Homogeneous Penalizers and Constraints in Convex Image Restoration[*]

R. Ciak[†]   B. Shafei[‡]   and G. Steidl[§]

November 8, 2012

Abstract

Recently convex optimization models were successfully applied for solving various problems in image analysis and restoration. In this paper, we are interested in relations between convex constrained optimization problems of the form $\text{argmin}\{\Phi(x)$ subject to $\Psi(x) \leq \tau\}$ and their penalized counterparts $\text{argmin}\{\Phi(x) + \lambda\Psi(x)\}$. We recall general results on the topic by the help of an epigraphical projection. Then we deal with the special setting $\Psi := \|L \cdot \|$ with $L \in \mathbb{R}^{m,n}$ and $\Phi := \varphi(H\cdot)$, where $H \in \mathbb{R}^{n,n}$ and $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ meet certain requirements which are often fulfilled in image processing models. In this case we prove by incorporating the dual problems that there exists a bijective function such that the solutions of the constrained problem coincide with those of the penalized problem if and only if τ and λ are in the graph of this function. We illustrate the relation between τ and λ for various problems arising in image processing. In particular, we point out the relation to the Pareto frontier for joint sparsity problems. We demonstrate the performance of the constrained model in restoration tasks of images corrupted by Poisson noise with the I-divergence as data fitting term φ and in inpainting models with the constrained nuclear norm. Such models can be useful if we have a priori knowledge on the image rather than on the noise level.

1 Introduction

In this paper, we are interested in the relation between the convex constrained optimization problem

$$(P_{1,\tau}) \qquad \underset{x \in \mathbb{R}^n}{\text{argmin}}\,\{\Phi(x) \quad \text{subject to} \quad \Psi(x) \leq \tau\} \tag{1}$$

and the non-constrained optimization problem

$$(P_{2,\lambda}) \qquad \underset{x \in \mathbb{R}^n}{\text{argmin}}\{\Phi(x) + \lambda\Psi(x)\}, \qquad \lambda \geq 0. \tag{2}$$

In the inverse problems and machine learning context these approaches are referred to as Ivanov regularization and Tikhonov regularization, respectively, see [35, 58].

[†]University Kaiserslautern, Dept. of Mathematics, Kaiserslautern, Germany
[‡]Fraunhofer ITWM, Fraunhofer Platz 1, 67663 Kaiserslautern, Germany
[§]University Kaiserslautern, Dept. of Mathematics, Kaiserslautern, Germany

Let $\mathrm{SOL}(P_\bullet)$ denote the set of solutions of problem $(P_\bullet)$. While it is rather clear that under mild conditions on Φ and Ψ a vector $\hat{x} \in \mathrm{SOL}(P_{2,\lambda})$, $\lambda > 0$ is also a solution of $(P_{1,\tau})$ exactly for $\tau = \Psi(\hat{x})$, the opposite direction has in general no simple explicit solution. At least it is known that for $\hat{x} \in \mathrm{SOL}(P_{1,\tau})$ there exists $\lambda \geq 0$ such that $\hat{x} \in \mathrm{SOL}(P_{2,\lambda})$. This result can be shown by using that under certain conditions the relation

$$\mathbb{R}_+ \, \partial\Psi(x) = \partial\iota_{\mathrm{lev}_{\Psi(x)}\Psi}(x)$$

holds true. In this paper, we prove the last equality using an epigraphical projection or briefly inf-projection, cf. [49, p. 18+], which allows to reduce the intrinsic problem to one dimension. Then we consider special problems where $\Psi := \|L \cdot \|$ with $L \in \mathbb{R}^{m,n}$ and $\Phi(x) := \phi(x_1)$. Here x_1 is the orthogonal projection of x onto a subspace X_1 of $\mathbb{R}^n$ and $\phi : X_1 \to \mathbb{R} \cup \{+\infty\}$ is a function which meets the following conditions: (i) $\mathrm{dom}\,\phi$ is an open subset of X_1 with $0 \in \overline{\mathrm{dom}\,\phi}$, (ii) ϕ is proper, convex and closed as well as strictly convex and essentially smooth, and (iii) ϕ has a minimizer. In our imaging applications, we will consider functions of the form $\Phi := \varphi(H\cdot)$ with a linear operator $H \in \mathbb{R}^{n,n}$. This fits into the above setting if X_1 is the range of H^*. We use the dual problems to prove that in a certain interval there is a one-to-one correspondence between τ and λ in the sense that $\mathrm{SOL}(P_{1,\tau}) = \mathrm{SOL}(P_{2,\lambda})$ exactly for the corresponding pairs. Furthermore, given τ, the value λ is determined by $\lambda := \|\hat{p}\|_*$, where $\hat{p}$ is any solution of the dual problem of $(P_{1,\tau})$.

Models with special functions Φ and Ψ of this kind were recently proposed for solving various image processing tasks. We will handle the following four example problems:

(E1) joint sparsity problems,

(E2) image restoration in the presence of Gaussian noise,

(E3) image restoration in the presence of Poissonian noise,

(E4) low rank matrix recovery such as matrix completion and tensor inpainting.

In particular, we are interested in the relation between τ and λ, i.e., in the τ-λ curve. For the numerical solution of the problems we restrict ourselves to forward-backward splitting (FBS) methods and alternating direction methods of multipliers (ADMM). We are aware of the various other techniques to speed up the computation, see, e.g. [2, 12, 19, 26, 44]; however the detection of the fastest algorithm is not the focus of this paper.

For the first task (E1), relations between slightly simpler non-constrained and constrained models were already stated in [21, 61] from another point of view. We point out the relation between the Pareto frontier considered in [61] and the τ-λ curve. Image processing results for the second task (E2) can be found in the recent paper [28]. In this paper we focus on the τ-λ curve. Since we are not aware of results using the constrained model $(P_{1,\tau})$ for the restoration of images corrupted by Poisson noise (E3) with the I-divergence as data fitting term φ we provide a deblurring result for this task in addition to the τ-λ curve. For the fourth task (E4) we give a numerical example for image inpainting via the constrained model with the nuclear norm.

This paper is organized as follows: First, we recall the relation between $(P_{1,\tau})$ and $(P_{2,\lambda})$ for a rather general setting in the next Section 2. In particular, we provide some proofs by incorporating an epigraphical projection. In Section 3 we restrict ourselves to homogeneous regularizers and to essentially smooth data terms, which are strictly convex on a certain

2

subspace of $\mathbb{R}^n$. We prove a relation between the parameters τ and λ such that the solution sets of the corresponding constrained and non-constrained problems coincide and determine λ corresponding to τ by duality arguments. Section 4 contains special examples from image processing where models of the described type are applied. It is not our intention to find the qualitatively best method or the fastest algorithm, but to illustrate the τ-λ relation. Conclusions are drawn in Section 5. Finally, the appendix describes the applied FBS and ADMM algorithms.

2 Penalizers and Constraints in Convex Problems

In this section we review general relations between constrained and non-constrained problems which will be specified for special settings appearing in image processing tasks in the subsequent sections.

Let $\mathbb{R}_+ := [0, +\infty)$ and let $\Gamma_0(\mathbb{R}^n)$ denote the space of proper, convex, closed functions mapping $\mathbb{R}^n$ into the extended real numbers $\mathbb{R} \cup \{+\infty\}$. For nonempty, affine subsets $X \subseteq \mathbb{R}^n$, we define $\Gamma_0(X)$ just analogously. The (lower) *level sets* of a function $\Psi \in \Gamma_0(\mathbb{R}^n)$ are given by $\mathrm{lev}_\tau \Psi := \{x : \Psi(x) \leq \tau\}$. Further, the *indicator function* ι_S of a set S is defined by

$$\iota_S(x) := \begin{cases} 0 & \text{if } x \in S, \\ +\infty & \text{otherwise.} \end{cases}$$

Using the indicator function, the constrained problem (1) can be rewritten as the following non-constrained one:

$$\underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \{\Phi(x) + \iota_{\mathrm{lev}_\tau \Psi}(x)\}. \tag{3}$$

For $x^* \in \mathbb{R}^n$ the *subdifferential* $\partial \Psi(x^*)$ of Ψ at x^* is the set

$$\partial \Psi(x^*) := \{p \in \mathbb{R}^n : \Psi(x^*) + \langle p, x - x^* \rangle \leq \Psi(x) \, \forall x \in \mathbb{R}^n\}.$$

If Ψ is proper, convex and $x^* \in \mathrm{ri}(\mathrm{dom}\,\Psi)$, then $\partial \Psi(x^*) \neq \emptyset$. Further we will need the Fenchel *conjugate function* of Ψ defined by

$$\Psi^*(p) := \sup_{x \in \mathbb{R}^n} \{\langle p, x \rangle - \Psi(x)\}.$$

For half-spaces, resp., hyperplanes we use the notation $H^\bullet_{p,\alpha} := \{x \in \mathbb{R}^n : \langle p, x \rangle \bullet \alpha\}$, where $\bullet \in \{\leq, >, =\}$.

In Lemma 2.2 we will provide a basic relation for understanding the correspondence between the constrained and penalized problems. To prove Lemma 2.2 we need the following auxiliary lemma which simple proof is left to the reader or can be found in [18].

Lemma 2.1. *Let* $\Psi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *be a proper, convex function,* $x^* \in \mathrm{int}(\mathrm{dom}\,\Psi)$ *and* $S := \mathrm{lev}_{\Psi(x^*)} \Psi$. *Let* $p \in \mathbb{R}^n$ *such that the half-space* $H^\leq_{p,\alpha}$ *with* $\alpha := \langle p, x^* \rangle$ *contains* S. *Then we have the equality*

$$\inf_{x \in H^=_{p,\alpha}} \Psi(x) = \Psi(x^*). \tag{4}$$

The following lemma will be used in our proof of Theorem 2.4.

Lemma 2.2. *Let $\Psi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper, convex function, $x^* \in \mathrm{dom}\,\Psi$ and $S := \mathrm{lev}_{\Psi(x^*)}\Psi$. Then we have*

$$\mathbb{R}_+ \, \partial\Psi(x^*) \subseteq \partial\iota_S(x^*). \tag{5}$$

Moreover, if $x^* \in \mathrm{int}(\mathrm{dom}\,\Psi)$ and x^* is not a minimizer of Ψ, then equality in (5) holds true.

A proof of a similar lemma for finite functions $\Psi : \mathbb{R}^n \to \mathbb{R}$ based on cone relations can be found, e.g., in [34, p. 245]. Here we provide a proof which uses the *epigraphical projection*, also known as *inf-projection* as defined in [49, p. 18+, p. 51]. For a function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$, the inf-projection is defined by $\nu(u) := \inf_x f(x,u)$. The name 'epigraphical projection' is due to the following fact: $\mathrm{epi}\,\nu$ is the image of $\mathrm{epi}\,f$ under the projection $(x, u, \alpha) \mapsto (u, \alpha)$, if $\mathrm{argmin}_x f(x, u)$ is attained for each $u \in \mathrm{dom}\,\nu$. (Note that this is not the projection onto epigraphs as used, e.g., in [1, p. 427].) The inf-projection is convexity preserving, i.e., if f is convex, then ν is also convex, cf. [49, Proposition 2.22].

Proof. 1. First we show that $\mathbb{R}_+ \, \partial\Psi(x^*) \subseteq \partial\iota_S(x^*)$. By definition of the subdifferential we obtain

$$
\begin{aligned}
q \in \partial\Psi(x^*) &\iff \langle q, x - x^* \rangle \leq \Psi(x) - \Psi(x^*) \qquad \forall x \in \mathbb{R}^n, \\
&\implies \langle q, x - x^* \rangle \leq 0 \qquad \forall x \in S.
\end{aligned}
$$

Then we obtain the above inclusion by

$$p \in \partial\iota_S(x^*) \iff \langle p, x - x^* \rangle \leq 0 \qquad \forall x \in S. \tag{6}$$

2. Next we prove $\partial\iota_S(x^*) \subseteq \mathbb{R}_+ \, \partial\Psi(x^*)$ if x^* is not a minimizer of Ψ and $x^* \in \mathrm{int}(\mathrm{dom}\,\Psi)$. Let $p \in \partial\iota_S(x^*)$. If p is the zero vector, then we are done since $\partial\Psi(x^*) \neq \emptyset$. In the following we assume that p is not the zero vector. It remains to show that there exists $h > 0$ such that $\frac{1}{h}p \in \partial\Psi(x^*)$. We can restrict our attention to $p = (0, \ldots, 0, p_n)^{\mathrm{T}}$ with $p_n > 0$. (Otherwise we can perform a suitable rotation of the coordinate system.) Then (6) becomes

$$p \in \partial\iota_S(x^*) \iff p_n x_n \leq p_n x_n^* \qquad \forall x = (\bar{x}, x_n) \in S. \tag{7}$$

Hence we can apply lemma 2.1 with $p = (0, \ldots, 0, p_n)^{\mathrm{T}}$ and obtain

$$\inf_{\{x \in \mathbb{R}^n : x_n = x_n^*\}} \Psi(x) = \Psi(x^*).$$

Introducing the inf-projection $\nu : \mathbb{R} \to \mathbb{R} \cup \{\pm\infty\}$ by

$$\nu(x_n) := \inf_{\bar{x} \in \mathbb{R}^{n-1}} \Psi(\bar{x}, x_n).$$

this can be rewritten as

$$\nu(x_n^*) = \Psi(x^*). \tag{8}$$

Therefore we have

$$
\begin{aligned}
\frac{1}{h}p = (0, \ldots, 0, \tfrac{1}{h}p_n)^{\mathrm{T}} \in \partial\Psi(x^*) &\iff \Psi(x) \geq \nu(x_n^*) + \frac{1}{h}p_n(x_n - x_n^*) \quad \forall x \in \mathbb{R}^n \\
&\iff \nu(x_n) \geq \nu(x_n^*) + \frac{1}{h}p_n(x_n - x_n^*) \quad \forall x_n \in \mathbb{R} \\
&\iff \frac{1}{h}p_n \in \partial\nu(x_n^*),
\end{aligned}
$$

