sup_{\xi, \lambda} \inf_x L(x, \xi, \lambda) \leq \inf_x \sup_{\xi, \lambda} L(x, \xi, \lambda) \tag{17}$$

(see [7, Subsection 5.4.1]).

We are interested in computing the parameters (ξ, λ) such that (W) is an equality, in which case (W) becomes:

$$\sup_{\substack{\xi \in \mathbb{R}^m \\ \lambda \geq 0}} H(\xi, \lambda) = \inf_{x \in K} F_0(x), \tag{S}$$

so that **strong duality** (S) holds for problem (13).

2.2 Subdifferential

Definition 2.2 (Subdifferential) Let $\Omega \subseteq \mathbb{R}^n$ be open and $g : \Omega \to \mathbb{R}$. Let $x_0 \in \Omega$. The **subdifferential** of g at x_0 is the set:

$$\partial g(x_0) := \{v \in \mathbb{R}^n \ g(x) \geq g(x_0) + v^T (x - x_0) \ \forall x \in \Omega\}.$$

We refer to any $v \in \partial g(x_0)$ as a **subgradient** of g at x_0.

We will use the following proposition.

Proposition 2.3 *Let $\Omega \subseteq \mathbb{R}^n$ be open and $g : \Omega \to \mathbb{R}$ be convex and continuous on Ω. Let $x_0 \in \Omega$. Then, $\partial g(x_0) \neq \varnothing$.*

2.3 Lagrange Formulation of Constrained Problems

Under the notation above, let $F(x) := (F_1(x), ..., F_M(x))$. In the convex framework, if the constraint $F(x) \preceq b$ does not reduce to $F(x) = b$, namely if for all $l = 1, \ldots, M$ the inequality $F_l(x) < b_l$ holds for some $x \in \mathbb{R}^n$, then strong duality holds.

Theorem 2.4 (Cf. [7], Section 5.3.2) *Assume that $F_0, F_1, \ldots, F_M$ are convex functions defined on $\mathbb{R}^n$. Let $x^\#$ be such that $F_0(x^\#) = \inf_{x \in \mathbb{R}^n} F_0(x)$. If:*

   (i) *there exists $\tilde{x} \in \mathbb{R}^n$ such that $\Psi\tilde{x} = y$ and $F(\tilde{x}) \prec b$ or,*
   (ii) *in absence of inequality constraints, if $K \neq \varnothing$ (i.e. if there exists $\tilde{x} \in \mathbb{R}^n$ such that $\Psi\tilde{x} = y$),*

then, there exists $(\xi^\#, \lambda^\#) \in \mathbb{R}^m \times [0, +\infty)^M$ such that $H(\xi^\#, \lambda^\#) = \sup_{\xi, \lambda} H(\xi, \lambda)$ and $H(\xi^\#, \lambda^\#) = F_0(x^\#)$.

The proof of Theorem 2.4 contains the fundamental construction we will use in the next sections and we report it for this reason. We refer to [7, Subsection 5.3.2] for the complete proof. First, we need a result from functional analysis, which is well-known as (geometrical) *Hahn-Banach theorem*.

Definition 2.5 (Separating hyperplane) Consider two subsets $\mathcal{A}, \mathcal{B} \subseteq \mathbb{R}^n$. A hyperplane $\Gamma := \{x \in \mathbb{R}^n \ \langle \xi, x \rangle = \alpha\}$ satisfying:

$$\langle \xi, x \rangle \leq \alpha < \langle \xi, y \rangle, \qquad x \in \mathcal{A}, \quad y \in \mathcal{B}, \tag{18}$$

is a **separating hyperplane** between $\mathcal{A}$ and $\mathcal{B}$.

Theorem 2.6 (Cf. [49] Theorem 3.4) *Let $\mathcal{A}, \mathcal{B} \subset \mathbb{R}^n$ be two convex and disjoint subsets. If $\mathcal{B}$ is open, there exists $\xi \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ such that (18) holds for all $x \in \mathcal{A}$ and all $y \in \mathcal{B}$.*

Idea of the proof of Theorem 2.4 First, one assumes that A has full row-rank. Moreover, one reduces to consider the situation in which $p^* := \inf_{x \in K} F_0(x) > -\infty$, otherwise the assertion is trivial.

Consider the set:

$$\mathcal{G} := \left\{ (F(x) - b, \Psi x - y, F_0(x)) \in \mathbb{R}^M \times \mathbb{R}^m \times \mathbb{R} \ : \ x \in \mathbb{R}^n \right\}, \quad (19)$$

where, with an abuse of notation, $\Psi x - y$ denotes the row vector with the same (ordered) entries of $\Psi x - y$, and $\mathcal{A}$ be defined as the epigraph:

$$\mathcal{A} := \mathcal{G} + ((\mathbb{R}_{\geq 0})^M \times \mathbb{R}^m \times \mathbb{R}_{\geq 0}) =$$
$$= \Big\{ (u, v, t) \in \mathbb{R}^M \times \mathbb{R}^m \times \mathbb{R} \ : \ u \succeq F(x) - b, \quad (20)$$
$$v = \Psi x - y, \ t \geq F_0(x) \ for \ some \ x \in \mathbb{R}^n \Big\}.$$

It is easy to verify that if $F_0, F_1, \ldots, F_M$ are convex, than $\mathcal{A}$ is convex. Then, consider the set:

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

$\mathcal{A}$ and $\mathcal{B}$ are clearly disjoint, $\mathcal{B}$ (which is an open half-line) being trivially convex and open. Therefore, the assumptions of Theorem 2.6 are satisfied and we conclude that there exists a triple of parameters $(\tilde{\lambda}, \tilde{\xi}, \mu) \neq 0$ and $\alpha \in \mathbb{R}$ such that:

$$(u, v, t) \in \mathcal{A} \implies \tilde{\lambda}^T u + \langle \tilde{\xi}, v \rangle + \mu t \geq \alpha, \quad (21)$$
$$(u, v, t) \in \mathcal{B} \implies \tilde{\lambda}^T u + \langle \tilde{\xi}, v \rangle + \mu t < \alpha. \quad (22)$$

It is easy to see that the definition of $\mathcal{A}$, together with (21), imply that $\tilde{\lambda}_l \geq 0$ for all $l = 1, \ldots, M$ and $\mu \geq 0$. Also, applying the definition of $\mathcal{B}$ to (22), one finds that $\mu t < \alpha$ for all $t < p^*$, which implies that $\mu p^* \leq \alpha$. Therefore, for all $x \in \mathbb{R}^n$,

$$\sum_{l=1}^{M} \tilde{\lambda}_l (F_l(x) - b_l) + \langle \tilde{\xi}, \Psi x - y \rangle + \mu F_0(x) \geq \alpha \geq \mu p^*. \quad (23)$$

If $\mu > 0$, then (23) gives that $L(x, \tilde{\xi}/\mu, \tilde{\lambda}/\mu) \geq p^*$ for all $x \in \mathbb{R}^n$, which implies that $H(\tilde{\xi}/\mu, \tilde{\lambda}/\mu) \geq p^*$. Since the other inequality holds trivially by the weak duality inequality, we conclude that $H(\tilde{\xi}/\mu, \tilde{\lambda}/\mu) = p^*$. Finally, using the assumptions on the rank of Ψ and on the existence of a point satisfying the strict inequality constraint, one proves by contradiction that it must be $\mu > 0$. $\qquad\square$

Definition 2.7 (Lagrange Multipliers) We refer to a couple $(\xi^{\#}, \lambda^{\#}) \in \mathbb{R}^m \times [0, +\infty)^M$ as to **Lagrange multipliers** for the problem (13) if $(\xi^{\#}, \lambda^{\#})$ attend the supremum in (S).

As a consequence of Theorem 2.4, we have the following result, which relates the minimizers of (13) and those of the dual problem $\max_{\xi, \lambda} H(\xi, \lambda)$, providing also the Lagrange multipliers, that may not be unique.

Corollary 2.8 (Cf. [23] Theorem B.28) *Let $F_0 : \mathbb{R}^n \to [0, +\infty)$ and $\phi : [0, +\infty) \to \mathbb{R}$ be such that ϕ is monotonically increasing and $\phi \circ F_0$ is convex. Let $\tau_j > 0$ $(j = 1, \ldots, M)$ and $\psi_j : \mathbb{R}^n \to \mathbb{R}$ $(j = 1, \ldots, M)$ be convex functions such that $\psi_j^{-1}([0, \tau_j)) \neq \varnothing$ for all $j = 1, \ldots, M$. Let $x^{\#}$ be a minimizer of the problem:*

$$\text{minimize} \quad F_0(x), \qquad x \in \mathbb{R}^n \ \psi(x) \preceq \tau, \tag{24}$$

where $\tau = (\tau_1, \ldots, \tau_M)$. Then, there exist $\lambda_j \geq 0$ $(j = 1, \ldots, M)$ such that $x^{\#}$ is a minimizer of:

$$\text{minimize} \quad \phi(F_0(x)) + \sum_{j=1}^{M} \lambda_j \psi_j(x). \tag{25}$$

Proof Since ϕ is monotonically increasing, (24) is obviously equivalent to:

$$\text{minimize} \quad \phi(F_0(x)), \qquad x \in \mathbb{R}^n \ \psi_j(x) \leq \tau_j, \tag{26}$$

$(j = 1, \ldots, M)$ whose Lagrangian is given by:

$$L(x, \lambda) = \phi(F_0(x)) + \sum_{j=1}^{M} \lambda_j(\psi_j(x) - \tau_j). \tag{27}$$

By the assumption, $\phi \circ F_0$ and each ψ_j are convex and the inequalities $\psi_j(\tilde{x}) < \tau_j$ are satisfied by some $\tilde{x} \in \mathbb{R}^n$ (observe that here we need $\tau_j > 0$), so we can apply Theorem 2.4 to get $H(\lambda^{\#}) = \phi(F_0(x^{\#}))$ for some $\lambda^{\#} \in [0, +\infty)^M$. By (17), for all $x \in \mathbb{R}^n$:

$$L(x^{\#}, \lambda^{\#}) \leq L(x, \lambda^{\#}),$$

so that $x^{\#}$ is also a minimizer of the function $x \in \mathbb{R}^n \mapsto L(x, \lambda^{\#})$. Since the constant terms $-\lambda_j \tau_j$ in (27) do not affect the set of minimizers, we have that $x^{\#}$ is a minimizer of:

$$\text{minimize} \quad \phi(F_0(x)) + \sum_{j=1}^{M} \lambda_j^{\#}(\psi_j(x) - \tau), \qquad x \in \mathbb{R}^n.$$

$\square$

Remark 2.9 Theorem 2.6 has a complex version that holds with $\Re\langle z, w \rangle = \Re\left(\sum_{j=1}^{n} \overline{z_j} w_j\right)$ ($\Re$ denotes the real part of a complex number) instead of $\langle \cdot, \cdot \rangle$. In particular, the entire theory presented in this work is applicable in the complex framework as well. This extension involves replacing the canonical real inner product of $\mathbb{R}^n$ with the real inner product on $\mathbb{C}^n$ defined above. Therefore, we do not need to study the complex case separately, as only the structure of $\mathbb{C}^n$ as a real vector space is involved.