4

so that it remains to show that $\partial\nu(x_n^*)$ contains a positive number. By (8) we verify that $\nu(x_n^*)$ is finite. Moreover, $x^* \in \text{int}(\text{dom}\Psi)$ implies $x_n^* \in \text{int}(\text{dom}\nu)$. Therefore, $\partial\nu(x_n^*) \neq \emptyset$. Let $q_n \in \partial\nu(x_n^*)$, i.e.,

$$q_n(x_n - x_n^*) \leq \nu(x_n) - \nu(x_n^*) \quad \forall x_n \in \mathbb{R}.$$

Since x^* is not a minimizer of Ψ, there exists $y \in \mathbb{R}^n$ with $\Psi(y) < \Psi(x^*)$ and we get by (7) that $y_n \leq x_n^*$. Since $y_n = x_n^*$ would by (8) imply that $\Psi(x^*) = \nu(y_n) \leq \Psi(y)$, we indeed have $y_n < x_n^*$. Thus

$$q_n(y_n - x_n^*) \leq \nu(y_n) - \nu(x_n^*) \leq \Psi(y) - \Psi(x^*) < 0$$

which implies $q_n > 0$ and we are done. $\qquad\square$

Remark 2.3. i) *The condition that x^* is not a minimizer of Ψ is essential to have equality in (5) as the following example illustrates. The function Ψ given by $\Psi(x) = x^2$ is minimal at $x^* = 0$. We have $S := \text{lev}_{\Psi(0)}\Psi = \{0\}$ so that*

$$\mathbb{R}_+\partial\Psi(x^*) = \{0\} \subset \mathbb{R} = \partial\iota_S(x^*).$$

ii) *In [18] it was shown that if x^* is not a minimizer of Ψ and $x^* \in \text{ri}(\text{dom}\,\Psi)$, then*

$$\partial\iota_S(x^*) = \overline{\mathbb{R}_+\partial\Psi(x^*)}.$$

However the condition $x^* \in \text{ri}(\text{dom}\,\Psi)$ is not sufficient to guarantee equality in (5): consider the function $\Psi : \mathbb{R}^2 \to \mathbb{R} \cup \{+\infty\}$, given by

$$\Psi(x_1, x_2) := \begin{cases} x_1 & \text{if } x_2 = 0, \\ +\infty & \text{if } x_2 \neq 0. \end{cases}$$

The affine hull $\text{aff}(\text{dom}\,\Psi) = \mathbb{R} \times \{0\}$ is a proper subset of $\mathbb{R}^2$. We have $S := \text{lev}_{\Psi(x^*)}\Psi = (-\infty, x_1^*] \times \{0\}$ for any $x^* = (x_1^*, 0) \in \mathbb{R} \times \{0\} = \text{aff}(\text{dom}\,\Psi) = \text{ri}(\text{dom}\,\Psi)$ and further $\partial\Psi(x^*) = \{(1, p_2)^T : p_2 \in (-\infty, +\infty)\}$ so that

$$\mathbb{R}_+\partial\Psi(x^*) = \{(0, 0)^T\} \cup \{(p_1, p_2)^T : p_1 \in (0, +\infty), p_2 \in (-\infty, +\infty)\}.$$

On the other hand we get

$$\partial\iota_S(x^*) = \{(p_1, p_2)^T : p_1 \in [0, +\infty), p_2 \in (-\infty, +\infty)\}$$

so that $\mathbb{R}_+\partial\Psi(x^*) \subset \partial\iota_S(x^*)$.

Using Lemma 2.2 it easy to prove the following theorem on the correspondence between the constrained problem $(P_{1,\tau})$ and the penalized problem $(P_{2,\lambda})$.

Theorem 2.4. i) *Let $\Phi, \Psi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper, convex functions. Assume that there exists a point in $\text{dom}\,\Phi \cap \text{lev}_\tau\Psi$ where one of the functions Φ or $\iota_{\text{lev}_\tau\Psi}$ is continuous. Let $\hat{x} \in \text{int}(\text{dom}\,\Psi)$ be a minimizer of $(P_{1,\tau})$, where $\hat{x}$ is not a minimizer of Ψ in case $\Psi(\hat{x}) = \tau$. Then there exists a parameter $\lambda \geq 0$ such that $\hat{x}$ is also a minimizer of $(P_{2,\lambda})$. If $\hat{x}$ is in addition not a minimizer of Φ, then $\lambda > 0$.*
ii) *For proper $\Phi, \Psi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ with $\text{dom}\,\Phi \cap \text{dom}\,\Psi \neq \emptyset$, let $\hat{x}$ be a minimizer of $(P_{2,\lambda})$. If $\lambda = 0$, then $\hat{x}$ is also a minimizer of $(P_{1,\tau})$ if and only if $\tau \geq \Psi(\hat{x})$. If $\lambda > 0$, then $\hat{x}$ is also a minimizer of $(P_{1,\tau})$ for $\tau := \Psi(\hat{x})$. Moreover, if Φ, Ψ are proper, convex functions and $\hat{x} \in \text{int}(\text{dom}\,\Psi)$, this τ is unique if and only if $\hat{x}$ is not a minimizer of Φ.*

5

Proof. i) Let $\hat{x} \in \mathrm{SOL}(P_{1,\tau}) \cap \mathrm{int}(\mathrm{dom}\,\Psi)$. Then $\Psi(\hat{x}) \leq \tau$ holds true.

In the case $\Psi(\hat{x}) < \tau$, the continuity of Ψ in $\mathrm{int}(\mathrm{dom}\,\Psi)$ assures that $\Psi(x) < \tau$ in a neighborhood of $\hat{x}$. Thus $\hat{x}$ is a local minimizer of Φ in this neighborhood. Since Φ is convex, $\hat{x}$ is also a global minimizer of Φ and hence a solution of $(P_{2,\lambda})$ at least for $\lambda = 0$.

Consider the case $\Psi(\hat{x}) = \tau$ and assume that $\hat{x}$ is not a minimizer of Ψ. By Fermat's rule, the regularity condition and Lemma 2.2 we obtain

$$0 \in \partial\big(\Phi + \iota_{\mathrm{lev}_\tau \Psi}\big)(\hat{x}) = \partial\Phi(\hat{x}) + \partial\iota_{\mathrm{lev}_\tau \Psi}(\hat{x}) = \partial\Phi(\hat{x}) + \mathbb{R}_+ \partial\Psi(\hat{x}). \tag{9}$$

This means that there exists $\lambda \geq 0$ such that $0 \in \partial\Phi(\hat{x}) + \lambda\partial\Psi(\hat{x}) \subseteq \partial\big(\Phi + \lambda\Psi\big)(\hat{x})$ so that by Fermat's rule $\hat{x}$ is a minimizer of $(P_{2,\lambda})$. If $\hat{x}$ is not a minimizer of Φ, then clearly $\lambda > 0$.

ii) Let $\hat{x} \in \mathrm{SOL}(P_{2,\lambda})$. Now $\hat{x}$ can only be a minimizer of $(P_{1,\tau})$ for $\tau \geq \Psi(\hat{x})$.

If $\lambda = 0$, then $\hat{x}$ is a minimizer of Φ and consequently of $(P_{1,\tau})$ for all these values of τ.

If $\lambda > 0$, then $\hat{x} \in \mathrm{SOL}(P_{1,\tau})$ at least for $\tau := \Psi(\hat{x})$ by the following reason: if there would exist $\tilde{x}$ with $\Phi(\tilde{x}) < \Phi(\hat{x})$ and $\Psi(\tilde{x}) \leq \tau$, then $\Phi(\tilde{x}) + \lambda\Psi(\tilde{x}) < \Phi(\hat{x}) + \lambda\Psi(\hat{x})$ which contradicts $\hat{x} \in \mathrm{SOL}(P_{2,\lambda})$. Finally, let in addition Φ, Ψ be convex and $\hat{x} \in \mathrm{int}(\mathrm{dom}\,\Psi)$ be not a minimizer of Φ. If there exists another $\tilde{\tau} > \tau = \Psi(\hat{x})$ such that $\hat{x}$ is a minimizer of $(P_{1,\tilde{\tau}})$, then we see as in the proof of part i) that $\hat{x}$ is a minimizer of Φ. This yields a contradiction. $\square$

If $\Psi : \mathbb{R}^n \to \mathbb{R}$ is a finite, convex function, then the regularity condition in part i) of the theorem (and clearly also the other assumptions) is fulfilled if there exists $x_0 \in \mathrm{dom}\,\Phi$ such that $\Psi(x_0) < \tau$. In this case, the existence of a Lagrange multiplier $\lambda \geq 0$ is assured by [48, Corollary 28.2.1].

For our general setting, the assumptions of the theorem are needed since $\Psi(x) < \tau$ does not imply $x \in \mathrm{ri}(\mathrm{lev}_\tau \Psi)$ as Fig. 1 shows.
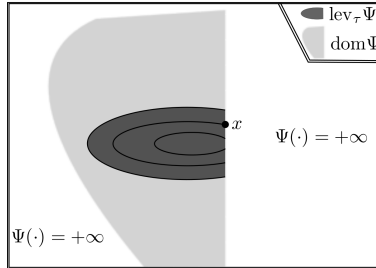

Figure 1: Example where $\Psi(x) < \tau$ does not imply $x \in \mathrm{ri}(\mathrm{lev}_\tau \Psi)$.

Remark 2.5. i) *Concerning part* i) *of the theorem, there may exist various parameters* λ *corresponding to the same parameter* τ *also if* $(P_{1,\tau})$ *has a unique solution: consider* $\Psi(x) := |x|$ *and*

$$\Phi(x) := \begin{cases} (x-2)^2 & \text{if } x \geq 1, \\ m(x-1) + 1 & \text{if } x < 1, \end{cases}$$

where $m \leq -2$. *The function* Φ *is differentiable for* $m = -2$. *Noting that* Φ *is strictly decreasing on* $(0, 2)$ *we see that*

$$\mathrm{argmin}_{x \in \mathbb{R}} \big\{ \Phi(x) \quad \text{subject to} \quad |x| \leq \tau \big\} = \{\tau\} \qquad \forall \tau \in (0, 2).$$

6

On the other hand, we get

$$\underset{x\in\mathbb{R}}{\operatorname{argmin}} \left\{ \Phi(x) + \lambda|x| \right\} = \begin{cases} \{2 - \frac{\lambda}{2}\} & \text{if } \lambda \in [0, 2), \\ \{1\} & \text{if } \lambda \in [2, -m), \\ [0, 1] & \text{if } \lambda = -m, \\ \{0\} & \text{if } \lambda \in (-m, +\infty), \end{cases}$$

so that $\tau = 1$ corresponds to $\lambda \in [2, -m]$. It is known that the set of Lagrange multipliers λ is a bounded, closed interval under certain assumptions, see [48, Corollary 29.1.5]
ii) Concerning Part ii) of the theorem in case that there are different minimizers of $(P_{2,\lambda})$, say $\hat{x}_1$ and $\hat{x}_2$, we notice that $\Psi(\hat{x}_1) \neq \Psi(\hat{x}_2)$ can appear as the following example shows: For $\Phi(x) := |x - 2|$ and $\Psi(x) := |x|$ and $\lambda = 1$ we have

$$(P_{2,1}) \qquad \Phi(x) + \Psi(x) = \begin{cases} -2(x - 1) & \text{if } x < 0, \\ 2 & \text{if } x \in [0, 2], \\ +2(x - 1) & \text{if } x > 2, \end{cases}$$

i.e., $\operatorname{argmin}_{x\in\mathbb{R}} \left\{ \Phi(x) + \Psi(x) \right\} = [0, 2]$. Hence we can choose, e.g., $\hat{x}_1 = 1$ and $\hat{x}_2 = \frac{3}{2}$ and obtain $\Psi(\hat{x}_1) = 1 \neq \frac{3}{2} = \Psi(\hat{x}_2)$.

Using duality arguments we will specify the relations between $(P_{1,\tau})$ and $(P_{2,\lambda})$ for special tasks appearing in image processing in the subsequent sections. In particular, we want to determine λ in part i) of Theorem 2.4. To this end, we need the following known Fenchel duality relation, compare, e.g., [49, p. 505] and [18].

Lemma 2.6. *Let $\Phi \in \Gamma_0(\mathbb{R}^n)$, $\Psi \in \Gamma_0(\mathbb{R}^m)$, $L \in \mathbb{R}^{m,n}$ and $\mu > 0$. Assume that the following conditions are fulfilled.*

   i) $\operatorname{ri}(\operatorname{dom}\Phi) \cap \operatorname{ri}(\operatorname{dom}\Psi(\mu L \cdot)) \neq \emptyset$,

   ii) $\mathcal{R}(L) \cap \operatorname{ri}(\operatorname{dom}\Psi(\mu \cdot)) \neq \emptyset$,

   iii) $\operatorname{ri}(\operatorname{dom}\Phi^*(-L^* \cdot)) \cap \operatorname{ri}(\operatorname{dom}\Psi^*(\frac{\cdot}{\mu})) \neq \emptyset$,

   iv) $\mathcal{R}(-L^*) \cap \operatorname{ri}(\operatorname{dom}\Phi^*) \neq \emptyset$.

Then, the primal problem

$$(P) \qquad \underset{x\in\mathbb{R}^n}{\operatorname{argmin}} \left\{ \Phi(x) + \Psi(\mu L x) \right\}, \qquad \mu > 0, \tag{10}$$

has a solution if and only if the dual problem

$$(D) \qquad \underset{p\in\mathbb{R}^m}{\operatorname{argmin}} \left\{ \Phi^*(-L^* p) + \Psi^*\left(\frac{p}{\mu}\right) \right\} \tag{11}$$

has a solution. Furthermore $\hat{x} \in \mathbb{R}^n$ and $\hat{p} \in \mathbb{R}^m$ are solutions of the primal and the dual problem, respectively, if and only if

$$\frac{1}{\mu}\hat{p} \in \partial\Psi(\mu L\hat{x}) \quad \text{and} \quad -L^*\hat{p} \in \partial\Phi(\hat{x}). \tag{12}$$

3 Homogeneous Penalizers and Constraints

In the rest of this paper, we deal with the functions

$$\Psi_1 := \iota_{lev_1\|\cdot\|} \quad \text{and} \quad \Psi_2 = \|\cdot\|,$$

where $\|\cdot\|$ denotes an arbitrary norm in $\mathbb{R}^m$ with dual norm $\|\cdot\|_* := \max_{\|x\|\leq 1}\langle\cdot, x\rangle$. Constraints and penalizers of this kind appear in many image processing tasks. Note that $\Psi_1(\tau^{-1}x) = \iota_{lev_\tau\|\cdot\|}(x) = \tau\iota_{lev_\tau\|\cdot\|}(x)$. The conjugate functions of Ψ_1 and Ψ_2 are

$$\Psi_1^* = \|\cdot\|_* \quad \text{and} \quad \Psi_2^* = \iota_{lev_1\|\cdot\|_*}.$$

Then the primal problems (P) in (10) with $\mu := \tau^{-1} > 0$ in the case $\Psi = \Psi_1$ and $\mu := \lambda > 0$ in the case $\Psi = \Psi_2$ become

$$(P_{1,\tau}) \quad \operatorname*{argmin}_{x\in\mathbb{R}^n}\{\Phi(x) \quad \text{subject to} \quad \|Lx\| \leq \tau\},$$

$$(P_{2,\lambda}) \quad \operatorname*{argmin}_{x\in\mathbb{R}^n}\{\Phi(x) + \lambda\|Lx\|\}$$

and the dual problems (D) in (11) read

$$(D_{1,\tau}) \quad \operatorname*{argmin}_{p\in\mathbb{R}^m}\{\Phi^*(-L^*p) + \tau\|p\|_*\},$$

$$(D_{2,\lambda}) \quad \operatorname*{argmin}_{p\in\mathbb{R}^m}\{\Phi^*(-L^*p) \quad \text{subject to} \quad \|p\|_* \leq \lambda\}.$$

We are interested in the relation between τ and λ in particular in the computation of λ given τ such that the problems have the same solution set. In the following, we suppose an orthogonal decomposition

$$\mathbb{R}^n = X_1 \oplus X_2$$

of $\mathbb{R}^n$ into subspaces X_i, $i = 1, 2$. In our imaging applications we will use the decomposition $\mathbb{R}^n = \mathcal{R}(H^*) \oplus \mathcal{N}(H)$, where $H : \mathbb{R}^n \to \mathbb{R}^m$ is a linear operator with *kernel* $\mathcal{N}(H)$ and $\mathcal{R}(H^*)$ is the *range* of H^*. We assume that Φ has the special form

$$\Phi(x) = \Phi(x_1 + x_2) = \phi(x_1), \tag{13}$$

where $\phi : X_1 \to \mathbb{R} \cup \{+\infty\}$ is a function meeting the conditions

(i) $\operatorname{dom}\phi$ is an open subset of X_1 with $0 \in \overline{\operatorname{dom}\phi}$,

(ii) ϕ belongs to $\Gamma_0(X_1)$ and is strictly convex and essentially smooth (compare [48, p.251]),

(iii) ϕ has a minimizer.

By (ii) and [48, p.253] it follows that ϕ^* belongs to $\Gamma_0(X_1)$ and is essentially smooth and essentially strictly convex. Straightforward computation gives

$$\partial\Phi(x) = \partial\Phi(x_1 + x_2) = \partial\phi(x_1) \tag{14}$$

and

$$\Phi^*(x^*) = \Phi^*(x_1^* + x_2^*) = \begin{cases} \phi^*(x_1^*) & \text{if } x_2^* = 0, \\ +\infty & \text{otherwise.} \end{cases} \tag{15}$$

Further, $\operatorname{argmin} \phi$ consists of just one element and is therefore a nonempty and bounded level set of ϕ. Consequently all level sets $\operatorname{lev}_\alpha \phi$, $\alpha \in \mathbb{R}$, are bounded, cf. [48, Corollary 8.7.1] which implies $0 \in \operatorname{int}(\operatorname{dom} \phi^*)$ by [48, Corollary 14.2.2]. Since $\operatorname{dom} \phi^* = \operatorname{dom} \Phi^*$ we finally obtain

$$0 \in \operatorname{int}(\operatorname{dom} \phi^*) \quad \text{and} \quad 0 \in \operatorname{ri}(\operatorname{dom} \Phi^*). \tag{16}$$

The rest of this section is organized as follows: In Theorem 3.4 we show that the problems $(P_{1,\tau})$, $(P_{2,\lambda})$, $(D_{1,\tau})$, $(D_{2,\lambda})$ have a solution for $\tau > 0$ and $\lambda > 0$, if certain conditions are fulfilled. In Lemma 3.5 we prove that under the same conditions there are intervals $(0, c)$ and $(0, d)$ such that $\operatorname{SOL}(P_{1,\tau})$, $\operatorname{SOL}(P_{2,\lambda})$, $\operatorname{SOL}(D_{1,\tau})$, $\operatorname{SOL}(D_{2,\lambda})$ have certain localization properties. This is refined in our main Theorem 3.6, where we obtain that $\operatorname{SOL}(P_{1,\tau}) = \operatorname{SOL}(P_{2,\lambda})$ if and only if (τ, λ) is in the graph of a certain function g.

The following two lemmas on properties of 'translation invariant' functions prepare the proofs of the subsequent theorems.

Lemma 3.1. *Let $\mathbb{R}^n$ be decomposed as direct sums $\mathbb{R}^n = U_1 \oplus U_2$ and $\mathbb{R}^n = V_1 \oplus V_2$ of subspaces U_1, U_2 and V_1, V_2, where $U_2 \cap V_2 = \{0\}$. Let $F, G \in \Gamma_0(\mathbb{R}^n)$ be functions with*

$$F(x) = F(x + u_2), \quad G(x) = G(x + v_2) \tag{17}$$

for all $x \in \mathbb{R}^n$, $u_2 \in U_2$ and $v_2 \in V_2$. Set $f := F|_{U_1}$ and $g := G|_{V_1}$.

   i) *If $\operatorname{lev}_\alpha f$, $\operatorname{lev}_\beta g$ are bounded for some $\alpha, \beta \in \mathbb{R}$, then $\operatorname{lev}_\alpha F \cap \operatorname{lev}_\beta G$ is bounded.*

   ii) *If $\operatorname{dom} F \cap \operatorname{dom} G \neq \emptyset$ and $\operatorname{lev}_{\tilde{\alpha}} f$, $\operatorname{lev}_{\tilde{\beta}} g$ are nonempty and bounded for some $\tilde{\alpha}, \tilde{\beta} \in \mathbb{R}$, then $F + G$ attains its finite minimum.*

Proof. i) Let $\operatorname{lev}_\alpha f$, $\operatorname{lev}_\beta g$ be bounded. Assume that $\operatorname{lev}_\alpha F \cap \operatorname{lev}_\beta G$ is unbounded. Then it contains an unbounded sequence $z^{(k)}$, $k \in \mathbb{N}$. Since $\operatorname{lev}_\alpha F = \operatorname{lev}_\alpha f \oplus U_2$ and $\operatorname{lev}_\beta G = \operatorname{lev}_\beta g \oplus V_2$ we obtain $z^{(k)} = u_1^{(k)} + u_2^{(k)} = v_1^{(k)} + v_2^{(k)}$ with bounded first components $u_1^{(k)} \in \operatorname{lev}_\alpha f$, $v_1^{(k)} \in \operatorname{lev}_\beta g$ and unbounded second components $u_2^{(k)} \in U_2$, $v_2^{(k)} \in V_2$. Further we have

$$\begin{aligned}
\|u_2^{(k)} - v_2^{(k)}\|_2^2 &= \|u_2^{(k)}\|_2^2 + \|v_2^{(k)}\|_2^2 - 2\langle u_2^{(k)}, v_2^{(k)} \rangle \\
&= \left(\|u_2^{(k)}\|_2 - \|v_2^{(k)}\|_2 \right)^2 + 2\|u_2^{(k)}\|_2 \|v_2^{(k)}\|_2 \left(1 - \left\langle \frac{u_2^{(k)}}{\|u_2^{(k)}\|_2}, \frac{v_2^{(k)}}{\|v_2^{(k)}\|_2} \right\rangle \right) \\
&\geq 2\|u_2^{(k)}\|_2 \|v_2^{(k)}\|_2 \left(1 - \left\langle \frac{u_2^{(k)}}{\|u_2^{(k)}\|_2}, \frac{v_2^{(k)}}{\|v_2^{(k)}\|_2} \right\rangle \right)
\end{aligned}$$

Using $U_2 \cap V_2 = \{0\}$ and the compactness of $\{x \in \mathbb{R}^d : \|x\|_2 = 1\} \cap U_2$ and $\{x \in \mathbb{R}^d : \|x\|_2 = 1\} \cap V_2$, we conclude by the Cauchy–Schwarz inequality for linearly independent vectors that the right-hand side goes to infinity as $k \to +\infty$. Now the unboundedness of the $\|u_2^{(k)} - v_2^{(k)}\|$ leads to the contradiction

$$\begin{aligned}
0 = \|z^{(k)} - z^{(k)}\|_2 &= \|u_1^{(k)} - v_1^{(k)} + u_2^{(k)} - v_2^{(k)}\|_2 \geq \|u_2^{(k)} - v_2^{(k)}\|_2 - \|u_1^{(k)} - v_1^{(k)}\|_2 \\
&\geq \|u_2^{(k)} - v_2^{(k)}\|_2 - (\|u_1^{(k)}\|_2 + \|v_1^{(k)}\|_2) \to +\infty.
\end{aligned}$$

ii) Let the assumptions on F and G be fulfilled. Because of $\operatorname{dom} F \cap \operatorname{dom} G \neq \emptyset$ we have $\operatorname{lev}_\alpha(F) \cap \operatorname{lev}_\beta(G) \neq \emptyset$ for $\alpha, \beta \in \mathbb{R}$ chosen large enough. Due to $\operatorname{lev}_\alpha f \oplus U_2 = \operatorname{lev}_\alpha F \neq \emptyset$ and

9

$\operatorname{lev}_\beta g \oplus V_2 = \operatorname{lev}_\beta G \neq \emptyset$ the compact level sets $\operatorname{lev}_\alpha f$ and $\operatorname{lev}_\beta g$ are nonempty. Since f and g are proper this implies that f and g are bounded from below. Without loss of generality we may therefore assume $f \geq 0$ and $g \geq 0$ so that $F \geq 0$ and $G \geq 0$. Next we show that $\operatorname{lev}_{\alpha+\beta}(F+G)$ is a nonempty compact set. We have $\operatorname{lev}_{\alpha+\beta}(F + G) \supseteq \operatorname{lev}_\alpha(F) \cap \operatorname{lev}_\beta(G) \neq \emptyset$. Furthermore $\operatorname{lev}_{\alpha+\beta}(F+G)$ is closed since $F+G$ is lsc. Since the nonempty level sets $\operatorname{lev}_{\tilde{\alpha}}(f)$ and $\operatorname{lev}_{\tilde{\beta}}(g)$ are bounded we obtain that $\operatorname{lev}_{\alpha+\beta}(f)$ and $\operatorname{lev}_{\alpha+\beta}(g)$ are bounded, cf. [48, Corollary 8.7.1]. Hence, by i), the nonempty and closed set $\operatorname{lev}_{\alpha+\beta}(F + G) \subseteq \operatorname{lev}_{\alpha+\beta}(F) \cap \operatorname{lev}_{\alpha+\beta}(G)$ is also bounded and thus compact. Therefore the lsc function $(F + G)|_{\operatorname{lev}_{\alpha+\beta}(F+G)} = F + G + \iota_{\operatorname{lev}_{\alpha+\beta}(F+G)}$ must be minimized by an $\check{u} \in \operatorname{lev}_{\alpha+\beta}(F + G)$ which clearly also minimizes $F + G$. $\square$

Lemma 3.2. *Let the Euclidean space $\mathbb{R}^n$ be decomposed into the direct sum $\mathbb{R}^n = U_1 \oplus U_2$ of two subspaces U_1, U_2 and let $F : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a convex function which is strictly convex on U_1 and which inheres the translation invariance $F(x) = F(x + u_2)$ for all $x \in \mathbb{R}^n$ and $u_2 \in U_2$. Furthermore, let $G : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be any convex function with $\operatorname{dom} F \cap \operatorname{dom} G \neq \emptyset$. Then all $\hat{x}, \tilde{x} \in \operatorname{argmin}_{x \in \mathbb{R}^n} \{F(x) + G(x)\}$ fulfill $\hat{x} - \tilde{x} \in U_2$ and $F(\hat{x}) = F(\tilde{x})$, $G(\hat{x}) = G(\tilde{x})$.*

Proof. 1. First we prove that for any $x, y \in \operatorname{dom} F$ and the line segment $l(x, y) := \{x + t(y - x) : t \in [0, 1]\}$ the following statements are equivalent:

   i) $F|_{l(x,y)}$ is constant,

   ii) $F|_{l(x,y)}$ is affine,

   iii) $y - x \in U_2$.