Remark 2.10 To sum up, Theorem 2.4 and Corollary 2.8 together tell that, up to the sign, the coefficients of any hyperplane separating the two sets:

$$\mathcal{A} = \left\{ (u, t) \in \mathbb{R}^{M+1} \; \middle| \; u \geq F(x) - b, \; t \geq F_0(x) \; for \; x \in \mathbb{R}^n \right\}$$

and

$$\mathcal{B} = \left\{ (0, t) \in \mathbb{R}^{M+1} \; \middle| \; t < \inf_{x \in K} F_0(x) \right\}$$

define Lagrange multipliers for problem (13), in absence of equality constraints, i.e. if $y = 0$ and $\Psi = 0$ in (13). This is the geometric idea that we will apply in the following sections to the weighted LASSO.

3 The Weighted LASSO

Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $\tau_1, \ldots, \tau_n \geq 0$. We denote with $a_{*,j}$ the j-th column of A and set $b = (b_1, \ldots, b_m)$. We consider the constrained minimization problem:

$$\text{minimize} \quad \|Ax - b\|_2^2, \qquad x \in \mathbb{R}^n, \; |x_j| \leq \tau_j, \; j = 1, \ldots, n. \tag{28}$$

We also assume that $\tau_j \neq 0$ for all $j = 1, \ldots, n$. In fact, if $\tau_j = 0$ for some $j = 1, \ldots, n$, then the solution $x = (x_1, \ldots, x_n)$ has $x_j = 0$. In this case, problem (28) reduces to

$$\text{minimize} \quad \left\| \tilde{A}y - b \right\|_2^2, \qquad y \in \mathbb{R}^{n-r}, \; |y_{i_j}| \leq \tau_{i_j}, \; j = 1, \ldots, n - r, \tag{29}$$

where $r = \text{card}\{j : \tau_j = 0\} \leq m$, $J = \{1 \leq i_1 < \ldots < i_{n-r} \leq n\} := \{j : \tau_j \neq 0\}$ and $\tilde{A} = (a_{*,j})_{j \in J} \in \mathbb{R}^{m \times (n-r)}$.

Let K denote the set of the feasible points of problem (28), that is:

$$K = \{x \in \mathbb{R}^n \; : \; |x_j| \leq \tau_j \; \forall j = 1, \ldots, n\} \tag{30}$$

and consider the Lagrange function associated to (28), i.e.

$$L(x, \lambda_1, \ldots, \lambda_n) = \|Ax - b\|_2^2 + \sum_{j=1}^{n} \lambda_j(|x_j| - \tau_j). \tag{31}$$

We are interested in a vector of Lagrange multipliers $\lambda^{\#} \succeq 0$ for (28). Based on the proofs of Theorem 2.4 and Corollary 2.8, $\lambda^{\#}$ can be chosen as the direction of any hyperplane separating the sets:

$$
\mathcal{A} = \Big\{ (u, t) \in \mathbb{R}^n \times \mathbb{R} \; : \; u_l \geq |x_l| - \tau_l \; (l = 1, \ldots, n),
$$
$$
t \geq \|Ax - b\|_2^2 \text{ for some } x \in \mathbb{R}^n \Big\}
\tag{32}
$$

and

$$
\mathcal{B} = \Big\{ (0, t) \in \mathbb{R}^n \times \mathbb{R} \; : \; t < p^* \Big\}
\tag{33}
$$

where $p^* := \inf_{x \in K} \|Ax - b\|_2^2$.

3.1 The Scalar Case

To clarify the general procedure, we focus on the simple case $m = n = 1$ first, in which (28) becomes:

$$
\text{minimize} \quad (Ax - b)^2, \qquad x \in \mathbb{R}, \; |x| \leq \tau,
\tag{34}
$$

where $A \in \mathbb{R} \setminus \{0\}$ and $b \in \mathbb{R}$. To find the Lagrange multipliers, we consider the set $\mathcal{G}$ of points $(u, t) \in \mathbb{R}^2$ that satisfy:

$$
\begin{cases}
u = |x| - \tau, \\
t = (Ax - b)^2,
\end{cases}
$$

which give a curve of the half-plane $U = \{ (u, t) \in \mathbb{R}^2 \; : \; u \geq -\tau, \; t \geq 0 \}$ parametrized by $x \in \mathbb{R}$. More precisely:

- if $x \geq 0$,

$$
\begin{cases}
x = u + \tau, \\
t = \big(A(u + \tau) - b \big)^2 = (Au + (A\tau - b))^2,
\end{cases}
$$

  which is a branch of parabola in U with vertex in $(\frac{b}{A} - \tau, 0)$.
- If $x < 0$

$$
\begin{cases}
x = -u - \tau, \\
t = \big(-A(u + \tau) - b \big)^2 = (Au + (A\tau + b))^2,
\end{cases}
$$

  which is, again, a branch of parabola in U, having its vertex in $(-\frac{b}{A} - \tau, 0)$.

Proposition 3.1 *Let $\tau > 0$, $A \in \mathbb{R}\backslash\{0\}$, $b \in \mathbb{R}$. A Lagrange multiplier for (34) is given by:*

$$\lambda^{\#} = \begin{cases} 2A^2(|b/A| - \tau) & if\, 0 < \tau < |b/A|, \\ 0 & if\, \tau \geq |b/A| \end{cases} = 2A^2(|b/A| - \tau)^{+}.$$

Namely, if $x^{\#}$ is a minimizer of (34), then it is also a minimizer for the problem:

$$minimize \quad (Ax - b)^2 + \lambda^{\#}|x|, \qquad x \in \mathbb{R}.$$

3.2 Properties of $\mathcal{A}$

Consider $A \in \mathbb{R}^{m \times n}$ and $b = (b_1, ..., b_m) \in \mathbb{R}^m$, with:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}.$$

We consider the problem (28) and the associated Lagrange function:

$$L(x, \lambda) := \|Ax - b\|_2^2 + \sum_{j=1}^{n} \lambda_j(|x_j| - \tau_j). \tag{35}$$

Recall that p^* was defined as $p^* := \min_{x \in K} \|Ax - b\|_2^2$, being K the set of the points $x \in \mathbb{R}^n$ such that $|x_j| \leq \tau_j$ for all $j = 1, \ldots, n$. It is not difficult to verify that:

$$p^* = \inf\left\{t \in \mathbb{R} \,:\, (u, t) \in \mathcal{G}, \, u_j \leq 0 \; \forall j = 1, \ldots, n\right\}. \tag{36}$$

Let $\mathcal{M}_n$ be the set of the n-dimensional signature matrices, that are the diagonal matrices $S = (s_{ij})_{i,j=1}^{n} \in \mathbb{R}^{n \times n}$ such that $|s_{jj}| = 1$ for all $j = 1, \ldots, n$. Observe that if $S \in \mathcal{M}_n$, then $S^2 = I_{n \times n}$, where $I_{n \times n}$ denotes the identity matrix in $\mathbb{R}^{n \times n}$, in particular S is invertible with $S^{-1} = S$. If $x \in \mathbb{R}^n$ and $S \in \mathcal{M}_n$ is such that $Sx \in \prod_{j=1}^{n}[0, +\infty)$, we write $S \in \text{sgn}(x)$.

Lemma 3.2 *Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $\tau_j > 0$ for $j = 1, \ldots, n$. Let $S \in \mathcal{M}_n$. There exists $u \in \prod_{j=1}^{n}[-\tau_j, 0]$ such that $ASu + AS\tau - b = 0$ if and only if $S \in \text{sgn}(x)$ for some $x \in \mathbb{R}^n$ such that $Ax = b$ and $|x_j| \leq \tau_j$.*

Proof Assume that there exists $u \in \prod_{j=1}^{n}[-\tau_j, 0]$ such that $ASu + AS\tau - b = 0$ and let $x := S(u + \tau)$. Then, $Sx = u + \tau \in \prod_{j=1}^{n}[0, \tau_j]$, so that $S \in \text{sgn}(x)$, $|x_j| \leq \tau_j$ for all $j = 1, \ldots, n$ and

$$0 = AS(u + \tau) - b = Ax - b.$$

Vice versa, assume that $Ax = b$ for some $x \in \prod_{j=1}^{n}[0, \tau_j]$. Let $S \in \text{sgn}(x)$ and $u := Sx - \tau$. Then, $u \in \prod_{j=1}^{n}[-\tau_j, 0]$ and

$$0 = Ax - b = A(Su + \tau) - b = ASu + AS\tau - b.$$

$\square$

Recall the definitions of the two sets $\mathcal{A}$ and $\mathcal{B}$ given in (32) and (33) respectively. First, if $\mathcal{G}$ is the set of the points $(u, t) \in \mathbb{R}^{n+1}$ such that:

$$\begin{cases} u_j = |x_j| - \tau_j & j = 1, \ldots, n, \\ t = \|Ax - b\|_2^2, \end{cases} \tag{37}$$

for some $x \in \mathbb{R}^n$, then

$$\mathcal{A} = \mathcal{G} + [0, +\infty)^{n+1},$$

that is, $(u, t) \in \mathcal{A}$ if and only if

$$\begin{cases} u_j \geq |x_j| - \tau_j & j = 1, \ldots, n, \\ t \geq \|Ax - b\|_2^2, \end{cases} \tag{38}$$

for some $x \in \mathbb{R}^n$. Finally, $(u, t) \in \mathcal{B}$ if and only if $t < p^* = \min_{x_j \leq \tau_j} \|Ax - b\|_2^2$.