We use the unique decompositions $x = x_1 + x_2$, $y = y_1 + y_2$ with $x_1, y_1 \in U_1$ and $x_2, y_2 \in U_2$.
i) $\Rightarrow$ ii): This is clear since a constant function is in particular an affine one.
ii) $\Rightarrow$ iii): If $F|_{l(x,y)}$ is affine, i.e.,

$$F(x + t(y - x)) = F(x) + t(F(y) - F(x)) \quad \text{for every } t \in [0, 1],$$

the translation invariance of F yields

$$F(x + t(y - x) - x_2 - t(y_2 - x_2)) = F(x - x_2) + t(F(y - y_2) - F(x - x_2)),$$
$$F(x_1 + t(y_1 - x_1)) = F(x_1) + t(F(y_1) - F(x_1)) \quad \text{for every } t \in [0, 1],$$

so that $F|_{l(x_1,y_1)}$ is affine as well. On the other hand F is also strictly convex on $l(x_1, y_1)$. Both can be simultaneously only true, if $x_1 = y_1$, which just means that $y - x = y_2 - x_2 \in U_2$.
iii) $\Rightarrow$ i): Assume that $y - x \in U_2$, i.e. $y_1 = x_1$, so that $y - x = y_2 - x_2$. Therefore and due to the translation invariance of F we get

$$F(x + t(y - x)) = F(x + t(y_2 - x_2)) = F(x)$$

even for all $t \in \mathbb{R}$. In particular F is constant on $l(x, y)$.
2. Now the assertion can be seen as follows: Due to the convexity of $F + G$ the whole segment $l(\hat{x}, \tilde{x})$ belongs to $\operatorname{argmin}\{F + G\}$ so that $F + G$ is constant on $l(\hat{x}, \tilde{x})$. Thus, the convex summands F and G must be affine on $l(\hat{x}, \tilde{x})$. Now part 1 of the proof tells us that $\hat{x} - \tilde{x} = -(\tilde{x} - \hat{x}) \in U_2$ and $F(\hat{x}) = F(\tilde{x})$. The remaining $G(\hat{x}) = G(\tilde{x})$ follows from the last equation and from $F(\hat{x}) + G(\hat{x}) = F(\tilde{x}) + G(\tilde{x})$ since only finite values occur. $\square$

Remark 3.3. *Provided that the stronger "overlapping" condition* $\mathrm{ri}(\mathrm{dom}\, F) \cap \mathrm{ri}(\mathrm{dom}\, G) \neq \emptyset$ *holds true, the assertion of the lemma is still valid, if F is essentially strictly convex on U_1 and essentially smooth on the affine hull of* $\mathrm{dom}\, F|_{U_1}$, *cf. [18]. Furthermore we have* $\mathrm{argmin}\{F + G\} \subseteq \mathrm{ri}(\mathrm{dom}\, F)$ *under these assumptions.*

The next theorem shows that the problems $(P_{1,\tau})$, $(D_{1,\tau})$, $(P_{2,\lambda})$, $(D_{2,\lambda})$ have a solution for $\tau > 0$ and $\lambda > 0$ if certain conditions on $\mathrm{argmin}\, \Phi$ and $\mathcal{N}(L) = \mathrm{argmin}\, \|L \cdot \|$ are fulfilled.

Theorem 3.4. *Let $\Phi \in \Gamma_0(\mathbb{R}^n)$ be a function fulfilling (13) and let $L \in \mathbb{R}^{m,n}$ so that $X_2 \cap \mathcal{N}(L) = \{0\}$ and $\mathrm{argmin}\, \Phi \cap \mathcal{N}(L) = \emptyset$. Then the sets* $\mathrm{SOL}(P_{1,\tau})$, $\mathrm{SOL}(D_{1,\tau})$, $\mathrm{SOL}(P_{2,\lambda})$, $\mathrm{SOL}(D_{2,\lambda})$ *are nonempty for $\tau \in (0, +\infty)$, $\lambda \in (0, +\infty)$ with corresponding finite minima.*

Proof. First we note that the requirements i) - iv) of Lemma 2.6 are fulfilled.
1. Let $\lambda > 0$. Since $\Phi^*(-L^* \cdot)$ is lsc on the compact ball $B_\lambda := \{p \in \mathbb{R}^m : \|p\|_* \leq \lambda\}$ we obtain $\mathrm{SOL}(D_{2,\lambda}) \neq \emptyset$. By (16) we have $0 \in B_\lambda \cap \mathrm{dom}(\Phi^*(-L^* \cdot))$ so that the attained minimum is finite. Lemma 2.6 ensures that also $\mathrm{SOL}(P_{2,\lambda}) \neq \emptyset$, where the attained minimum is finite, since $\mathrm{dom}(\Phi + \lambda\|L \cdot \|) = \mathrm{dom}\, \Phi \neq \emptyset$.
2. Let $\tau > 0$. To get $\mathrm{SOL}(P_{1,\tau}) \neq \emptyset$, we want to apply Lemma 3.1 ii), with $F := \Phi$, $U_i := X_i$, $i = 1, 2$ and $G := \iota_{\mathrm{lev}_\tau \|L \cdot \|}$, $V_1 := \mathcal{R}(L^*)$, $V_2 := \mathcal{N}(L)$. By assumption we have $U_2 \cap V_2 = \{0\}$. By $\overline{\mathrm{dom}\, \phi} \cap \mathrm{int}(\mathrm{lev}_\tau \|L \cdot \|) \neq \emptyset$ we obtain $\mathrm{dom}\, \Phi \cap \mathrm{lev}_\tau \|L \cdot \| \neq \emptyset$ which implies $\mathrm{lev}_\alpha F \cap \mathrm{lev}_0 G = \mathrm{lev}_\alpha \Phi \cap \mathrm{lev}_\tau \|L \cdot \| \neq \emptyset$ for sufficiently large α. Denoting the unique minimizer of the strictly convex function $\phi = \Phi|_{X_1}$ by $\check{x}$ and setting $\tilde{\alpha} := \phi(\check{x})$ we see that $\mathrm{lev}_{\tilde{\alpha}}(F|_{U_1}) = \mathrm{lev}_{\tilde{\alpha}}(\phi) = \{\check{x}\}$ is nonempty and bounded. Finally we have that $\mathrm{lev}_0(G|_{V_1}) = \{0\}$ is nonempty and bounded. Thus, by Lemma 3.1 ii), $\mathrm{SOL}(P_{1,\tau}) \neq \emptyset$ with corresponding finite minimum. Lemma 2.6 implies $\mathrm{SOL}(D_{1,\tau}) \neq \emptyset$, where the minimum is finite since $\mathrm{dom}(\Phi^*(-L^* \cdot) + \tau\| \cdot \|_*) = \mathrm{dom}\, \Phi^*(-L^* \cdot) \neq \emptyset$. $\qquad\square$

The next lemma states that there are three main areas where $\mathrm{SOL}(P_{1,\tau})$ and $\mathrm{SOL}(P_{2,\lambda})$ are located: either they are completely contained in $\mathrm{argmin}\, \|L \cdot\| = \mathcal{N}(L)$ or $\mathrm{argmin}\, \Phi$, or they are located "between" them, in the sense of $\mathrm{SOL}(P_\bullet) \cap \mathcal{N}(L) = \emptyset$ and $\mathrm{SOL}(P_\bullet) \cap \mathrm{argmin}\, \Phi = \emptyset$. Similar relations hold true for $\mathrm{SOL}(D_\bullet)$.

Lemma 3.5. *Let $\Phi \in \Gamma_0(\mathbb{R}^n)$ be a function fulfilling (13) and let $L \in \mathbb{R}^{m,n}$ so that $X_2 \cap \mathcal{N}(L) = \{0\}$ and $\mathrm{argmin}\, \Phi \cap \mathcal{N}(L) = \emptyset$. Then the values*

$$c := \inf_{x \in \mathrm{argmin}\, \Phi} \|Lx\| = \min_{x \in \mathrm{argmin}\, \Phi} \|Lx\|, \tag{18}$$

$$d := \inf_{p \in \mathrm{argmin}\, \Phi^*(-L^* \cdot)} \|p\|_* = \min_{p \in \mathrm{argmin}\, \Phi^*(-L^* \cdot)} \|p\|_* \tag{19}$$

are positive, where $d := +\infty$ if $\mathrm{argmin}\, \Phi^*(-L^* \cdot) = \emptyset$ and the following relations hold true:

$$\mathrm{SOL}(P_{1,\tau}) \subseteq \mathcal{N}(L), \qquad \mathrm{SOL}(D_{1,\tau}) \subseteq \mathrm{argmin}\, \Phi^*(-L^* \cdot), \qquad \text{if } \tau = 0,$$

$$\left.\begin{array}{l} \mathrm{SOL}(P_{1,\tau}) \cap \mathcal{N}(L) = \emptyset \\ \mathrm{SOL}(P_{1,\tau}) \cap \mathrm{argmin}\, \Phi = \emptyset \end{array}\right\}, \quad \left.\begin{array}{l} \mathrm{SOL}(D_{1,\tau}) \cap \mathrm{argmin}\, \Phi^*(-L^* \cdot) = \emptyset \\ \mathrm{SOL}(D_{1,\tau}) \cap \{0\} = \emptyset \end{array}\right\}, \quad \text{if } \tau \in (0, c),$$

$$\mathrm{SOL}(P_{1,\tau}) \subseteq \mathrm{argmin}\, \Phi, \qquad \mathrm{SOL}(D_{1,\tau}) \subseteq \{0\}, \qquad \text{if } \tau \in [c, +\infty)$$

and

$$\mathrm{SOL}(P_{2,\lambda}) \subseteq \mathcal{N}(L), \qquad\qquad \mathrm{SOL}(D_{2,\lambda}) \subseteq \operatorname{argmin}\Phi^*(-L^*\cdot), \qquad\quad \textit{if } \lambda \in [d,+\infty),$$

$$\left.\begin{cases} \mathrm{SOL}(P_{2,\lambda}) \cap \mathcal{N}(L) = \emptyset \\ \mathrm{SOL}(P_{2,\lambda}) \cap \operatorname{argmin}\Phi = \emptyset \end{cases}\right\}, \quad \left.\begin{cases} \mathrm{SOL}(D_{2,\lambda}) \cap \operatorname{argmin}\Phi^*(-L^*\cdot) = \emptyset \\ \mathrm{SOL}(D_{2,\lambda}) \cap \{0\} = \emptyset \end{cases}\right\}, \quad \textit{if } \lambda \in (0,d),$$

$$\mathrm{SOL}(P_{2,\lambda}) \subseteq \operatorname{argmin}\Phi, \qquad\qquad \mathrm{SOL}(D_{2,\lambda}) \subseteq \{0\}, \qquad\qquad\qquad \textit{if } \lambda = 0.$$

Proof. 1. We show that d is a minimum if $\operatorname{argmin}\Phi^*(-L^*\cdot) \neq \emptyset$. Let $p_0 \in \operatorname{argmin}\Phi^*(-L^*\cdot)$ and set $r := \|p_0\|_*$. The set $\operatorname{argmin}\Phi^*(-L^*\cdot)$ is closed so that $C := \operatorname{argmin}\Phi^*(-L^*\cdot) \cap B_r$ is a nonempty compact set, which must provide a minimizer $\check{p} \in \operatorname{argmin}\Phi^*(-L^*\cdot)$ for the continuous function $\|\cdot\|_*|_C$. Clearly we also have $\|\check{p}\|_* = \inf_{p \in \operatorname{argmin}\Phi^*(-L^*\cdot)} \|p\|_*$.

2. We show that c is a minimum, i.e., that the function $\iota_{\operatorname{argmin}\Phi} + \|L\cdot\|$ attains its minimum. Denoting the unique minimizer of ϕ by $\check{x}_1$ we obtain $\operatorname{argmin}\Phi = \{\check{x}_1\} \oplus X_2$. Now the assertion follows by Lemma 3.1 ii) with $F := \iota_{\operatorname{argmin}\Phi}$, $U_1 := X_1$, $U_2 := X_2$ and $G := \|L\cdot\|$, $V_1 := \mathcal{R}(L^*)$, $V_2 := \mathcal{N}(L)$. Concerning the assumption of the lemma note that $\mathrm{lev}_\alpha F \cap \mathrm{lev}_\beta G \neq \emptyset$ for $\alpha = \beta := G(\check{x}_1)$. Further, for $\tilde{\alpha} := F(\check{x}_1) = 0$, the set $\mathrm{lev}_{\tilde{\alpha}}(F|_{U_1}) = \{\check{x}_1\}$ is nonempty and bounded. Finally $\mathrm{lev}_{\tilde{\beta}}(G|_{V_1})$ is nonempty and bounded for any $\tilde{\beta} \geq 0$.

3. Next we prove $c > 0$ and $d > 0$. By definition we have $c \geq 0$ and $d \geq 0$. By

$$c = 0 \quad \Leftrightarrow \quad \min_{x \in \operatorname{argmin}\Phi} \|Lx\| = 0 \quad \Leftrightarrow \quad \operatorname{argmin}\Phi \cap \mathcal{N}(L) \neq \emptyset$$

and assumption we see that $c > 0$. Concerning d we obtain by Fermat's rule and the chain rule which can be applied since $0 \in \mathrm{ri}(\operatorname{dom}\Phi^*)$ that

$$d = 0 \quad \Leftrightarrow \quad 0 \in \operatorname{argmin}\Phi^*(-L^*\cdot) \quad \Leftrightarrow \quad 0 \in -L\partial\Phi^*(-L^*0)$$

Thus there exists $x \in \Phi^*(-L^*0)$ such that $0 = -Lx$. Since $x \in \partial\Phi^*(x^*) \Leftrightarrow x^* \in \partial\Phi(x)$, see [48, Corollary 23.5.1] this is equivalent to $0 \in \partial\Phi(x)$ and $x \in \mathcal{N}(L)$. Thus, by Fermat's rule, $x \in \operatorname{argmin}\Phi \cap \mathcal{N}(L)$ which contradicts our assumption on these sets.

4. For the technical proof of the location properties of the solution sets we refer to [18]. $\square$

The following theorem specifies the relations between $(P_{1,\tau})$, $(P_{2,\lambda})$, $(D_{1,\tau})$ and $(D_{2,\lambda})$. We will see that for every $\tau \in (0,c)$ there exists a *uniquely determined* λ such that the *solution sets* of $(P_{1,\tau})$ and $(P_{2,\lambda})$ coincide. Note that by Remark 2.5 this is not the case for general functions $\Phi, \Psi \in \Gamma_0(\mathbb{R}^n)$. Moreover, we want to determine for given τ, the value λ such that $(P_{2,\lambda})$ has the same solutions as $(P_{1,\tau})$. Here Theorem 2.4 i) was not constructive. By $\mathrm{gr}\,g$ we denote the graph of the function g.

Theorem 3.6. *Let $\Phi \in \Gamma_0(\mathbb{R}^n)$ be of the form (13) and let $L \in \mathbb{R}^{m,n}$ such that $X_2 \cap \mathcal{N}(L) = \{0\}$ and $\operatorname{argmin}\Phi \cap \mathcal{N}(L) = \emptyset$. Define c by (18) and d by (19). Then, for $\tau \in (0,c)$ and $\lambda \in (0,d)$, the problems $(P_{1,\tau})$, $(P_{2,\lambda})$, $(D_{1,\tau})$, $(D_{2,\lambda})$ have solutions with finite minima. Further there exists a bijective mapping $g : (0,c) \to (0,d)$ such that for $\tau \in (0,c)$ and $\lambda \in (0,d)$ we have*

$$\begin{cases} \mathrm{SOL}(P_{1,\tau}) = \mathrm{SOL}(P_{2,\lambda}) \\ \mathrm{SOL}(D_{1,\tau}) = \mathrm{SOL}(D_{2,\lambda}) \end{cases} \quad \textit{if} \quad (\tau,\lambda) \in \mathrm{gr}\,g \qquad\qquad (20)$$

and for $\tau \in (0, c)$, $\lambda \in [0, +\infty)$ or $\lambda \in (0, d)$, $\tau \in [0, +\infty)$,

$$\left. \begin{array}{l} \mathrm{SOL}(P_{1,\tau}) \cap \mathrm{SOL}(P_{2,\lambda}) = \emptyset \\ \mathrm{SOL}(D_{1,\tau}) \cap \mathrm{SOL}(D_{2,\lambda}) = \emptyset \end{array} \right\} \quad if \quad (\tau, \lambda) \notin \mathrm{gr}\, g.$$

For $(\tau, \lambda) \in \mathrm{gr}\, g$ any solutions $\hat{x}$ and $\hat{p}$ of the primal and dual problems, resp., fulfill

$$\tau = \|L\hat{x}\| \quad and \quad \lambda = \|\hat{p}\|_*.$$

Proof. First we note that the requirements i) - iv) of Lemma 2.6 are fulfilled. Further Theorem 3.4 ensures that for $\tau \in (0, c)$ and $\lambda \in (0, d)$ all sets $\mathrm{SOL}(P_{1,\tau})$, $\mathrm{SOL}(P_{2,\lambda})$, $\mathrm{SOL}(D_{1,\tau})$, $\mathrm{SOL}(D_{2,\lambda})$ are not empty with corresponding finite minima.

The core of the proof consists of two main steps: in the first step we construct mappings $g : (0, c) \to (0, d)$ and $f : (0, d) \to (0, c)$ such that

$$\forall \tau \in (0, c) : \left. \begin{array}{l} \mathrm{SOL}(P_{1,\tau}) \subseteq \mathrm{SOL}(P_{2,g(\tau)}) \\ \mathrm{SOL}(D_{1,\tau}) \subseteq \mathrm{SOL}(D_{2,g(\tau)}) \end{array} \right\}, \tag{21}$$

$$\forall \lambda \in (0, d) : \left. \begin{array}{l} \mathrm{SOL}(P_{2,\lambda}) \subseteq \mathrm{SOL}(P_{1,f(\lambda)}) \\ \mathrm{SOL}(D_{2,\lambda}) \subseteq \mathrm{SOL}(D_{1,f(\lambda)}) \end{array} \right\}. \tag{22}$$

In the second step we verify that $f \circ g = \mathrm{id}_{(0,c)}$ and $g \circ f = \mathrm{id}_{(0,d)}$ so that g is bijective and (21) and (22) actually hold true with equality. Finally, we deal in a third part with $(\tau, \lambda) \notin \mathrm{gr} g$.
1. First we show that for all $\hat{x} \in \mathbb{R}^n \setminus \mathcal{N}(L)$, $\hat{p} \in \mathbb{R}^m \setminus \{0\}$ and for all $\lambda, \tau > 0$ the following equivalence holds true:

$$\left. \begin{array}{r} \hat{x} \in \mathrm{SOL}(P_{1,\tau}), \\ \hat{p} \in \mathrm{SOL}(D_{1,\tau}), \\ \lambda = \|\hat{p}\|_* \end{array} \right\} \Leftrightarrow \left. \begin{array}{r} \hat{x} \in \mathrm{SOL}(P_{2,\lambda}), \\ \hat{p} \in \mathrm{SOL}(D_{2,\lambda}), \\ \tau = \|L\hat{x}\| \end{array} \right\}. \tag{23}$$

On the one hand, we obtain by Lemma 2.6 with $\Psi := \Psi_1 = \iota_{\mathrm{lev}_1 \|\cdot\|}$ and $\mu := \tau^{-1}$ and the Fenchel equality [48, Theorem 23.5] the equivalences

$$\hat{x} \in \mathrm{SOL}(P_{1,\tau}), \ \hat{p} \in \mathrm{SOL}(D_{1,\tau})$$
$$\Leftrightarrow \tau \hat{p} \in \partial \Psi_1(\tau^{-1} L\hat{x}), \qquad\qquad - L^* \hat{p} = \nabla \Phi(\hat{x})$$
$$\Leftrightarrow \Psi_1(\tau^{-1} L\hat{x}) + \Psi_1^*(\tau \hat{p}) = \langle \tau^{-1} L\hat{x}, \tau \hat{p} \rangle, \ - L^* \hat{p} = \nabla \Phi(\hat{x})$$
$$\Leftrightarrow \|L\hat{x}\| \leq \tau, \ \tau \|\hat{p}\|_* = \langle L\hat{x}, \hat{p} \rangle, \qquad - L^* \hat{p} = \nabla \Phi(\hat{x}),$$
$$\Leftrightarrow \|L\hat{x}\| = \tau, \ \tau \|\hat{p}\|_* = \langle L\hat{x}, \hat{p} \rangle, \qquad - L^* \hat{p} = \nabla \Phi(\hat{x}),$$
$$\Leftrightarrow \|L\hat{x}\| = \tau, \ \|L\hat{x}\| \|\hat{p}\|_* = \langle L\hat{x}, \hat{p} \rangle, \quad - L^* \hat{p} = \nabla \Phi(\hat{x}),$$

On the other hand, we get similarly for $\lambda > 0$ the equivalence

$$\hat{x} \in \mathrm{SOL}(P_{2,\lambda}), \ \hat{p} \in \mathrm{SOL}(D_{2,\lambda})$$
$$\Leftrightarrow \|\hat{p}\|_* = \lambda, \ \|L\hat{x}\| \|\hat{p}\|_* = \langle L\hat{x}, \hat{p} \rangle, \ - L^* \hat{p} = \nabla \Phi(\hat{x}).$$

Consequently, (23) holds true.
Now we introduce the function g on $(0, c)$ by

$$g(\tau) := \|\hat{p}\|_*$$

13

for any $\hat{p} \in \mathrm{SOL}(D_{1,\tau})$. Since $\mathrm{SOL}(D_{1,\tau}) \neq \emptyset$ and Lemma 3.2 or rather Remark 3.3 ensures $\|\hat{p}\|_* = \|\hat{q}\|_*$ for any other $\hat{q} \in \mathrm{SOL}(D_{1,\tau})$ this function is well-defined. Further, Lemma 3.5 assures for $\tau \in (0, c)$ that $\mathrm{SOL}(P_{1,\tau}) \cap \mathcal{N}(L) = \emptyset$, $\mathrm{SOL}(D_{1,\tau}) \cap \{0\} = \emptyset$ and $\mathrm{SOL}(D_{1,\tau}) \cap \mathrm{argmin} \, \Phi^*(-L^*\cdot) = \emptyset$, so that for all $\hat{x} \in \mathrm{SOL}(P_{1,\tau})$, $\hat{p} \in \mathrm{SOL}(D_{1,\tau})$ the relations

$$\|L\hat{x}\| > 0, \qquad\qquad \|\hat{p}\|_* > 0, \qquad\qquad \|\hat{p}\|_* < d \qquad\qquad (24)$$

hold true. To verify the last inequality note that Theorem 2.4 ii) ensures $\hat{p} \in \mathrm{SOL}(D_{2,\|\hat{p}\|_*})$. If $\|\hat{p}\|_* \geq d$, then Lemma 3.5 would imply $\hat{p} \in \mathrm{SOL}(D_{2,\|\hat{p}\|}) \subseteq \mathrm{argmin} \, \Phi^*(-L^*\cdot)$ which results in the contradiction $\hat{p} \in \mathrm{SOL}(D_{1,\tau}) \cap \mathrm{argmin} \, \Phi^*(-L^*\cdot) = \emptyset$. By the second and third inequality in (24) we see that $g : (0, c) \to (0, d)$. The inclusions (21) follow from (23) which can be applied by the first and second inequality in (24).

Similarly, we can prove that the function f on $(0, d)$ defined by

$$f(\lambda) := \|L\hat{x}\|$$

for any $\hat{x} \in \mathrm{SOL}(P_{2,\lambda})$ is well-defined, maps into $(0, c)$ and gives rise to (22).

2. First we prove that

$$\mathrm{SOL}(P_{1,\tau}) \cap \mathrm{SOL}(P_{1,\tau'}) = \emptyset, \qquad\qquad (25)$$

$$\mathrm{SOL}(D_{2,\lambda}) \cap \mathrm{SOL}(D_{2,\lambda'}) = \emptyset \qquad\qquad (26)$$

for all distinct $\tau, \tau' \in (0, c)$ and all distinct $\lambda, \lambda' \in (0, d)$, respectively. Assume that there exist $\lambda, \lambda' \in (0, d)$ with $\lambda < \lambda'$ and $\hat{p} \in \mathrm{SOL}(D_{2,\lambda}) \cap \mathrm{SOL}(D_{2,\lambda'})$. Then $\|\hat{p}\|_* \leq \lambda < \lambda'$ and $\hat{p} \in \mathrm{argmin}\{\Phi^*(-L^*\cdot) \text{ subject to } \|\hat{p}\|_* \leq \lambda'\}$ implies that $\hat{p}$ is a local minimizer of $\Phi^*(-L^*\cdot)$ and hence, by the convexity of $\Phi^*(-L^*\cdot)$, a global minimizer of $\hat{p} \in \mathrm{argmin} \, \Phi^*(-L^*\cdot)$. This contradicts $\mathrm{argmin} \, \Phi^*(-L^*\cdot) \cap \mathrm{SOL}(D_{2,\lambda'}) = \emptyset$ in Lemma 3.5. Similarly, we can show (25).

Next we prove the bijectivity of $g : (0, c) \to (0, d)$ by showing $f \circ g = \mathrm{id}_{(0,c)}$ (injectivity) and $g \circ f = \mathrm{id}_{(0,d)}$ (surjectivity). Let $\tau \in (0, c)$ be arbitrarily chosen and set $\tau' := f(g(\tau))$. Then (21) and (22) with $\lambda = g(\tau)$ yields

$$\mathrm{SOL}(P_{1,\tau}) \subseteq \mathrm{SOL}(P_{2,g(\tau)}) \subseteq \mathrm{SOL}(P_{1,\tau'}),$$

$$\mathrm{SOL}(D_{1,\tau}) \subseteq \mathrm{SOL}(D_{2,g(\tau)}) \subseteq \mathrm{SOL}(D_{1,\tau'}).$$

Since $\mathrm{SOL}(P_{1,\tau}) \neq \emptyset$ we must have $\tau = \tau'$ in order to avoid a contradiction to (25). Similarly we can prove for an arbitrarily chosen $\lambda \in (0, d)$ and $\lambda' := g(f(\lambda))$ that $\lambda = \lambda'$.

3. It remains to show $\mathrm{SOL}(P_{1,\tau}) \cap \mathrm{SOL}(P_{2,\lambda}) = \emptyset$ and $\mathrm{SOL}(D_{1,\tau}) \cap \mathrm{SOL}(D_{2,\lambda}) = \emptyset$ for $(\tau, \lambda) \in [(0, c) \times [0, +\infty)] \cup [[0, +\infty) \times (0, d)]$ with $(\tau, \lambda) \notin \mathrm{gr} \, g$. By Lemma 3.5 we can restrict us to those $(\tau, \lambda) \notin \mathrm{gr} \, g$ with $(\tau, \lambda) \in (0, c) \times (0, d)$. For such τ, λ we have $\tau \neq g^{-1}(\lambda)$ and $\lambda \neq g(\tau)$. Therefore we have by (25) and (26) that $\mathrm{SOL}(P_{1,\tau}) \cap \mathrm{SOL}(P_{1,g^{-1}(\lambda)}) = \emptyset$ and $\mathrm{SOL}(D_{2,\lambda}) \cap \mathrm{SOL}(D_{2,g(\tau)}) = \emptyset$. By (20) we can replace $\mathrm{SOL}(P_{1,g^{-1}(\lambda)})$ by $\mathrm{SOL}(P_{2,\lambda})$ and $\mathrm{SOL}(D_{2,g(\tau)})$ by $\mathrm{SOL}(D_{1,\tau})$ and we are done. $\qquad\square$

By the following corollary the function g is decreasing.