We will prove that the equations (37) defining $\mathcal{G}$ can be written in terms of $\mathcal{M}_n$.

Lemma 3.3 *Let $\tau_1, \ldots, \tau_n > 0$ and let $\mathcal{G}$ be the set of points satisfying (37). Then,*

(i) *$\mathcal{G}$ is closed.*

(ii) *$(u, p^*) \in \mathcal{G}$ for some $u \in \mathbb{R}^n$ such that $-\tau_j \leq u_j \leq 0$ for all $j = 1, \ldots, n$. Moreover, $p^* = \min \left\{ t \in \mathbb{R} : (u, t) \in \mathcal{G}, \ u_j \leq 0 \ \forall j = 1, \ldots, n \right\}$.*

(iii) *For every $(u, t) \in \mathcal{G}$ there exists $S \in \mathcal{M}_n$ such that $t = \|ASu + (AS\tau - b)\|_2^2$. Viceversa, if $t = \|ASu + (AS\tau - b)\|_2^2$ for some $u \in \mathbb{R}^n$ such that $u_j \geq -\tau_j$ and some $S \in \mathcal{M}_n$, then $(u, t) \in \mathcal{G}$.*

Proof We prove that $\mathcal{G}$ is closed. For, let $(u^k, t^k) \in \mathcal{G}$ converge to $(u, t) \in \mathbb{R}^{n+1}$. We prove that $(u, t) \in \mathcal{G}$. Let $x^k \in \mathbb{R}^n$ be such that (37) is satisfied for (u^k, t^k). Then, $|x_j^k| = u_j^k - \tau_j \leq u_j + 1 - \tau_j$ for j sufficiently large. In particular, the sequence $\{x^k\}_k$ is bounded and, thus, it converges up to subsequences. Without loss of generality, we may assume that $(x^k)_k$ converges to $x := \lim_{k \to +\infty} x^k$ in $\mathbb{R}^n$. Then, for all $j = 1, \ldots, n$,

$$|x_j| = \lim_{k \to +\infty} |x_j^k| = \lim_{k \to +\infty} u_j^k + \tau_j = u_j + \tau_j$$

and, by continuity,

$$\|Ax - b\|_2^2 = \lim_{k \to +\infty} \left\|Ax^k - b\right\|_2^2 = \lim_{k \to +\infty} t^k = t.$$

This proves that $(u, t) \in \mathcal{G}$ and, thus, that $\mathcal{G}$ is closed. (ii) follows by (i) and (36).

It remains to check (iii). If $(u, t) \in \mathcal{G}$, there exists $x \in \mathbb{R}^n$ satisfying (37). Let $S \in \mathcal{M}_n$ be such that $|x| = Sx$, where $|x| := (|x_1|, \ldots, |x_n|)$. Then, using the fact that $S^{-1} = S$,

$$|x| = u + \tau \implies Sx = (u + \tau) \implies x = S(u + \tau).$$

By the last equation of (37), we have:

$$t = \|Ax - b\|_2^2 = \|ASu + (AS\tau - b)\|_2^2.$$

Viceversa, assume that $t = \|ASu + (AS\tau - b)\|_2^2$ for some $S \in \mathcal{M}_n$ and $u \in \mathbb{R}^n$ is such that $u \succeq -\tau$. Let $x := S(u + \tau)$, then $|x_j| = |u_j + \tau_j| = u_j + \tau_j$ for all $j = 1, \ldots, n$ and $t = \|Ax - b\|_2^2$. This proves that $(u, t) \in \mathcal{G}$ and the proof of (iii) is concluded. $\qquad\square$

Lemma 3.4 *Let* $u \in \prod_{j=1}^n [-\tau_j, +\infty)$,

$$h_G(u) := \min_{S \in \mathcal{M}_n} \|ASu + AS\tau - b\|_2^2 \tag{39}$$

and

$$g_G(u) := \min_{(u,s) \in \mathcal{G}} s.$$

Then, $h_G(u) = g_G(u)$.

Proof By Lemma 3.3 (iii), if $(u, s) \in \mathcal{G}$, then $s = \|AS_0 u + AS_0 \tau - b\|_2^2$ for some $S_0 \in \mathcal{M}_n$. Hence,

$$h_G(u) = \min_{S \in \mathcal{M}_n} \|ASu + AS\tau - b\|_2^2 \leq \|AS_0 u + AS_0 \tau - b\|_2^2 = s$$

for all s such that $(u, s) \in \mathcal{G}$. Taking the minimum, we get $h_G(u) \leq g_G(u)$. On the other hand, $(u, h_G(u)) \in \mathcal{G}$ by Lemma 3.3 (iii). Therefore, $g_G(u) \leq h_G(u)$ by definition of g_G. $\qquad\square$

Lemma 3.5 *Let* $\mathcal{G}$ *be the set of points satisfying* (37) *and* $\mathcal{A}$ *be the set of points satisfying* (38). *Then,*

  (i) $\mathcal{G} \subseteq \mathcal{A}$;
  (ii) $\mathcal{A}$ *is closed.*

Proof (i) is obvious. We prove (ii).

Let $(u^k, t^k) \in \mathcal{A}$ be a sequence such that $(u^k, t^k) \xrightarrow[k \to +\infty]{} (u, t)$ in $\mathbb{R}^{n+1}$. We need to prove that $(u, t) \in \mathcal{A}$. For all k, let $x^k \in \mathbb{R}^n$ be such that:

$$\begin{cases} u_1^k \geq |x_1^k| - \tau_1, \\ \vdots \\ u_n^k \geq |x_n^k| - \tau_n, \\ t^k \geq \|Ax^k - b\|_2^2. \end{cases}$$

The sequence $\{x^k\}_k$ is bounded, in fact for all $j = 1, \ldots, n$, $|x_j^k| \leq u_j^k + \tau_j \leq u_j + 1 + \tau_j$ for k sufficiently large. Therefore, up to subsequences, we can assume $x^k \xrightarrow[k \to +\infty]{} x$ in $\mathbb{R}^n$. For all $j = 1, \ldots, n$,

$$|x_j| = \lim_{k \to +\infty} |x_j^k| \leq \lim_{k \to +\infty} u_j^k + \tau_j = u_j + \tau_j.$$

Moreover, by continuity,

$$\|Ax - b\|_2^2 = \lim_{k \to +\infty} \|Ax^k - b\|_2^2 \leq \lim_{k \to +\infty} t^k = t.$$

$\square$

Lemma 3.6 *Let $\mathcal{A}$ be the set of points satisfying* (38). *Then,*

(i) $\mathcal{A}$ is the epigraph of a convex non-negative function function $g : \prod_{j=1}^n [-\tau_j, +\infty) \to \mathbb{R}$ which is continuous in $\prod_{j=1}^n (-\tau_j, +\infty)$;

(ii) $\partial g(0) \neq \varnothing$;

(iii) $g(u) = 0$ if and only if $(u, t) \in \mathcal{A}$ for all $t \geq 0$.

Proof First, observe that $\mathcal{A} \subseteq \{(u, t) \ t \geq 0\}$ since $t \geq \|Ax - b\|_2^2 \geq 0$ for some $x \in \mathbb{R}^n$ whenever $(u, t) \in \mathcal{A}$.

For the sake of completeness, we check that $\mathcal{A}$ is the epigraph of the function:

$$g(u) = \min_{(u,s) \in \mathcal{A}} s \quad \left(u \in \prod_j [-\tau_j, +\infty) \right), \tag{40}$$

which is well defined by Lemma 3.5.

By the observation at the beginning of the proof, $g(u) \geq 0$. Let

$$epi(g) := \{(u, t) \ t \geq g(u)\}$$

be the epigraph of g. If $(u, t) \in \mathcal{A}$, then $t \geq \min_{(u,s) \in \mathcal{A}} s = g(u)$, this means that $(u, t) \in epi(g)$. On the other hand, if $(u, t) \in epi(g)$, then $t \geq s$ for some $(u, s) \in \mathcal{A}$. But, if $t \geq s$ (and $(u, s) \in \mathcal{A}$), then $(u, t) \in \mathcal{A}$ as well, since $\mathcal{A}$ contains the vertical upper half-lines having their origins in (u, s), namely $(u, s) + (\{0\} \times [0, +\infty))$.

This proves that $\mathcal{A}$ is an epigraph. Moreover, g is convex because $\mathcal{A}$ is convex (see [48] Proposition 2.4). The continuity of g on $\prod_j(-\tau_j, +\infty)$ follows from [47], Theorem 10.1. This proves (i).

Moreover, since $\tau_j > 0$ for all $j = 1, \ldots, n$, $0 \in \mathbb{R}^n$ is an interior point of $\prod_j[-\tau_j, +\infty)$. Since g is continuous and convex in $\prod_j(-\tau_j, +\infty)$, the subdifferential of g in 0 is non-empty and (ii) follows.

To prove (iii), assume that $g(u) = 0$. Then, $\min_{(u,s)\in\mathcal{A}} s = 0$ implies $(u, 0) \in \mathcal{A}$. Since for all $t \geq 0$, $(u, 0) + (\{0\} \times [0, +\infty)) \in \mathcal{A}$, we have that $(u, t) \in \mathcal{A}$ for all $t \geq 0$. For the converse, assume that $(u, t) \in \mathcal{A}$ for all $t \geq 0$. Then, $(u, 0) \in \mathcal{A}$, so that (by the non-negativity of g) $0 \leq g(u) \leq 0$. This proves the equivalence in (iii). $\square$

Remark 3.7 As we observed in the general theory situation, $(0, s) \in \mathcal{A}$ if and only if $s \geq p^*$. This tells that $g(0) = p^*$ and $(0, p^*) \in \mathcal{A}$.