Corollary 3.7. *Let the assumptions of Theorem 3.6 be fulfilled. Then the bijection $g : (0, c) \to (0, d)$ is strictly decreasing and continuous.*

Proof. Since decreasing bijections between open intervals are strict decreasing and continuous we need only to show that $f = g^{-1} : (0, d) \to (0, c)$ is decreasing. Let $0 < \lambda_1 < \lambda_2 < d$ and

14

$\hat{x}_i \in \mathrm{SOL}(P_{2,\lambda_i})$, $i = 1, 2$. Then $f(\lambda_i) = \|L\hat{x}_i\|$, $i = 1, 2$. Assuming that $f(\lambda_1) < f(\lambda_2)$ we obtain with $\lambda_2 = \lambda_1 + \varepsilon$, $\varepsilon > 0$ the contradiction

$$
\begin{aligned}
\Phi(\hat{x}_2) + \lambda_2 \|L\hat{x}_2\| &= \Phi(\hat{x}_2) + \lambda_1 \|L\hat{x}_2\| + \varepsilon\|L\hat{x}_2\| \\
&\geq \Phi(\hat{x}_1) + \lambda_1 \|L\hat{x}_1\| + \varepsilon f(\lambda_2) \\
&> \Phi(\hat{x}_1) + \lambda_1 \|L\hat{x}_1\| + \varepsilon f(\lambda_1) \\
&= \Phi(\hat{x}_1) + \lambda_2 \|L\hat{x}_1\|. \qquad \square
\end{aligned}
$$

Remark 3.8. *The function g is in general neither differentiable nor convex as the following example shows: The strictly convex function Φ, given by*

$$
\Phi(x) := \begin{cases} (x-4)^2 & \text{for } x \leq 2 \\ 2(x-3)^2 + 2 & \text{for } x > 2 \end{cases}
$$

has exactly one minimizer, namely $x_0 = 3$. For $\tau \in (0, x_0)$ we have

$$
\operatorname*{argmin}_{x \in \mathbb{R}} \{\Phi(x) \text{ subject to } |x| \leq \tau\} = \{\tau\} =: \{\hat{x}\}.
$$

By theorem 3.6 there is only one $\lambda \geq 0$ with $\operatorname{argmin}(\Phi(\cdot) + \lambda|\cdot|) = \{\tau\}$, namely $\lambda = g(\tau)$. Fermat's rule gives $0 \in \partial(\Phi(\cdot) + g(\tau)|\cdot|)|_\tau = (\{\Phi'(\cdot)\} + g(\tau)\partial|\cdot|)|_\tau = \{\Phi'(\tau) + g(\tau)\}$, so that

$$
g(\tau) = -\Phi'(\tau) = \begin{cases} 2(4-\tau) & \text{for } 0 < \tau < 2 \\ 4 & \text{for } \tau = 2 \\ 4(3-\tau) & \text{for } 2 < \tau < x_0 \end{cases}
$$

Obviously g is neither differentiable nor convex.

4 Problems from Image Processing

In our imaging applications the function $\Phi \in \Gamma_0(\mathbb{R}^n)$ has the form

$$
\Phi(x) = \varphi(Hx), \tag{27}
$$

where φ is a shifted squared Euclidean norm or the *I*-divergence with $\mathrm{dom}\,\varphi = \mathbb{R}^n$ and $\mathrm{dom}\,\varphi = \mathbb{R}^n_{>0}$, respectively. Thus φ fulfills

- $\mathrm{dom}\,\varphi$ is an open subset of $\mathbb{R}^n$ with $0 \in \overline{\mathrm{dom}\,\varphi}$
- $\varphi \in \Gamma_0(\mathbb{R}^n)$ is strictly convex and essentially smooth.
- φ has a minimizer.

Moreover, H must be chosen such that $\mathcal{R}(H) \cap \mathrm{dom}\,\varphi \neq \emptyset$ in order to guarantee $\Phi \not\equiv +\infty$. Noting that $H|_{\mathcal{R}(H^*)} : \mathcal{R}(H^*) \to \mathcal{R}(H)$ is bijective, it is not hard to check that Φ is of the form (13) with $X_1 := \mathcal{R}(H^*)$, $X_2 := \mathcal{N}(H)$ and $\phi(x_1) := \varphi(Hx_1)$ for $x_1 \in X_1$.

For a simple matrix-vector notation we assume images to be column-wise reshaped as a vector. Thus, having an image $x \in \mathbb{R}^{N,M}$ we use the vectorized image $\mathrm{vec}(x) \in \mathbb{R}^n$ with $n = MN$. We retain the notation x for both the original and the reshaped image. We will consider four different models, where the mixed ℓ_1 and ℓ_∞ norms $\|\,|\cdot|\,\|_1$ and $\|\,|\cdot|\,\|_\infty$ are used. These

norms are the ℓ_1 and ℓ_∞ norms applied to the vector $|p| \in \mathbb{R}^n$ which follows from the actual vector $p \in \mathbb{R}^m$ with $m = \kappa n$, $\kappa \in \mathbb{N}$ by

$$|p| := \left(\|\boldsymbol{p}_j\|_2 \right)_{j=1}^n, \quad \boldsymbol{p}_j = (p_{kn+j})_{k=0}^{\kappa-1}. \tag{28}$$

For $\kappa = 1$ these are the ordinary ℓ_1 and ℓ_∞ norms. The mixed ℓ_1 and ℓ_∞ norms are dual to each other. Moreover, they have the useful property that the orthogonal projections onto the balls $B_{\alpha,r} := \{p \in \mathbb{R}^m : \| |p| \|_\alpha \le r\}$, $\alpha \in \{1, \infty\}$ can be computed in a simple way:

- The *projection onto* $B_{\infty,r}$ of $p \in \mathbb{R}^m$ is given separately for the subvectors $\boldsymbol{p}_j$ by

$$P_{B_{\infty,r}}(\boldsymbol{p}_j) := \min \left\{ \frac{r}{\|\boldsymbol{p}_j\|_2}, 1 \right\} \boldsymbol{p}_j, \quad j = 1, \ldots, n.$$

This projection is related to the *(coupled) soft shrinkage operator* S_r by $S_r = I - P_{B_{\infty,r}}$, i.e.,

$$S_r(\boldsymbol{p}_j) = \max \left\{ 1 - \frac{r}{\|\boldsymbol{p}_j\|_2}, 0 \right\} \boldsymbol{p}_j, \quad j = 1, \ldots, n.$$

- The *projection onto* $B_{1,r}$ is given by

$$P_{B_{1,r}}(p) := \begin{cases} p & \text{if} \| |p| \|_1 \le r, \\ S_\mu(p) & \text{otherwise} \end{cases}$$

with $\mu := \frac{\|\boldsymbol{p}_{\pi(1)}\|_2 + \ldots + \|\boldsymbol{p}_{\pi(\nu)}\|_2 - r}{\nu}$, where $\|\boldsymbol{p}_{\pi(1)}\| \ge \ldots \ge \|\boldsymbol{p}_{\pi(n)}\|_2 \ge 0$ are the sorted norms of the vectors $\boldsymbol{p}_j$ and $\nu \le n$ is the largest index such that $\|p_{\pi(\nu)}\|_2 > 0$ and

$$\frac{\|\boldsymbol{p}_{\pi(1)}\|_2 + \ldots + \|\boldsymbol{p}_{\pi(\nu)}\|_2 - r}{\nu} \le \|\boldsymbol{p}_{\pi(\nu)}\|_2,$$

see [24, 54].

Next we consider four models.

4.1 Joint Sparsity Problems

The following task arises in (joint) sparsity problems in the presence of white Gaussian noise [42]. Here one is looking for a minimizer $\hat{p}$ of one of the problems

$$(D_{2,\lambda}) \quad \operatorname*{argmin}_{p \in \mathbb{R}^m} \left\{ \frac{1}{2} \|Ap - b\|_2^2 \quad \text{subject to} \quad \| |p| \|_1 \le \lambda \right\},$$

$$(D_{1,\tau}) \quad \operatorname*{argmin}_{p \in \mathbb{R}^m} \left\{ \frac{1}{2} \|Ap - b\|_2^2 + \tau \| |p| \|_1 \right\}$$

where $A \in \mathbb{R}^{n,m}$, $n \ll m$ and $0 \ne b \in \mathcal{R}(A)$. We are interested in obtaining τ for given λ. Problem $(D_{2,\lambda})$ is also known as LASSO (least absolute shrinkage and selection operator) [57] and can be also tackled by basis pursuit [16, 60]. The above problems can be seen as dual problems of the primal problems

$$(P_{2,\lambda}) \quad \operatorname*{argmin}_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} \|x - b\|_2^2 + \lambda \| |A^*x| \|_\infty \right\},$$

$$(P_{1,\tau}) \quad \operatorname*{argmin}_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} \|x - b\|_2^2 \quad \text{subject to} \quad \| |A^*x| \|_\infty \le \tau \right\}.$$

These problems fit into the setting of the previous section with $L := A^*$, $H := I$ and the strictly convex function $\Phi(x) := \frac{1}{2}\|x - b\|_2^2$ which has $\Phi^*(p) = \frac{1}{2}\|p + b\|_2^2 - \frac{1}{2}\|b\|_2^2$ as dual function. By (12) the solutions $\hat{x}$ and $\hat{p}$ of the primal and dual problems, resp., are related by

$$\hat{x} = b - A\hat{p}. \tag{29}$$

Now Theorem 3.6 implies the following corollary.

Corollary 4.1. *Let*

$$c := \| \, |A^*b| \, \|_\infty \quad \text{and} \quad d := \min_{Ap=b} \| \, |p| \, \|_1. \tag{30}$$

Then, for $\lambda \in (0, d)$, a solution $\hat{p}$ of $(D_{2,\lambda})$ is also a solution of $(D_{1,\tau})$ iff

$$\tau = \| \, | \underbrace{A^*(b - A\hat{p})}_{=:\hat{q}} | \, \|_\infty = \|\hat{\boldsymbol{q}}_j\|_2 \quad \text{for all } j \text{ with } \hat{\boldsymbol{p}}_j \neq 0$$

where $\hat{\boldsymbol{q}}_j := (\hat{q}_{kn+j})_{k=0}^{\kappa-1}$ as in (28).

Relations between the above problems were also considered in another way for $\kappa = 1$ in [21], see also [40]. The problem of finding d or better a minimizer of the right-hand side problem in (30) appears, e.g., in compressed sensing [23]

Proof. By Theorem 3.6 and (29), we know that $\mathrm{SOL}(D_{1,\tau}) = \mathrm{SOL}(D_{2,\lambda})$ if and only if

$$\tau = \| \, |A^*\hat{x}| \, \|_\infty = \| \, |A^*(b - A\hat{p})| \, \|_\infty.$$

To prove the rest of the assertion we use that $\hat{p}$ is also a solution of $(D_{1,\tau})$. Solving $(D_{1,\tau})$ with the FBS algorithm described in the Appendix A.1 with $f_1 := \frac{1}{2}\|A \cdot -b\|_2^2$ and $f_2 := \tau\| \, |\cdot| \, \|_1$ for which $\nabla f_1(p) = A^*(Ap - b)$ and $\mathrm{prox}_{\eta f_2} = S_{\eta\tau}$, cf., [20, 62], we see that $\hat{p}$ solves

$$\hat{p} = S_{\eta\tau}(\hat{p} + \eta A^*(b - A\hat{p})) = S_{\eta\tau}(\hat{p} + \eta\hat{q})$$

for any $\eta \in (0, 2/\|A\|_2^2)$. By definition of the coupled soft thresholding operator we distinguish two cases: If $\|\hat{\boldsymbol{p}}_j + \eta\hat{\boldsymbol{q}}_j\|_2 \leq \eta\tau$ for some η then $\hat{\boldsymbol{p}}_j = 0$. If $\|\hat{\boldsymbol{p}}_j + \eta\hat{\boldsymbol{q}}_j\|_2 > \eta\tau$ for all η, then

$$\hat{\boldsymbol{p}}_j = \hat{\boldsymbol{p}}_j + \eta\hat{\boldsymbol{q}}_j - \eta\tau \frac{\hat{\boldsymbol{p}}_j + \eta\hat{\boldsymbol{q}}_j}{\|\hat{\boldsymbol{p}}_j + \eta\hat{\boldsymbol{q}}_j\|_2},$$

$$\hat{\boldsymbol{q}}_j = \tau \frac{\hat{\boldsymbol{p}}_j + \eta\hat{\boldsymbol{q}}_j}{\|\hat{\boldsymbol{p}}_j + \eta\hat{\boldsymbol{q}}_j\|_2}.$$

Since $\eta > 0$ can be chosen arbitrarily small, we see that $\hat{\boldsymbol{q}}_j = \tau \frac{\hat{\boldsymbol{p}}_j}{\|\hat{\boldsymbol{p}}_j\|_2}$. Taking the ℓ_2 norm at both sides, we obtain the rest of the assertion. $\qquad\square$

Example 4.2. *We want to illustrate the graph of the function g^{-1} which maps λ to τ for a sparsity problem with $\kappa = 1$, i.e., for the ordinary ℓ_1 norm. We use the matrix A obtained by taking 200 rows of the (1000,1000) DCT-II matrix and b resulting from a noisy vector Ax, where x has 25 nonzero random entries of dynamical range 100. To find the values τ corresponding to $\lambda \in \{1, 2, 3, \ldots, \lfloor d \rfloor\}$ we have computed a solution $\hat{p}$ of $(D_{2,\lambda})$ for each λ by the FBS algorithm described in Appendix A.1 with $f_1 := \frac{1}{2}\|A \cdot -b\|_2^2$ and $f_2 := \iota_{B_{1,\lambda}}$ for which $\nabla f_1(p) = A^*(Ap - b)$ and $\mathrm{prox}_{\eta f_2} = P_{B_{1,\lambda}}$. We have $d \approx 563$. Since $\|A\|_2 = 1$ we can choose $\eta \in (0, 2)$. Then, by Corollary 4.1, we have set $\tau := \|A^*(b - A\hat{p})\|_\infty$. Fig. 2 shows the corresponding function $g^{-1} : (0, d) \to (0, c)$.*
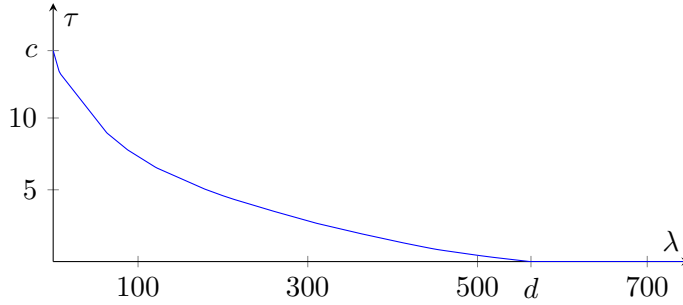
Figure 2: Function g^{-1} from Theorem 3.6 for the described joint sparsity problem.

In [61] problem $(D_{2,\lambda})$ was considered with the ordinary ℓ_1-norm. The authors were interested in the Pareto frontier or rather curve with respect to $\|p\|_1$ and $\|Ap-b\|_2$. Except for endpoints, that Pareto frontier is the graph of the function $\chi : (0,d) \to \mathbb{R}$, given by

$$\chi(\lambda) := \inf_{\|p\|_1 \leq \lambda} \|Ap - b\|_2 = \inf_{\|p\|_1 = \lambda} \|Ap - b\|_2. \tag{31}$$

This function is related to the above $\lambda - \tau$ curve $f = g^{-1}$ by

$$f(\lambda) = -\|Ap_\lambda - b\|_2 \chi'(\lambda), \quad p_\lambda \in \mathrm{SOL}(D_{2,\lambda}).$$

Fitting more to the origins of g, we also have a look at the Pareto frontier with regard to $\|p\|_1$ and $\frac{1}{2}\|Ap-b\|_2^2$. Except for endpoints it is the graph of the function $\psi : (0,d) \to \mathbb{R}$, given by

$$\psi(\lambda) := \inf_{\|p\|_1 \leq \lambda} \frac{1}{2} \|Ap - b\|_2^2 = \frac{1}{2}(\chi(\lambda))^2, \quad \lambda \in (0,d). \tag{32}$$

This function is related to f by

$$f(\lambda) = -\psi'(\lambda). \tag{33}$$

In [61] it was proved that χ is convex and continuously differentiable which holds also true for ψ. Thus, by (33), the function f is continuous and decreasing; this is in agreement with the statement of Corollary 3.7 for the special setting in the current subsection. However, in general the function f is not differentiable. To see this fact we prove the following lemma on the derivative of ψ which also shows that ψ is continuously differentiable.

Lemma 4.3. *For $A \in \mathbb{R}^{m,n}$, $b \in \mathcal{R}(A)\backslash\{0\}$ and $d := \min_{Ap=b}\|p\|_1$, let $\psi : (0,d) \to \mathbb{R}$ be given by (32). Then ψ is differentiable with derivative*

$$\psi'(\lambda) = 2\lambda \, (^1h)\left(\frac{b}{\lambda}\right) - \frac{1}{\lambda}\|b\|_2^2 + \left\langle P_{AB_{1,1}}\left(\frac{b}{\lambda}\right), b\right\rangle, \tag{34}$$

where $^1h(v) := \inf_u\{\frac{1}{2}\|u - v\|_2^2 + \iota_{AB_{1,1}}(u)\}$ denotes the Moreau envelope to parameter 1 of the function $h := \iota_{AB_{1,1}}$ and $P_{AB_{1,1}}$ the orthogonal projector onto $AB_{1,1}$.

Proof. We rewrite ψ as

$$\psi(\lambda) = \inf_{s \in AB_{1,\lambda}} \frac{1}{2}\|s - b\|_2^2 = \lambda^2 \inf_{q \in AB_{1,1}} \frac{1}{2}\left\|q - \frac{b}{\lambda}\right\|_2^2$$

$$= \lambda^2 (^1h)\left(\frac{b}{\lambda}\right).$$

18

Since h is proper, closed and convex its Moreau envelope ${}^1h$ has a Lipschitz continuous gradient which is given by

$$\nabla({}^1h)(v) = v - \text{prox}_h(v) = v - P_{AB_{1,1}}(v),$$

see, e.g., [1, Proposition 12.29]. Using this relation, we obtain

$$
\begin{aligned}
\psi'(\lambda) &= 2\lambda({}^1h)\left(\frac{b}{\lambda}\right) + \lambda^2 \langle \nabla({}^1h)\left(\frac{b}{\lambda}\right), b \rangle \left(-\frac{1}{\lambda^2}\right) \\
&= 2\lambda({}^1h)\left(\frac{b}{\lambda}\right) - \frac{1}{\lambda}\|b\|_2^2 + \langle P_{AB_{1,1}}\left(\frac{b}{\lambda}\right), b \rangle. \qquad \square
\end{aligned}
$$

While the first two summands in (34) are again continuously differentiable for $\lambda \in (0, d)$, the third summand is in general not differentiable as the following example shows.

Example 4.4. *Let $m = n = 2$, $A = I$ and $\frac{1}{2} < b_1 < 1$, $b_2 = 1 - b_1$. Then $d = b_1 + b_2 = 1$ and we obtain by straightforward computation that*

$$P_{B_{1,1}}\left(\frac{b}{\lambda}\right) = \begin{pmatrix} \frac{b_1 - b_2}{2\lambda} + \frac{1}{2} \\ \frac{b_2 - b_1}{2\lambda} + \frac{1}{2} \end{pmatrix} \quad \text{for} \quad \lambda \in (b_1 - b_2, 1)$$

and $P_{B_{1,1}}\left(\frac{b}{\lambda}\right) = (1, 0)^T$ for $\lambda \in (0, b_1 - b_2]$. Consequently,

$$\langle P_{B_{1,1}}\left(\frac{b}{\lambda}\right), b \rangle = \begin{cases} b_1 & \text{for} \quad \lambda \in (0, b_1 - b_2], \\ \frac{b_1 + b_2}{2} + \frac{(b_1 - b_2)^2}{2\lambda} & \text{for} \quad \lambda \in (b_1 - b_2, 1) \end{cases}$$

which is not differentiable at $\lambda = b_1 - b_2 \in (0, 1) = (0, d)$.

4.2 Restoration of Images Corrupted by Gaussian Noise

In this subsection, we are interested in problems of the form

$$(P_{1,\tau}) \qquad \underset{x \in \mathbb{R}^n}{\text{argmin}} \left\{ \frac{1}{2}\|Hx - b\|_2^2 \quad \text{subject to} \quad \| |Lx| \|_1 \leq \tau \right\},$$

$$(P_{2,\lambda}) \qquad \underset{x \in \mathbb{R}^n}{\text{argmin}} \left\{ \frac{1}{2}\|Hx - b\|_2^2 + \lambda \| |Lx| \|_1 \right\},$$

where $L \in \mathbb{R}^{m,n}$, $H \in \mathbb{R}^{n,n}$ and $b := Hx_{\text{orig}} + \eta$ is the transformed original image corrupted by white Gaussian noise. Model $(P_{2,\lambda})$ arises from Bayesian statistical inference, i.e., the MAP (maximum a posteriori) approach, see, e.g., [14, Section 2.4] and [3]. The model incorporates the noise statistics in the data term and some information on the image in the regularizing term. In the literature there appear for example the following settings:

(H1) Denoising with $H = I$.

(H2) Deblurring with an invertible, component–by–component non-negative matrix H which is diagonalizable by the DCT-II transform [47] and fulfills $H^*1_n = 1_n$. Here $1_n \in \mathbb{R}^n$ is the vector with all entries one.

(H3) Inpainting with a diagonal matrix H having only the values zero and one at its diagonal.

Image restoration models frequently use the regularizing term introduced by Rudin, Osher and Fatemi (ROF) [50] or "ROF-like" regularizers as:

(L1) The discrete gradient $L := \nabla = (\partial_x, \partial_y)^{\mathrm{T}}$ which belongs indeed to a discrete ROF model, see [10].

Here $m = 2n$, i.e., $\kappa = 2$.

(L2) The operator associated with the Frobenius norm of the Hessian $L := (\partial_{xx}, \partial_{yy}, \partial_{xy}, \partial_{yx})^{\mathrm{T}}$ in [55] – here $m = 4n$, i.e. $\kappa = 4$ – and some of its relatives, see, e.g., [7, 13, 41, 37].

(L3) Operators L arising from nonlocal means, see [31, 32, 56].

For invertible H, the dual problems read

$$(D_{1,\tau}) \qquad \operatorname*{argmin}_{p \in \mathbb{R}^m} \left\{ \frac{1}{2} \|(H^{-1})^* L^* p - b\|_2^2 + \tau \| \, |p| \, \|_\infty \right\},$$

$$(D_{2,\lambda}) \qquad \operatorname*{argmin}_{p \in \mathbb{R}^m} \left\{ \frac{1}{2} \|(H^{-1})^* L^* p - b\|_2^2 \quad \text{subject to} \quad \| \, |p| \, \|_\infty \le \lambda \right\}.$$

By Remark A.1, for general H, the dual problems can be expressed in their saddle point formulation

$$\operatorname*{argmax}_{p \in \mathbb{R}^m} \min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} \|Hx - b\|_2^2 + \langle p, Lx \rangle - \tau \| \, |p| \, \|_\infty \right\},$$

$$\operatorname*{argmax}_{p \in \mathbb{R}^m} \min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} \|Hx - b\|_2^2 + \langle p, Lx \rangle - \iota_{B_{\infty,\lambda}}(p) \right\}.$$

By Theorem 3.6 we obtain the following corollary.

Corollary 4.5. *Let $L \in \mathbb{R}^{m,n}$ and $H \in \mathbb{R}^{n,n}$ such that $\mathcal{N}(L) \cap \mathcal{N}(H) = \{0\}$. Assume further that there does not exist $x \in \mathcal{N}(L)$ with $H^*(Hx - b) = 0$. Set*

$$c := \min_{H^*(Hx-b)=0} \| \, |Lx| \, \|_1$$

which reduces to $c = \| \, |LH^{-1}b| \, \|_1$ if H is invertible. Then, for $\tau \in (0, c)$, a solution $\hat{x}$ of $(P_{1,\tau})$ is also a solution of $(P_{2,\lambda})$ iff $\lambda = \| \, |\hat{p}| \, \|_\infty$, where $\hat{p}$ is any solution of $(D_{1,\tau})$.

While the model $(P_{2,\lambda})$ was considered in many papers, we have found the treatment of the non-constrained problem $(P_{1,\tau})$ only in [28]. In the following example, we just illustrate the relation between τ and λ for a denoising problem and an inpainting model. For applications in image processing we refer to [28] and the references therein.

Example 4.6. *First, we consider a denoising problem with $H := I$, $L := \nabla$ and b shown in Fig. 3 a). Here $c \approx 795$. In order to find the values λ corresponding to $\tau \in \{1, 10, 20, 30, \ldots, 790, 800\}$ we have computed a solution $\hat{p}$ of $(D_{1,\tau})$ for each τ by the FBS algorithm described in Appendix A.1 with $f_1 := \frac{1}{2}\|L^* \cdot - b\|_2^2$ and $f_2 := \tau \| \, |\cdot| \, \|_\infty$ for which $\nabla f_1(p) = L(L^* p - b)$ and $\operatorname{prox}_{\eta f_2} = I - P_{B_{1,\tau}}$. Since $\|\nabla\|_2^2 < 8$ we can choose $\eta \in (0, 1/4)$ and have used $\eta = 1/8$. Fig. 3 b) shows the result of the curve g .*
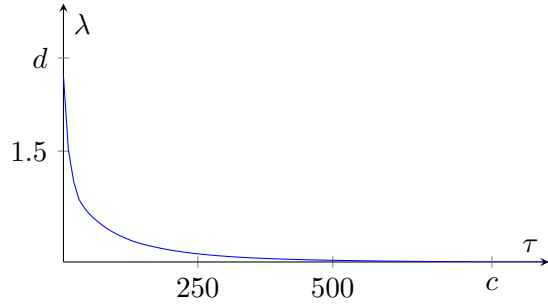
(a) Input b.

(b) Plot of the function g for b shown in Fig. 3 a).

Figure 3: Function g from Theorem 3.6 for the described denoising problem.

Example 4.7. *Next, we deal with an inpainting problem with a diagonal matrix H having zero diagonal entries corresponding to the mask in Fig. 4 (a) and ones otherwise and $L := \nabla$. We do not intend to provide the qualitatively best inpainting model among the large number of methods in this field, but to present a restoration example with a matrix H which is not invertible. Moreover, it may be interesting to see the difference to the tensor inpainting method in Fig. 8. The corrupted image b given by the Fig. 4 (a). For comparison, we denote by x_{orig} the original image and use the abbreviation $TV(x) := \| |\nabla x| \|_1$. Fig. 4 shows the inpainting result by solving $(P_{1,\tau})$ by the scaled ADMM algorithm described in Appendix A.2 with the splitting $Lx = y$ and $\gamma = 10$. Note that the first step of the algorithm requires the solution of a linear system with coefficient matrix $H^*H + \gamma L^*L = H + \gamma L^*L$, $\gamma > 0$ which is invertible since the matrix is irreducible diagonal-dominant. To illustrate the relation between τ and λ we have computed a solution $\hat{p}$ of the dual problem of $(P_{1,\tau})$ with reduced input image* `b(201:300,201:300)` *from Fig. 4(a) for each $\tau \in \{1, 10, 20, 30, \ldots, 990, 1000, 1500, 2000, \ldots, 9500, 10000\}$. Here we have $c \approx 95200$. Note that $\hat{p} = \gamma \hat{q}$ with $\hat{q}$ computed by the scaled ADMM algorithm. Fig. 5 shows the graph of the function g.*

4.3 Restoration of Images Corrupted by Poisson Noise

Next we are interested in restoration problems for images corrupted by Poisson noise, more precisely we focus on the denoising and deblurring problems with H as in (H1) and (H2). Note that in both cases H is invertible. Then b is just the transformed original image corrupted by Poisson noise. In particular we have that $b > 0$. In the MAP approach we have to replace the squared ℓ_2 data term from the previous subsection by the discrete *I-divergence*

$$I(b,x) := \begin{cases} \langle 1_n, b \log \frac{b}{x} - b + x \rangle & \text{if } x > 0, \\ +\infty & \text{otherwise.} \end{cases}$$

The *I*-divergence is also known as *generalized Kullback-Leibler divergence* and is the Bregman distance of the Boltzman-Shannon entropy. The Fenchel conjugate of $\varphi(x) := I(b, \cdot)$ is given by

$$\varphi^*(p) = \begin{cases} -\langle b, \log(1_n - p) \rangle & \text{if } p < 1_n, \\ +\infty & \text{otherwise.} \end{cases}$$

21

(a) Corrupted image b.

(b) $\hat{x}$ with $TV(\hat{x}) \approx TV(x_{\mathrm{orig}})$.

(c) $\hat{x}$ with $TV(\hat{x}) \approx \frac{1}{2}TV(x_{\mathrm{orig}})$.

(d) $\hat{x}$ with $TV(\hat{x}) \approx \frac{1}{4}TV(x_{\mathrm{orig}})$.