We want to prove formally that $g(u)$ defines the boundary $\partial\mathcal{A}$ of $\mathcal{A}$ in a neighborhood of $u = 0$ and, then, find an explicit formula for $g(u)$. Observe that, $\mathcal{A} = \partial\mathcal{A} \cup \mathring{\mathcal{A}}$, where $\mathring{\mathcal{A}}$ denotes the topologic interior of $\mathcal{A}$. Since $\mathcal{A}$ is closed and convex in $\mathbb{R}^n$, $\mathring{\mathcal{A}}$ coincides with the algebraic interior of $\mathcal{A}$, which is defined as follows:

Definition 3.8 Let X be a vector space and $\mathcal{A} \subseteq X$ be a subset. The **algebraic interior** of $\mathcal{A}$ is defined as:

$$\text{a-int}(\mathcal{A}) := \{a \in \mathcal{A} \ \forall x \in X \ \exists \varepsilon_x > 0 \ s.t. \ a + tx \in \mathcal{A} \ \forall t \in (-\varepsilon_x, \varepsilon_x)\}.$$

Lemma 3.9 *Let $\mathcal{A}$ be as in Lemma 3.5. Then,*

$$\begin{aligned} \partial\mathcal{A} = &\{(u, t) \in \mathcal{A} \ : \ t = g(u), \ u_j > -\tau_j \ \forall j = 1, \ldots, n\} \cup \\ &\cup \{(u, t) \in \mathcal{A} \ : \ u_j = -\tau_j \ for \ some \ j = 1, \ldots, n\} \end{aligned} \tag{41}$$

and the union is disjoint. Moreover,

$$\begin{aligned} \{(u, t) \in \mathcal{A} \ u_j &= -\tau_j \ for \ some \ j = 1, \ldots, n\} \\ &= \{(u, t) \in \partial\mathcal{A} \ (u, t + \alpha) \in \partial\mathcal{A} \ \forall \alpha \geq 0\}. \end{aligned}$$

Proof Observe that the union in (41) is clearly disjoint. We first prove (41).

($\supseteq$) None of the sets on the RHS of (41) is contained in $\mathring{\mathcal{A}}$. In fact,

- the definition of $g(u)$ implies that for all $\varepsilon > 0$, $-\varepsilon < t < \varepsilon$, $(u, t) \in \mathcal{A}$ if and only if $t \geq 0$, so that $(u, g(u)) \notin \text{a-int}(\mathcal{A}) = \mathring{\mathcal{A}}$. This proves that the graph of g in $\prod_j(-\tau_j, +\infty)$ is a subset of $\partial\mathcal{A}$.
- Analogously, assume that $u_j = -\tau_j$ for some $j = 1, \ldots, n$, and for all $\varepsilon > 0$ consider the point (u_ε, t), where $(u_\varepsilon)_l = u_l$ for all $l \neq j$ and $(u_\varepsilon)_j = -\tau_j - \varepsilon$. But g is defined on $\prod_j[-\tau_j, +\infty)$ and $\mathcal{A}$ is its epigraph, hence all the points of $\mathcal{A}$ must be in the form $(u, g(u) + \alpha)$ for some $u \in \prod_j[-\tau_j, +\infty)$, $t = g(u) + \alpha$ ($\alpha \geq 0$), hence $(u_\varepsilon, t) \notin \mathcal{A}$ and this proves that $(u, t) \notin \text{a-int}(\mathcal{A})$.

The fact that $\partial\mathcal{A} = \mathcal{A} \setminus \mathring{\mathcal{A}}$ proves the first inclusion.

($\subseteq$) We prove that the complementary of the RHS of (41) in $\mathbb{R}^{n+1}$ is contained in $\mathring{A}$. Let (u, t) be such that $u > -\tau_j$ for all j and $t > g(u)$ (as it is easy to check, these are the conditions for (u, t) to belong to the complementary of the union of the two set at the LHS of (41)).

Let $d := t - g(u) > 0$. Since g is continuous on $\prod_j(-\tau_j, +\infty)$, there exists $\delta > 0$ such that $|g(u) - g(v)| < d/4$ for all $v \in B_\delta(u) := \{w \in \mathbb{R}^n : |w - u| < \delta\}$. In particular, for all $v \in B_\delta(u)$, $g(v) < t - \frac{3}{4}d < t$. Then, $B_\delta(u) \times (t - \frac{3}{4}d, +\infty)$ is all contained in $\mathcal{A}$ (because $\mathcal{A}$ is the epigraph of g) and it is an open neighborhood of (u, t). Hence, $(u, t) \in \mathring{A} = \mathcal{A} \setminus \partial\mathcal{A}$.

Next, we check the second part of the lemma:

($\subseteq$) assume $(u, t) \in \mathcal{A}$ is such that $u_j = -\tau_j$ for some j. Then, by the first part of this Lemma, $(u, t + \alpha) \in \partial\mathcal{A}$ for all $\alpha \geq 0$, since (41) is a partition of $\partial\mathcal{A}$.

($\supseteq$) Assume that $(u, t + \alpha) \in \partial\mathcal{A}$ for all $\alpha \geq 0$. Then, $(u, t) \in \partial\mathcal{A}$. Assume by contradiction that $u_j > -\tau_j$ for all j. Then, since (41) is a partition of $\partial\mathcal{A}$, $g(u) = t + \alpha$ for all $\alpha \geq 0$, which cannot be the case.

$\square$

The function g, defined in Lemma 3.6, can be expressed in terms of the function h_G of Lemma 3.4, as shown in the following result.

Theorem 3.10 *Let $\mathcal{A}$ be the set of points satisfying* (38), *h_G and g be the functions defined in* (39) *and* (40), *respectively. For $u \in \prod_{j=1}^n [-\tau_j, +\infty)$, $u = (u_1, \ldots, u_n)$, let $Q(u) := \prod_{j=1}^n [-\tau_j, u_j]$ and*

$$h(u) := \min_{S \in \mathcal{M}_n,\, v \in Q(u)} \|AS(v + \tau) - b\|_2^2 = \min_{v \in Q(u)} h_G(v). \tag{42}$$

Then, $h(u) = g(u)$ for all $u \in \prod_j [-\tau_j, +\infty)$.

Proof We first prove that $g(u) \leq h(u)$. For, it is enough to prove that $(u, h(u)) \in \mathcal{A}$, so that $g(u) \leq h(u)$ would follow by the definition of g. By definition of h, there exist $S_0 \in \mathcal{M}_n$ and $v \in Q(u)$ so that:

$$h(u) = \|AS_0 v + AS_0 \tau - b\|_2^2.$$

By Lemma 3.3 (iii), $(v, h(u)) \in \mathcal{G}$. Since $u_j \geq v_j$ for all $j = 1, \ldots, n$, it follows that $(u, h(u)) \in \mathcal{A}$ by definition of $\mathcal{A}$.

For the converse, since $(u, g(u)) \in \mathcal{A}$, there exists $(v', t) \in \mathcal{G}$ such that $v'_j \leq u_j$ for all $j = 1, \ldots, n$ and $g(u) \geq t$. In particular, $v' \in Q(u)$. By Lemma 3.3 (iii), $t = \|AS_1 v' + AS\tau - b\|_2^2$ for some $S_1 \in \mathcal{M}_n$. Therefore,

$$g(u) \geq \|AS_1 v' + AS_1 \tau - b\|_2^2 \geq \min_{S \in \mathcal{M}_n,\, v \in Q(u)} \|ASv + AS\tau - b\|_2^2 = h(u).$$

This concludes the proof. $\square$

Even if $g = h$, in what follows we still distinguish h and g when we want to stress the explicit definitions of both. Namely, we write $g(u)$ when we refer to $\min_{(u,s)\in\mathcal{A}} s$ and $h(u)$ when we refer to (42).

Corollary 3.11 *Under the same notation as above,*

$$g(u) = \min_{-u-\tau \preceq v \preceq u+\tau} \|Av - b\|_2^2. \tag{43}$$

Proof Using the second expression in (42),

$$g(u) = \min_{S\in\mathcal{M}_n} \min_{v\in Q(u)} \|AS(v + \tau) - b\|_2^2.$$

But,

$$f_S(v) = \|AS(v + \tau) - b\|_2^2 = f(S(v + \tau)),$$

for $f(v) = \|Av - b\|_2^2$, that gives:

$$\min_{-\tau \preceq v \preceq u} f_S(v) = \min_{v\in Q(u)} f(S(v + \tau)) = \min_{v\in S(Q(u)+\tau)} \|Av - b\|_2^2,$$

so that:

$$\min_{S\in\mathcal{M}_n} \min_{-\tau \preceq v \preceq u} f_S(v) = \min_{\bigcup_{S\in\mathcal{M}_n} S(Q(u)+\tau)} \|Av - b\|_2^2$$

and the assertion follows by observing that

$$\bigcup_{S\in\mathcal{M}_n} S(Q(u) + \tau) = \{v \in \mathbb{R}^n : -u - \tau \preceq v \preceq u + \tau\}.$$

$\square$

3.3 A Result Under Conditions on the Gradient of $\|Ax - b\|_2^2$

In general, the geometry of $\mathcal{A}$ is so complicated that expressing g explicitly may turn into a tough task. Nevertheless, it is obvious that if u is itself one of the minimizers of (42), then $g(u) = h_G(u) = \min_{S\in\mathcal{M}_n} \|ASu + AS\tau - b\|_2^2$. So, under further assumptions on $\nabla(\|ASu - b\|_2^2)$ granting the equality $g(u) = h_G(u)$ holds in a neighborhood of 0, we can compute explicitly the Lagrange multipliers.

Theorem 3.12 *Let* $f(v) = \|Av - b\|_2^2$ *and assume that for all* $k = 1, \ldots, n$ *the condition:*

$$\sum_{j=1}^{n} u_j \langle a_{*,j}, a_{*,k} \rangle \leq \langle b, a_{*,k} \rangle \qquad (-\tau \preceq u \preceq \tau) \tag{44}$$

holds. Then, $g(u) = f(u + \tau)$ for all $u \in Q(0)$ and $\lambda^{\#} = A^T(b - A\tau)$ is a set of Lagrange multipliers for problem (28).

Proof The set of conditions (44) is equivalent to $(Au - b)^T A \preceq 0$ for all $-\tau \preceq u \preceq \tau$, that is $\nabla f(u) \preceq 0$ for $-\tau \preceq u \preceq \tau$. We prove that, under this further condition, $g(u) = f(u + \tau)$ for all $u \in Q(0)$. Let $u \in Q(0)$ and $\mathfrak{n} \succ 0$ be a unit vector. For all $t \in \mathbb{R}$, define:

$$f_{\mathfrak{n}}(t) := f(u + \tau + t\mathfrak{n}) = \|A(u + \tau + t\mathfrak{n}) - b\|_2^2$$
$$= \|A\mathfrak{n}\|_2^2 t^2 + 2\langle A(u + \tau) - b, A\mathfrak{n}\rangle t + \|A(u + \tau) - b\|_2^2,$$

which is the restriction of f to the line $\{u + t\mathfrak{n} : t \in \mathbb{R}\}$. If $\mathfrak{n} \in \ker(A)$, then $f_{\mathfrak{n}} \equiv 0$ and it has a global minimum in $t = 0$. Assume $\mathfrak{n} \notin \ker(A)$. The intersection of this line with $\{-\tau \preceq v \preceq \tau\}$ is contained in $(-\infty, 0]$. If we prove that, for all $\mathfrak{n} \succ 0$, $f_{\mathfrak{n}}$ has a constrained minimum in $t = 0$, we get the first assertion. For, it's enough to observe that

$$f_{\mathfrak{n}}'(0) = \nabla f(u + \tau) \cdot \mathfrak{n} \le 0,$$

because if $u \in Q(0)$, then $\{-u - \tau \preceq v \preceq u + \tau\} \subseteq \{-\tau \preceq v \preceq \tau\}$. This proves that $g(u) = f(u + \tau)$ for all $u \in Q(0)$. In particular,

$$-\nabla g(0) = -\nabla f(\tau) = (b - A\tau)^T A \succeq 0$$

is a set of Lagrange multipliers for (28).   □

Remark 3.13 It is not difficult to generalize Theorem 3.12 a bit further. If the hyperparallelogram $\{-\tau \preceq u \preceq \tau\}$ is all contained in the region $\{u \in \mathbb{R}^n : S\nabla f(u) \preceq 0\}$ for some $S \in \mathcal{M}_n$, then $g(u) = f(S(u + \tau))$ for all $u \in Q(0)$ and

$$\lambda^{\#} = -\nabla g(0) = -S\nabla f(S(u + \tau))^T$$

defines a vector of Lagrange multipliers for (28). The proof goes exactly as in Theorem 3.12.

3.4 Decoupling the Variables

In this subsection, we focus on the situation in which $A^T A$ is a diagonal matrix. Since:

$$A^T A = \begin{pmatrix} \|a_{*,1}\|_2^2 & \langle a_{*,1}, a_{*,2}\rangle & \dots & \langle a_{*,1}, a_{*,n}\rangle \\ \langle a_{*,2}, a_{*,1}\rangle & \|a_{*,2}\|_2^2 & \dots & \langle a_{*,2}, a_{*,n}\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle a_{*,n}, a_{*,1}\rangle & \langle a_{*,n}, a_{*,2}\rangle & \dots & \|a_{*,n}\|_2^2 \end{pmatrix}$$

and the rank of $A^T A$ is equal to that of A, it follows that in this case:

$$A^T A = \text{diag}(\|a_{*,1}\|_2^2, ..., \|a_{*,n}\|_2^2). \tag{45}$$

Remark 3.14 If $m \leq n$ and $A^T A$ is diagonal, $n - m$ of the norms in (45) above vanish. In this case, we assume that $a_{*,m+1} = ... = a_{*,n} = 0$, so that A can be written in terms of its columns as:

$$A = \left(A' | 0_{m \times (n-m)}\right),$$

where $A' = (a_{*,1}|...|a_{*,m}) \in GL(m, \mathbb{R})$. Observe that:

$$\|Ax - b\|_2^2 = \|A'x' - b\|_2^2,$$

where $x' = (x_1, ..., x_m)^T$, so that $x^\#$ is a mimizer of (28) if and only if $(x^\#)' = (x_1^\#, ..., x_m^\#)$ is a minimizer of the problem:

$$\text{minimize} \quad \|A'y - b\|_2^2, y \in \mathbb{R}^m, \ |y_j| \leq \tau_j, \ j = 1, ..., m, \tag{46}$$

under the further condition that the remaining coordinates of x vanish.

For this reason, for the rest of this subsection, we focus on (46), both for the cases $n \leq m$ and $m \leq n$, and provide the Lagrange multipliers.

Remark 3.15 We point out that in this situation the Lagrange multipliers can be computed directly from Proposition 3.1. Indeed, under the *orthogonality* assumption on A, the target function in problem (46) becomes:

$$\sum_{j=1}^{m}(\|a_{*,j}\|_2^2 y_j^2 - 2y\langle a_{*,j}, n \rangle y_j) + \|b\|_2^2.$$

Since the variables of all the addenda are decoupled, and the addenda are non-negative,

$$\min_{y} \sum_{j=1}^{m}(\|a_{*,j}\|_2^2 y_j^2 - 2y\langle a_{*,j}, n \rangle y_j) + \|b\|_2^2$$

$$= \sum_{j=1}^{m} \min_{y_j} \left(\|a_{*,j}\|_2^2 y_j^2 - 2y\langle a_{*,j}, n \rangle y_j + \frac{\|b\|_2^2}{m}\right)$$

and a minimizer of (46) is also a minimizer of the problem:

$$\text{minimize} \quad \|a_{*,j}\|_2^2 y_j^2 - 2y\langle a_{*,j}, n \rangle y_j + \frac{\|b\|_2^2}{m}, \qquad |y_j| \leq \tau_j$$

for all $j = 1, ..., m$. In other words, it is enough to treat (46) as m 1-dimensional constrained minimization problems. However, our interest is testing the tools presented in the previous section, computing the function g and the separating hyperplane.

To exhibit a vector of Lagrange multipliers, we start by the set

$$\mathcal{G} := \{(u, t) \in \mathbb{R}^{m+1} \mid u_j = |y_j| - \tau_j \ (j = 1, \ldots, m),$$
$$t = \left\| A'y - b \right\|_2^2 \text{ for some } y \in \mathbb{R}^m\}.$$

By Lemma 3.3 (iii), $(u, t) \in \mathcal{G}$ if and only if $u \succeq -\tau$ and $t = \left\| A'S(u + \tau) - b \right\|_2^2$ for some $S \in \mathcal{M}_m$. Let $f_S(u) = \left\| A'S(u + \tau) - b \right\|_2^2$ and observe that:

$$f_S(u) = \sum_{j=1}^{m} \left\| a_{*,j} \right\|_2^2 (u_j + \tau_j)^2 - 2 \sum_{j=1}^{m} s_{jj} \langle b, a_{*,j} \rangle (u_j + \tau_j) + \|b\|_2^2.$$

The functions f_S are the equivalent of the parabolas in the 1-dimensional case and they describe elliptic paraboloids. As it clear by Sect. 3.1, we need to understand what is $h_G(u) := \min_{S \in \mathcal{M}_m} f_S(u)$. Observe that for all $S \in \mathcal{M}_n$,

$$f_S(u) \geq \sum_{j=1}^{m} \left\| a_{*,j} \right\|_2^2 (u_j + \tau_j)^2 - 2 \sum_{j=1}^{m} |\langle b, a_{*,j} \rangle| (u_j + \tau_j) + \|b\|_2^2 = f_{S_\beta}(u), \quad (47)$$

where $S_\beta = (s_j^\beta)_{j=1}^m \in \mathcal{M}_m$ is a diagonal matrix such that $s_j^\beta \langle b, a_{*,j} \rangle \geq 0$.

Lemma 3.16 *Under the notation and the assumptions of this subsection,*

$$h_G(u) = f_{S_\beta}(u) = \sum_{j=1}^{m} \left\| a_{*,j} \right\|_2^2 (u_j + \tau_j)^2 - 2 \sum_{j=1}^{m} |\langle b, a_{*,j} \rangle| (u_j + \tau_j) + \|b\|_2^2.$$

h_G defines an elliptic paraboloid whose vertex $V = (c, 0) \in \mathbb{R}^{m+1}$ is characterized both by $c = -\tau + S_\beta (A')^{-1} b$ and

$$c_j = -\tau_j + \frac{|\langle b, a_{*,j} \rangle|}{\left\| a_{*,j} \right\|_2^2}$$

$(j = 1, \ldots, m)$. Moreover,

$$h_G(u) = \sum_{j=1}^{m} \left\| a_{*,j} \right\|_2^2 (u_j - c_j)^2. \quad (48)$$

Proof We already proved the first part of the Lemma. We only need to compute the vertex of f_{S_β}. For, observe that the minimum of f_{S_β} is $(c, 0)$, where c satisfies $f_S(c) =$

0. This equation is satisfied if and only if $c = -\tau + S_\beta(A')^{-1}b$. Moreover, the minimum of f_{S_β} is also characterized by $\nabla f_{S_\beta}(c) = 0$, that is:

$$c_j + \tau_j - \frac{|\langle b, a_{*,j}\rangle|}{\|a_{*,j}\|_2^2} = 0$$

($j = 1, \ldots, m$). Finally, using the first characterization of c,

$$h_G(u) = \|AS_\beta(u + \tau) - b\|_2^2 = \|AS_\beta(u - c) + AS_\beta(c + \tau) - b\|_2^2$$
$$= \|AS_\beta(u - c)\|_2^2 =$$
$$= \sum_{j=1}^m \|a_{*,j}\|_2^2 (u_j - c_j)^2.$$

This concludes the proof.   □

In order to compute the Lagrange multipliers for the decoupled problem, we observe that $\mathcal{A} + [0, +\infty)^{m+1}$ is the epigraph of the function $g(u)$ whose first properties are proved in Lemma 3.6. Hence, this function describes the lower boundary of $\mathcal{A}$, that is the part of $\mathcal{A}$ we need to compute a separating hyperplane. By (42), $g(u) = \min_{v \in Q(u)} h_G(v)$, where $Q(u) = \prod_{j=1}^m [-\tau_j, u_j]$.

Theorem 3.17 *Under the notation and the assumptions of this subsection,*

$$g(u) = h_G(Pu),$$

where $P : \prod_{j=1}^m [-\tau_j, +\infty) \to Q(c)$ *is the projection defined for all* $u \in \prod_{j=1}^m [-\tau_j, +\infty)$ *by*

$$(Pu)_j = \begin{cases} u_j & \text{if } -\tau_j \leq u_j \leq c_j, \\ c_j & \text{if } u_j > c_j \end{cases} = \min\{c_j, u_j\} \tag{49}$$

($j = 1, \ldots, m$). *Explicitly, under the assumptions of this subsection,*

$$g(u) = \sum_{j=1}^m \|a_{*,j}\|_2^2 (u_j - c_j)^2 \chi_{[-\tau_j, c_j]}(u_j). \tag{50}$$

In particular, $g \in \mathcal{C}^1(\prod_{j=1}^n (-\tau_j, +\infty))$ *with:*

$$\frac{\partial g}{\partial u_j}(u) = 2\|a_{*,j}\|_2^2 (u_j - c_j)\chi_{[-\tau_j, c_j]}(u_j) \tag{51}$$

for all $u \in \prod_{j=1}^n (-\tau_j, +\infty)$.

Proof Obviously, P is a projection of $\prod_{j=1}^{n}[-\tau_j, +\infty)$ onto $Q(c)$. For all $j = 1, \ldots, m$,

$$\text{argmin}_{-\tau_j \leq v_j \leq u_j} \left\| a_{*,j} \right\|_2^2 (v_j - c_j)^2 = \begin{cases} u_j & \text{if } -\tau_j \leq u_j \leq c_j, \\ c_j & \text{otherwise} \end{cases} = (Pu)_j.$$

Hence,

$$g(u) = \min_{v \in Q(u)} h_G(v) = \sum_{j=1}^{m} \min_{-\tau_j \leq v_j \leq u_j} \left\| a_{*,j} \right\|_2^2 (v_j - c_j)^2$$

$$= \sum_{j=1}^{m} \left\| a_{*,j} \right\|_2^2 ((Pu)_j - c_j)^2 = h_G(Pu).$$

The explicit definition of Pu gives (50) and (51). The differentiability and formula (51) are obvious by the expression (50) of g. □

Remark 3.18 As a consequence of Theorem 3.17,

$$p^* = g(0) = \sum_{j=1}^{m} \left\| a_{*,j} \right\|_2^2 \left(-\tau_j + \frac{|\langle b, a_{*,j} \rangle|}{\left\| a_{*,j} \right\|_2^2} \right)^2 \chi_{[-\tau_j, c_j]}(0).$$

Then, observe that:

$$-\tau_j \leq 0 \leq -\tau_j + \frac{|\langle b, a_{*,j} \rangle|}{\left\| a_{*,j} \right\|_2^2} \iff 0 \leq \tau_j \leq \frac{|\langle b, a_{*,j} \rangle|}{\left\| a_{*,j} \right\|_2^2}, \qquad (52)$$

so that:

$$p^* = \sum_{j=1}^{m} \left\| a_{*,j} \right\|_2^2 \left(-\tau_j + \frac{|\langle b, a_{*,j} \rangle|}{\left\| a_{*,j} \right\|_2^2} \right)^2 \chi_{\left[0, \frac{|\langle b, a_{*,j} \rangle|}{\left\| a_{*,j} \right\|_2^2} \right]}(\tau_j).$$

Theorem 3.19 *Under the notation of this subsection, the vector* $\lambda^{\#} \in [0, +\infty)^m$ *given by*

$$\lambda_j^{\#} = 2 \left\| a_{*,j} \right\|_2^2 \left(\frac{|\langle b, a_{*,j} \rangle|}{\left\| a_{*,j} \right\|_2^2} - \tau_j \right)^+$$

defines a vector of Lagrange multipliers for (46).

Proof We apply (51) to $u = 0$ and use (52). Namely,

$$t = p^* + \langle \nabla g(0), u \rangle$$

is the tangent hyperplane of g in $u = 0$, which is also the hyperplane that separates $\mathcal{A}$ and $\mathcal{B}$. The direction of this hyperplane is $(\nabla g(0), -1)$, so that:

$$\lambda^{\#} = -\nabla g(0),$$

i.e. the assertion. $\square$

Remark 3.20 As far as the original problem (28) with $m \leq n$ is concerned, we get the Lagrange multipliers for free by Theorem 3.19 simply observing that if $A = (a_{*1}|\ldots|a_{*m}|0|\ldots|0) \in \mathbb{R}^{m \times n}$, $A' = (a_{*1}|\ldots|a_{*m})$ and $x = (x', x'') \in \mathbb{R}^m \times \mathbb{R}^{n-m}$, then

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 + \sum_{j=1}^{n} \lambda_j(|x_j| - \tau_j) = \min_{x' \in \mathbb{R}^m} \|A'x' - b\|_2^2 + \sum_{j=1}^{m} \lambda_j(|x_j'| - \tau_j) +$$

$$+ \underbrace{\min_{x'' \in \mathbb{R}^{n-m}} \sum_{j=m+1}^{n} \lambda_j(|x_j| - \tau_j)}_{=} - \sum_{j=m+1}^{n} \lambda_j \tau_j,$$

so that, if $\lambda^{\#} \in \mathbb{R}^m$ defines a vector of Lagrange multipliers for (46), then $(\lambda^{\#}|0) \in \mathbb{R}^m \times \mathbb{R}^{n-m}$ defines a vector of Lagrange multipliers for (28).

3.5 Explicit Solution

The conditions $|x_j| \leq \tau_j$ are equivalent to $x_j^2 \leq \tau_j^2$. Under this point of view, (28) can be restated as:

$$\text{minimize} \quad \|Ax - b\|_2^2, \quad x_j^2 \leq \tau_j^2, \tag{53}$$

that can be interpreted as a weighted Tikhonov problem. Assume that $\lambda^{\#}$ is a vector of Lagrange multipliers for (28) or, equivalently, for (53). We are interested in computing

$$x^{\#} = \arg\min_x L(x, \lambda),$$

where L is the Lagrange function associated to (53), i.e.

$$L(x, \lambda^{\#}) = \|Ax - b\|_2^2 + \sum_{j=1}^{n} \lambda_j^{\#}(x_j^2 - \tau_j^2).$$

Since $L \in \mathcal{C}^{\infty}(\mathbb{R}^n)$ and it is convex, they satisfy $\nabla L(x, \lambda^{\#}) = 0$, that is:

$$(A^T A + \Delta_\lambda)x = A^T b,$$

where $\Delta_\lambda = \text{diag}(\lambda_1^{\#}, \ldots, \lambda_n^{\#})$. Hence, $x^{\#}$ satisfies:

$$(A^T A + \Delta_\lambda)x^{\#} = A^T b, \tag{54}$$

that is, $x^\# \in (A^T A + \Delta_\lambda)^{-1} A^T b$.

Remark 3.21 Another way to compute the Lagrange multipliers associated to (28), or equivalently to (53), can be by means of strong duality condition, namely using:

$$\lambda^\# = \arg \max_{\lambda \succeq 0} \min_x L(x, \lambda).$$

However, we stress that the explicit value of $\min_x L(x, \lambda)$ is still hard to compute since the implicit relation (54) satisfied by $x^\#$ cannot be made explicit by means of Dini's theorem.

4 Considerations and Conclusions

4.1 Applications

Despite the apparently heavy assumptions on A, Theorem 3.19 has itself interesting applications. For instance, it can be applied to denoising problems, where $A = I_{n \times n}$, i.e. problems in the form:

$$\text{minimize} \quad \|x - b\|_2^2, \qquad x \in \mathbb{R}^n, \ |x_j| \leq \tau_j, \ j = 1, \ldots, n. \tag{55}$$

By Theorem 3.19, $\lambda^\# = (\lambda_j^\#)_{j=1}^n$ is a vector of Lagrange multipliers for (55), where:

$$\lambda_j^\# = 2(|b_j| - \tau_j)^+. \tag{56}$$

We can also apply Theorem 3.19 to the discrete Fourier transform, i.e. given a noisy fully-sampled signal $b \in \mathbb{C}^n$, we want to find a vector $z \in \mathbb{C}^n$ such that $\|\Phi z - b\|_2^2$ is minimized under the constrains $|z_j| \leq \tau_j$, where $\Phi \in \mathbb{C}^{n \times n}$ denotes the (complex) DFT matrix. Since $\Phi^* \Phi = I_{n \times n}$, we can apply Theorem 3.19 to deduce that a set of Lagrange multipliers for this problem is:

$$\lambda_j^\# = 2 \left(|\langle b, \phi_{*,j} \rangle| - \tau_j \right)^+,$$

($j = 1, \ldots, n$), being $\phi_{*,j}$ the j-th column of Φ.

The question that naturally arises in the applications is whether the dependence of $\lambda_1, \ldots, \lambda_n$ on $\tau_1, \ldots, \tau_n$ can be a critical issue in the applicability of the theory. Indeed, $\tau_1, \ldots, \tau_n$ are upper bounds for $|x_1|, \ldots, |x_n|$ respectively, which are not available in the practice. However, whenever it is possible to estimate these local upper bounds, our result may lead to high-quality imaging perfomances. For instance, for denoising, (56) may be approximated by replacing $\tau_1, \ldots, \tau_n$ with the voxel values obtained by applying a Gaussian filter (or other types of filtering) to the noisy image. This opens the question of which filtering technique could lead to optimal approximations of the $\tau_1, \ldots, \tau_n$ depending on the field of research in which (28) can be implemented. We intend to investigate this topic in the immediate future.

4.2 Open Problems

As long as $A^T A$ is not a diagonal matrix, the geometries of the sets $\mathcal{G}$ and $\mathcal{A}$ of the points satisfying (37) and (38) respectively, become more involved, along with the possible casuistry. However, the general case in which $A^T A$ is not diagonal would be of great importance in applications. Indeed, as long as Lagrange multipliers are proved to act as effective tuning parameters, the behavior of Lagrange multipliers for the weighted LASSO problem (13) in terms of voxel-wise estimates would provide a way to control the tuning parameters via estimates of the τ_j.

Another open problem is whether it is possible to apply the same procedure to compute the Lagrange multipliers for (3). Clearly, the corresponding sets $\mathcal{G}$ and $\mathcal{A}$ lie in $\mathbb{R}^2$ so that $g : \mathbb{R} \to \mathbb{R}$. Despite this simplifying fact, the set $\mathcal{G}$ is characterized by:

$$\begin{cases} u = s(x)^T x - \tau, \\ t = \|Ax - b\|_2^2 \end{cases} \qquad \text{for some } x \in \mathbb{R}^n,$$

where $s(x) \in \mathbb{R}^n$ is a vector such that $\text{diag}(s(x)_j) \in \text{sgn}(x)$ and, in this case, u and x belong to different spaces and a closed form for $t = t(u)$ is even more difficult to provide.

The possibility of using Lagrange multipliers as tuning parameters in disciplines that apply LASSO problems, such as MRI, is still open. Lagrange multipliers for (6), however, depend on its constraint:

$$\|Ax - b\|_2 \leq \eta.$$

Consequently, even if the utilization of Lagrange multipliers as tuning parameters were feasible in applications, unless a method for accurately estimating η is provided, the focus would simply shift from Lagrange multipliers to estimating the ℓ_2 norm of the noise. It is therefore crucial to determine whether slight perturbations in these estimates lead to significant variations, for instance, of the quality of retrieved images in MRI.

Finally, we stress that it would be important to generalize (28) up to consider different inner products on $\mathbb{R}^n$. Namely, this is the situation that occurs in MRI when the undersampling pattern is non-cartesian. Problem (13) in this case becomes:

$$\min_x \|Ax - b\|_W^2 + \sum_j \lambda_j |x_j|,$$

where

$$\|x\|_W = x^T W^T W x \qquad (x \in \mathbb{R}^n),$$

for a definite positive diagonal matrix W. Since this topic falls beyond the purpose of this work, we limit ourselves to mention the very mathematical reason why the weighted norm shall definitely replace the Euclidean norm over $\mathbb{R}^n$ when sampling

is not performed on a cartesian grid. Indeed, non-cartesian sampling patterns require appropriate discretizations of the Fourier transform integral. Roughly speaking,

$$\hat{f}(\xi) \approx \sum_j f(x_j) e^{-2\pi i \xi \cdot x_j} \Delta x_j = \langle f, e^{2\pi i \xi \cdot} \rangle_W,$$

where Δx_j is the Lebesgue measure of an adequate neighborhood of x_j, weighting the contribution of the sample x_j, and W is the diagonal matrix whose entries are $\sqrt{\Delta x_j}$. The inversion formula of the Fourier transform shall be modified accordingly. For instance, if the sampling follows a spiral trajectory, Δx_j shall be bigger the further x_j is from the origin, since this value serves as an avarage of f on a portion of sphere that is larger as x_j is far from the origin. All the above-mentioned problems will be object of our future investigations.

Declarations

Conflict of interest The authors have not disclosed any competing interests.

Appendix

Since we did not find a direct proof in the existing literature, we provide a formal proof of the existence of the minimizer of the generalized LASSO problem:

$$\arg \min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 + \lambda \|\Phi x\|_1, \tag{57}$$

where $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$ and $\Phi \in \mathbb{R}^{N \times n}$.

We state the result in the general framework of finite-dimensional vector spaces: we denote by X, Y, Z three finite-dimension real vector spaces. We denote by $\langle \cdot, \cdot \rangle_X$ an inner product on X and with $\|\cdot\|_X$ the induced norm. Analogous notation will be used for Y, whereas we set:

$$\|z\|_p = \left(\sum_{j=1}^{\dim(Z)} |z_j|^p \right)^{1/p}, \qquad z \in Z$$

for $0 < p \leq \infty$. Recall that $\|\cdot\|_p$ is a Banach quasi-norm (meaning that there exists $C_p \geq 1$ such that $\|x + y\|_p \leq C_p(\|x\|_p + \|y\|_p)$ for all $x, y \in Z$) replaces the triangular inequality for $0 < p < 1$, and it is a norm for $1 \leq p \leq \infty$.

Then, for given $\lambda \geq 0$, $b \in Y$, $A : X \to Y$ and $\Phi : X \to Z$ linear, we define for all $x \in X$,

$$f(x) = \|Ax - b\|_Y^2 + \lambda \|\Phi x\|_p.$$

Theorem *For all $0 < p \leq \infty$, $\lambda \geq 0$ there exists $x^\# \in X$ such that*

$$\inf_{x \in X} f(x) = f(x^\#).$$

In particular, the generalized LASSO problem (57) *has at least one solution.*

Proof Clearly, the function $f(x) = \|b\|_Y^2 + \lambda \|\Phi x\|_p$ attains its minimum in $x^\# = 0$. Hence, we may assume that

$$\mathrm{Im}(A) \neq \{0\}.$$

Since $Image(A)$ is a vector subspace of Y, for all $b \in Y$ there exists a unique $y^\# \in Image(A)$ such that $\inf_{y \in Y} \|y - b\|_Y^2 = \|y^\# - b\|_Y^2$. By definition, $y^\# = Ax^\#$ for some $x^\# \in X$. Hence,

$$\inf_{x \in X} \|Ax - b\|_Y^2 = \inf_{y \in Image(A)} \|y - b\|_Y^2 = \min_{y \in Y} \|y - b\|_2^2 = \left\| Ax^\# - b \right\|_Y^2$$

and the assertion follows also for the case in which $\lambda = 0$ or $Image(B) = \{0\}$. We will thereby assume that $\mathrm{Im}(A) \neq \{0\}$, $\mathrm{Im}(B) \neq \{0\}$ and $\lambda > 0$. Let $L := \ker(A) \cap \ker(B) = \{x \in X : Ax = 0, \Phi x = 0\}$ and denote the closed ball of X of center 0 and radius $r > 0$ by $B_X(0, r) = \{x \in X : \|x\|_X \leq r\}$. The rest of the proof is devided into three graded steps.

Step 1. We prove that if $L = \{0\}$, then $\lim_{\|x\|_X \to +\infty} f(x) = +\infty$.

By convexity of $\|\cdot\|_Y$,

$$\|y_1\|_Y^2 \leq 2(\|y_1 - y_2\|_Y^2 + \|y_2\|_Y^2)$$

for all $y_1, y_2 \in Y$. Therefore,

$$\|Ax - b\|_2^2 \geq \frac{1}{2} \|Ax\|_Y^2 - \|b\|_Y^2,$$

so that:

$$\|Ax - b\|_Y^2 + \lambda \|\Phi x\|_p \geq \frac{1}{2} \|Ax\|_Y^2 + \lambda \|\Phi x\|_p - \|b\|_Y^2 .$$

Let

$$\$_X := \{x \in X : \|x\|_X = 1\}$$

denote the unit sphere of X and set $\eta := \min_{x \in \$_X} \frac{1}{2} \|Ax\|_Y^2 + \lambda \|\Phi x\|_p$. If $\eta = 0$, then,

$$\min_{x \in \$_X} \frac{1}{2} \|Ax\|_Y^2 + \lambda \|\Phi x\|_p = 0,$$

together with the assumptions on $\mathrm{Im}(A)$, $\mathrm{Im}(B)$ and λ, yields to the existence of $x^\# \in \$_X$ such that $\frac{1}{2} \left\|Ax^\#\right\|_Y^2 + \lambda \left\|\Phi x^\#\right\|_p = 0$. But $\|\cdot\|_Y$ and $\|\cdot\|_p$ are (quasi-)norms, so $x^\# = 0 \notin \$_X$. This is a contradiction. Hence, $\eta > 0$.

Next, for $\|x\|_X > 1$, we have:

$$\frac{1}{2} \|Ax\|_Y^2 + \lambda \|\phi x\|_p - \|b\|_Y^2 = \frac{1}{2} \|x\|_X^2 \left\|A \frac{x}{\|x\|_X}\right\|_Y^2 + \lambda \|x\|_X \left\|\Phi \frac{x}{\|x\|_X}\right\|_p - \|b\|_Y^2$$

$$> \|x\|_X \left(\frac{1}{2} \left\|A \frac{x}{\|x\|_X}\right\|_Y^2 + \lambda \left\|B \frac{x}{\|x\|_X}\right\|_p \right) - \|b\|_Y^2$$

$$\geq \eta \|x\|_X - \|b\|_Y^2 .$$

Therefore, for all $x \in X$ such that $\|x\|_X > 1$,

$$f(x) = \|Ax - b\|_Y^2 + \lambda \|\Phi x\|_p > \eta \|x\|_X + \|b\|_Y^2$$

and the assertion follows, since $\eta > 0$ implies that the right hand-side goes to $+\infty$ as $\|x\|_X \to +\infty$.

Step 2. We prove the assertion for $L = \{0\}$.

Let $m := \inf_{x \in X} f(x)$. By Step 1, there exist $R > 0$ such that $f(x) > m + 1$ for $\|x\|_X > R$. $B_X(0, R)$ is compact and convex, and $\inf_{x \in X} f(x) = \inf_{x \in B_X(0,R)} f(x)$ by definition of R. Let $(x_j)_j \subseteq B_X(0, R)$ be a minimizing sequence. By compactness, it admits a converging subsequence and, without loss of generality, we may assume that $\lim_{j \to +\infty} x_j = x^\# \in B_X(0, R)$. By continuity, $f(x^\#) = \lim_{j \to +\infty} f(x_j) = m$.

Step 3. We prove the assertion for $L \neq \{0\}$.

Recall that $X = L \oplus L^\perp$, where the orthogonality is defined with respect to the inner product $\langle \cdot, \cdot \rangle_X$. By definition of direct sum, for all $x \in X$ there exist unique $x_1 \in L$ and $x_2 \in L^\perp$ such that $x = x_1 + x_2$. Observe that since $x_1 \in L$,

$$f(x) = \|Ax_2 - b\|_Y^2 + \lambda \|\Phi x_2\|_p = f(x_2).$$

In particular,

$$\inf_{x \in X} f(x) = \inf_{x \in L^\perp} f(x).$$

The restrictions of A and Φ to $L^\perp$ are linear mappings between vector spaces. We denote them with $A|_{L^\perp} L^\perp \to Y$ and $\Phi|_{L^\perp} : L^\perp \to Z$ respectively and set $f|_{L^\perp} : L^\perp \to Y$ as the restriction of f to $L^\perp$. Obviously,

$$f|_{L^\perp}(x) = \left\| A|_{L^\perp} x - b \right\|_Y^2 + \lambda \left\| \Phi|_{L^\perp} x \right\|_p = f(x)$$

for all $x \in L^\perp$, so that:

$$\inf_{x \in X} f(x) = \inf_{x \in L^\perp} f(x) = \inf_{x \in L^\perp} f|_{L^\perp}(x).$$

Obviously,

$$L^\perp := \ker(A|_{L^\perp}) \cap \ker(\Phi|_{L^\perp}) = \ker(A) \cap \ker(B) \cap L^\perp = L \cap L^\perp = \{0\}.$$

Therefore, by Step 2, it follows that there exists $x^\# \in L^\perp$ such that:

$$\inf_{x \in L^\perp} f|_{L^\perp}(x) = f|_{L^\perp}(x^\#).$$

This implies that:

$$\inf_{x \in X} f(x) = f|_{L^\perp}(x^\#) = f(x^\#),$$

since $x^\# \in L^\perp$. □

References

1. Adcock, B., Hansen, A.C.: Compressive Imaging: Structure, Sampling, Learning, p. 2021. Cambridge University Press, UK (2021)
2. Arridge, S., Maass, P., Öktem, O., Schönlieb, C.: Solving inverse problems using data-driven models. Acta Numer. **28**, 1–174 (2019). https://doi.org/10.1017/S0962492919000059
3. Berk, A., Brugiapaglia, S., Hoheisel, T.: LASSO reloaded: a variational analysis perspective with applications to compressed sensing. arXiv preprint arXiv:2205.06872
4. Berk, A., Plan, Y., Yilmaz, Ö.: Parameter instability regimes in sparse proximal denoising programs. In: 2019 13th International conference on Sampling Theory and Applications (SampTA) pp. 1-5. IEEE (2019, July)
5. Berk, A., Plan, Y., Yilmaz, Ö.: Sensitivity of ℓ_1 minimization to parameter choice. Inf. Inference: J. IMA **10**(2), 397–453 (2021)
6. Bertrand, Q., Klopfenstein, Q., Blondel, M., Vaiter, S., Gramfort, A., Salmon, J.: Implicit differentiation of Lasso-type models for hyperparameter optimization. International Conference on Machine Learning, pp. 810–821. PMLR (2020, November)
7. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, UK (2004)

8. Brady, D.J., Choi, K., Marks, D.L., Horisaki, R., Lim, S.: Compressive holography. Opt. Express **17**(15), 13040–13049 (2009)
9. Calatroni, L. et al.: Adaptive parameter selection for weighted-TV image reconstruction problems. J. Phys.: Conf. Ser. 1476 012003 (2020)
10. Calatroni, L., Lanza, A., Pragliola, M., Sgallari, F.: A flexible space-variant anisotropic regularization for image restoration with automated parameter selection. SIAM J. Imaging Sci. **12**(2), 1001–1037 (2019)
11. Calvetti, D., Morigi, S., Reichel, L., Sgallari, F.: Tikhonov regularization and the L-curve for large discrete ill-posed problems. J. Comput. Appl. Math. **123**(1–2), 423–446 (2000)
12. Candés, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly Incomplete Frequency Information. IEEE Trans. Inf. Theo (2006)
13. Candés, E.J.: The Restricted Isometry Property and its Implications for Compressed Sensing. C. R. Acad. Sci. Paris, Ser. I **346**, 589–592 (2008)
14. Chan, W.L.: Terahertz imaging with compressive sensing. Rice University, USA (2010)
15. Chen, S., Donoho, D.: Basis pursuit. Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers. (1994). https://doi.org/10.1109/ACSSC.1994.471413
16. Chen, S.S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. SIAM Rev. **43**(1), 129–159 (2001)
17. Correa, C.V., Arguello, H., Arce, G.R.: Compressive spectral imaging with colored-patterned detectors. In: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) pp. 7789–7793 (2014, May) IEEE
18. Deledalle, C., Vaiter, S., Fadili, J.M., Peyré, G.: Stein Unbiased GrAdient estimator of the Risk (SUGAR) for multiple parameter selection. SIAM J. Imaging Sci. Soc. Indust. Appl. Math. **7**(4), 2448–2487 (2014). (**hal-00987295v2**)
19. Di Sopra, L., Piccini, D., Coppo, S., Stuber, M., Yerly, J.: An automated approach to fully self-gated free-running cardiac and respiratory motion-resolved 5D whole-heart MRI. Magn. Reson. Med. **82**(6), 2118–2132 (2019)
20. Donoho, D.L., Elad, M., Temlyakov, V.N.: Stable Recovery of Sparse Overcomplete Representations in the Presence of Noise. IEEE Trans. Inf. Theo. 52(1), (2006)
21. Donoho, D.L.: Compressed Sensing. IEEE Trans. Inf. Theo. 52(4), (2006)
22. Feng, L.I., Coppo, S., Piccini, D., Yerly, J., Lim, R.P., Masci, P.G., Otazo, R.: 5D whole-heart sparse MRI. Magnet. Reson. Med. **79**(2), 826–838 (2018)
23. Foucart, S., Rauhut, H.: A Mathematical Introduction to Compressive Sensing. Springer, Birkhäuser (2013)
24. Franceschiello, B., Di Sopra, L., Minier, A., Ionta, S., Zeugin, D., Notter, M.P., Bastiaansen, J.A.M., Jorge, J., Yerly, J., Stuber, S., Murray, M.M.: 3-Dimensional magnetic resonance imaging of the freely moving human eye. Prog. Neurobiol. **194**, 101885 (2020)
25. Gao, L., Liang, J., Li, C., Wang, L.V.: Single-shot compressed ultrafast photography at one hundred billion frames per second. Nature **516**(7529), 74–77 (2014)
26. Golub, G.H., Heath, M., Wahba, G.: Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics **21**(2), 215–223 (1979)
27. Guerquin-Kern, M., Van De Ville, D., Vonesch, C., Baritaux, J.C., Pruessmann, K.P., Unser, M.: Wavelet-regularized reconstruction for rapid MRI. In: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro pp. 193–196. (2009, June). IEEE
28. Hasen, P.C.: The L-curve and its use in the numerical treatment of inverse problems. https://www.sintef.no/globalassets/project/evitameeting/2005/lcurve.pdf
29. He, C., Hu, C., Zhang, W., Shi, B.: A Fast Adaptive Parameter Estimation for Total Variation Image Restoration. IEEE Trans. Image Process. **23**(12), 4954–4967 (2014). https://doi.org/10.1109/TIP.2014.2360133
30. Hintermüller, M., Papafitsoros, K.: Generating structured nonsmooth priors and associated primal-dual methods. Handbook of numerical analysis (Vol. 20, pp. 437–502). Elsevier (2019)
31. Kim, Y.C., Narayanan, S.S., Nayak, K.S.: Accelerated three-dimensional upper airway MRI using compressed sensing. magnetic resonance in medicine: an official journal of the Int. Soc. Magnet. Reson. Med. **61**(6), 1434–1440 (2009)
32. Lanza, A., Pragliola, M., Sgallari, F.: Residual whiteness principle for parameter-free image restoration (2020)

33. Lian, L., Liu, A., Lau, V.K.N.: Weighted lasso for Sparse Recovery with Statistical Prior Support Information. IEEE Trans. Signal Process. **66**(6), 1607–1618 (2018)

34. Lin, F.H., Kwong, K.K., Belliveau, J.W., Wald, L.L.: Parallel imaging reconstruction using automatic regularization. Magnetic Resonance in Medicine: An Official Journal of the Int. Soc. Magnet. Reson. Med. **51**(3), 559–567 (2004)

35. Liu, X., Kang, J.U.: Compressive SD-OCT: the application of compressed sensing in spectral domain optical coherence tomography. Opt. Express **18**(21), 22010–22019 (2010)

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

37. Lustig, M., Donoho, D.L., Santos, J.M., Pauly, J.M.: Compressed sensing MRI. IEEE Signal Process. Mag. **25**(2), 72–82 (2008)

38. Mansour, H., Yilmaz, Ö: Weighted-l1 Minimization with Multiple Weighting Sets. In: Wavelets and Sparsity XIV, volume 8138, pp. 52-64. SPIE (2011)

39. Marim, M.M., Atlan, M., Angelini, E., Olivo-Marin, J.C.: Compressed sensing with off-axis frequency-shifting holography. Opt. Lett. **35**(6), 871–873 (2010)

40. McEwen, J.D., Wiaux, Y.: Compressed sensing for wide-field radio interferometric imaging. Mon. Notices Royal Astron. Soc. **413**(2), 1318–1332 (2011)

41. Nekhili, R.R.D., Descombes, X., Calatroni, L.: A hybrid approach combining cnns and variational modelling for blind image denoising (2022)

42. Oymak, S., Thrampoulidis, C., Hassibi, B.: The squared-error of generalized lasso: A precise analysis. In: 2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton) (pp. 1002-1009). IEEE (2013, October)

43. Oymak, S., Hassibi, B.: Sharp MSE Bounds for Proximal Denoising. Found. Comput. Math. **16**(4), 965–1029 (2016)

44. Plan, Y., Vershynin, R.: The Generalized Lasso With Non-Linear Observations. IEEE Trans. Inf. Theo. **62**(3), 1528–1537 (2016)

45. Pragliola, M., Calatroni, L., Lanza, A., Sgallari, F.: ADMM-based residual whiteness principle for automatic parameter selection in super-resolution problems (2021)

46. Pragliola, M., Calatroni, L., Lanza, A., Sgallari, F.: On and beyond Total Variation regularisation in imaging: the role of space variance. (2021). arXiv:2104.03650

47. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Second Printing, UK (1972)

48. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis, 3rd edn. Springer-Verlag, Berlin Heidelberg (2009)

49. Rudin, W.: Functional Analysis, 2nd edn. McGrow Hill, USA (1990)

50. Santosa, F., Symes, W.W.: Linear inversion of band-limited reflection seismograms. SIAM J. Sci. Stat. Comput. SIAM. **7**(4), 1307–1330 (1986). https://doi.org/10.1137/0907087

51. Soubies, E., Blanc-Féraud, L., Aubert, G.: A continuous exact ℓ_0 penalty (CEL0) for least squares regularized problem. SIAM J. Imaging Sci. **8**(3), 1607–1639 (2015)

52. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser B (Methodological) **58**(1), 267–288 (1996)

53. Unser, M.: A unifying representer theorem for inverse problems and machine learning. Found. Comput. Math. **21**(4), 941–960 (2021)

54. Unser, M., Aziznejad, S.: Convex optimization in sums of Banach spaces. Appl. Comput. Harmon. Anal. **56**, 1–25 (2022)

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

56. Yu, H., Wang, G.: Compressed sensing based interior tomography. Phys. Med. Biol. **54**(9), 2791 (2009)

57. Zhu, L., Zhang, W., Elnatan, D., Huang, B.: Faster STORM using compressed sensing. Nat. Methods **9**(7), 721–723 (2012)