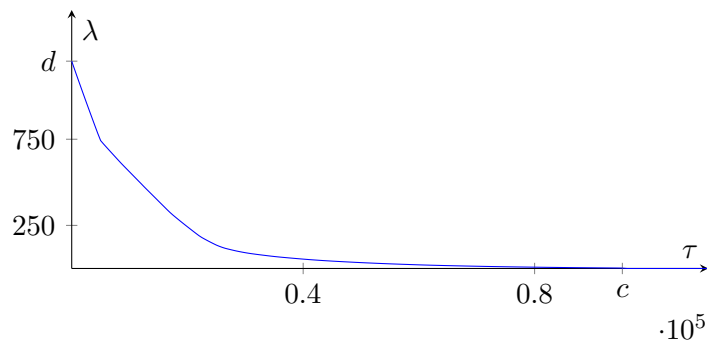Figure 4: Solutions of the described TV-inpainting problem $(P_{1,\tau})$ for various values of τ.

Figure 5: Function g from Theorem 3.6 for the described TV-inpainting problem.

Using the same regularizers (L1) - (L3) as in the previous subsection, the minimization problems read

22

$$(P_{1,\tau}) \qquad \underset{Hx>0}{\operatorname{argmin}} \left\{ \langle 1_n, Hx - b\log(Hx) \rangle \quad \text{subject to} \quad \||Lx|\|_1 \le \tau \right\},$$

$$(P_{2,\lambda}) \qquad \underset{Hx>0}{\operatorname{argmin}} \{ \langle 1_n, Hx - b\log(Hx) \rangle \ + \ \lambda\||Lx|\|_1 \}.$$

Model $(P_{2,\lambda})$ was considered, e.g., in [29, 53, 56]. So far we have not seen model $(P_{1,\tau})$ in the literature. Note that very recently, in [17], the authors have dealt with a constrained minimization model which data term consists of a mixed norm instead of the I-divergence. Interestingly, they have used projections onto epigraphs for the solution. For invertible H the dual problems are given by

$$(D_{1,\tau}) \qquad \underset{1_n+(H^{-1})^*L^*p>0}{\operatorname{argmin}} \left\{ -\langle b, \log(1_n + (H^{-1})^*L^*p) \rangle \ + \ \tau\||p|\|_\infty \right\},$$

$$(D_{2,\lambda}) \qquad \underset{1_n+(H^{-1})^*L^*p>0}{\operatorname{argmin}} \left\{ -\langle b, \log(1_n + (H^{-1})^*L^*p) \rangle \quad \text{subject to} \quad \||p|\|_\infty \le \lambda \right\},$$

and by (12) the solutions $\hat{x}$ and $\hat{p}$ of the primal and dual problems, resp., are related by

$$\hat{x} = H^{-1}\left(\frac{b}{1_n + (H^{-1})^*L^*\hat{p}} \right).$$

By Theorem 3.6 we obtain the following relation.

Corollary 4.8. *Let $L \in \mathbb{R}^{m,n}$ and $H \in \mathbb{R}^{n,n}$ be given by (H2) and $c := \||LH^{-1}b|\|_1$. Assume further that $H^{-1}b \notin \mathcal{N}(L)$. Then, for $\tau \in (0,c)$, a solution $\hat{x}$ of $(P_{1,\tau})$ is also a solution of $(P_{2,\lambda})$ iff $\lambda = \||\hat{p}|\|_\infty$, where $\hat{p}$ is any solution of $(D_{1,\tau})$.*

Since we are not aware of the application of model $(P_{1,\tau})$ for image processing tasks in the literature, we provide an image processing example together with the illustration of the function g which maps τ to λ. Of course model $(P_{1,\tau})$ makes sense if we have some a priori knowledge, e.g., on the (discrete) total variation of the original image rather than on the noise level.

Example 4.9. *Fig. 6 shows deblurring results by solving $(P_{1,\tau})$ for different values of τ. Remember that $\tau = \||L\hat{x}|\|_1$ for the solution $\hat{x}$. To this end, we have blurred the original image x_{orig} by a Gaussian kernel with standard deviation $\sigma = 1$ and mirror (Neumann) boundary conditions which results in the operator H. Then we have corrupted the blurred image by Poisson noise by scaling the image by the factor $3000b_{\max}/10^{12}$, where $b_{\max}$ is the maximal value of the blurred input image and using the MATLAB function* `imnoise(3000,'poisson')`. *Then we have restored the image by applying the scaled ADMM for the multi-splitting in (41) with $\gamma = \tau/30$. Note that for L in (L1) and (L2), the linear system of equations arising in the first step of the ADMM algorithm can be solved in a fast way by the DCT-II. In this example we have used $L := \nabla$ and the abbreviation $TV(x) := \||\nabla x|\|_1$.*
Next we have computed the corresponding curve g which maps τ to λ for the above image part $b(40 : 80, 100 : 140)$ and $\tau \in \{1, 10, 60, 110, \ldots, 910, 960, 1500, 2000, 2500, \ldots 270000\}$. We have used the ADMM algorithm for the multi-splitting (41), where we have set $\hat{p} := \gamma\hat{q}_2$ by Remark A.1. The curve is plotted in Fig. 7.

(a) Noisy and blurry image b.

(b) $\hat{x}$ with $TV(\hat{x}) \approx \frac{1}{2}TV(x_{\mathrm{orig}})$.

(c) $\hat{x}$ with $TV(\hat{x}) \approx \frac{1}{4}TV(x_{\mathrm{orig}})$.

(d) $\hat{x}$ with $TV(\hat{x}) \approx \frac{1}{8}TV(x_{\mathrm{orig}})$.

Figure 6: Solutions of the described deblurring problem $(P_{1,\tau})$ with Gaussian blur and Poisson noise for various values of τ.
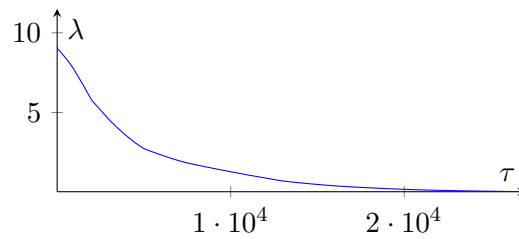

Figure 7: Function g from Theorem 3.6 for the described deblurring problem with Poisson noise for different scales of τ.

4.4 Matrix Completion and Tensor Inpainting Problems

Finally, we consider minimization problems involving *unitarily invariant matrix norms*, i.e., matrix norms $\|\cdot\|_\bullet$ fulfilling

$$\|X\|_\bullet = \|UXV^T\|_\bullet$$

for all unitary matrices $U \in \mathbb{R}^{N,N}, V \in \mathbb{R}^{M,M}$. By a result of von Neumann (1937) the unitarily invariant matrix norms are exactly those matrix norms which can be written in the form

$$\|X\|_\bullet = s_\bullet(\sigma(X)), \tag{35}$$

where $\sigma(X)$ is the vector of singular values of X and $s_\bullet$ is a symmetric gauge function, see [63] and for Hermitian matrices [38]. Note that a symmetric gauge function $s : \mathbb{R}^n \to \mathbb{R}^+$ is a norm satisfying $s(\varepsilon_1 x_{i_1}, \ldots, \varepsilon_n x_{i_n}) = s(x)$, where $\varepsilon_i = \pm 1$ for all i and $i_1, \ldots, i_n$ is a permutation of $1, \ldots, n$. We are interested in the *Schatten-p norms* for $p = 1, 2, \infty$ which are defined for $X \in \mathbb{R}^{N,M}$ and $t := \min\{N, M\}$ by

$$\|X\|_n \;\; := \;\; \sum_{i=1}^{t} \sigma_i(X) = s_*(\sigma(X)) = \|\sigma(X)\|_1 \qquad \text{(nuclear norm)},$$

$$\|X\|_F \;\; := \;\; (\sum_{i=1}^{N}\sum_{j=1}^{M} x_{ij}^2)^{\frac{1}{2}} = (\sum_{i=1}^{t} \sigma_i^2(X))^{\frac{1}{2}} = s_F(\sigma(X)) = \|\sigma(X)\|_2 \quad \text{(Frobenius norm)},$$

$$\|X\|_2 \;\; := \;\; \max_{i=1,\ldots,t} \sigma_i(X) = s_2(\sigma(X)) = \|\sigma(X)\|_\infty, \qquad \text{(spectral norm)}.$$

Let $A : \mathbb{R}^{N,M} \to \mathbb{R}^{N,M}$ be a linear operator and $B \in \mathbb{R}^{N,M}$ be not contained in $\mathcal{N}(A^*)$. We are interested in the minimization problems

$$(D_{2,\lambda}) \qquad \operatorname*{argmin}_{P \in \mathbb{R}^{N,M}} \left\{ \frac{1}{2}\|A(P) - B\|_F^2 \quad \text{subject to} \quad \|P\|_n \leq \lambda \right\},$$

$$(D_{1,\tau}) \qquad \operatorname*{argmin}_{P \in \mathbb{R}^{N,M}} \left\{ \frac{1}{2}\|A(P) - B\|_F^2 \;+\; \tau\|P\|_n \right\}.$$

Problem $(D_{1,\tau})$ with the identity operator A was considered in connection with low rank matrix recovery such as matrix completion [59, 8]. With a pointwise multiplication operator A with entries in $\{0, 1\}$ it appears in tensor inpainting [39]. We are not aware of applications of model $(D_{2,\lambda})$ in the literature. Using von Neumann's characterization (35) we can check that the dual norm of a unitarily invariant matrix norm is just given by the dual gauge function. Thus, the dual norm of the nuclear norm is the spectral norm, $\|X\|_2$ and the corresponding primal problems can be written as

$$(P_{2,\lambda}) \qquad \operatorname*{argmin}_{X \in \mathbb{R}^{N,M}} \left\{ \frac{1}{2}\|X - B\|_F^2 \;+\; \lambda\|A^*(X)\|_2 \right\},$$

$$(P_{1,\tau}) \qquad \operatorname*{argmin}_{X \in \mathbb{R}^{N,M}} \left\{ \frac{1}{2}\|X - B\|_F^2 \quad \text{subject to} \quad \|A^*(X)\|_2 \leq \tau \right\}.$$

These problems fit into our setting with $H := I$, $L := A^*$ and $\Phi(X) := \frac{1}{2}\|X - B\|_F^2$ which has $\Phi^*(P) := \frac{1}{2}\|P + B\|_F^2 - \frac{1}{2}\|B\|_F^2$ as dual function. By (12) the solutions $\hat{X}$ and $\hat{P}$ of the primal and dual problems, resp., are related by

$$\hat{X} = B - A(\hat{P}). \tag{36}$$

Now Theorem 3.6 implies the following corollary.

Corollary 4.10. *Let*

$$c := \|A^*(B)\|_2 \quad \text{and} \quad d := \min_{A^*(B)=A^*A(P)} \|P\|_n.$$

Then, for $\lambda \in (0,d)$, a solution $\hat{P}$ of $(D_{2,\lambda})$ is also a solution of $(D_{1,\tau})$ iff $\tau = \|A^*(B - A(\hat{P}))\|_2$.

For solving $(D_{2,\lambda})$ we can use the following proposition.

Proposition 4.11. *Let $B = U_B \Sigma_B V_B^T$ be a singular value decomposition of B. Then the minimizer of the proximity problem*

$$\operatorname*{argmin}_{P \in \mathbb{R}^{N,M}} \left\{ \frac{1}{2} \|B - P\|_F^2 + \iota_{\operatorname{lev}_\lambda \|\cdot\|_\bullet}(P) \right\} \tag{37}$$

is given by $\hat{P} = U_B \Sigma_{\hat{P}} V_B^T$, where the singular values $\sigma(\hat{P})$ in $\Sigma_{\hat{P}}$ are determined by

$$\sigma(\hat{P}) = \operatorname*{argmin}_{\sigma \in \mathbb{R}^t} \left\{ \frac{1}{2} \|\sigma(B) - \sigma\|_2^2 + \iota_{\operatorname{lev}_\lambda s_\bullet}(\sigma) \right\} \tag{38}$$

with the symmetric gauge function $s_\bullet$ corresponding to $\|\cdot\|_\bullet$.

The proof follows by the duality relation (36) (with $A = I$) and a known relation for the dual problem of the proximity problem (37), see, e.g., [54, Proposition 2.2]. Note that a numerical solution of $(D_{1,\tau})$, resp. $(P_{1,\tau})$ without singular value decomposition was proposed in [9].

Example 4.12. *Fig. 8 shows solutions of the inpainting problem modeled by $(D_{2,\lambda})$ for different values λ. The operator A is determined by the mask in Fig. 8 (a). For solving $(D_{2,\lambda})$ we have used the FBS algorithm in Appendix A.1 with $f_1 := \frac{1}{2}\|A(\cdot) - B\|_F^2$ and $f_2 := \iota_{B_{*,\lambda}}$ for which $\nabla f_1(P) = A^*(A(P) - B)$ and $\operatorname{prox}_{\eta f_2} = P_{\operatorname{lev}_\lambda \|\cdot\|_*}$. Here $P_{\operatorname{lev}_\lambda \|\cdot\|_*}$ can be computed by solving the proximity problem (37) in Proposition 4.11 via (38). Further, we have chosen $\eta := 1 \in (0,2)$. The original image of the facade without corruption is denoted by P_{orig}.*
Finally, we want to illustrate the relation between λ and τ given by Corollary 4.10, i.e., the curve g^{-1}. To this end, we have used the same settings as in Fig. 8 but with a part of the facade image. More precisely, the input is given by the MATLAB command B(101:200,101:200). Fig. 9 shows the function g^{-1} for
$\lambda \in \{1, 10, 20, 30, \ldots, 990, 1000, 1500, 2000, 2500, 3000, \ldots, 26500, 27000\}$.

(a) Corrupted image B.

(b) $\hat{P}$ with $\|\hat{P}\|_n \approx \|P_{\mathrm{orig}}\|_n$.

(c) $\hat{P}$ with $\|\hat{P}\|_n \approx \frac{2}{3}\|P_{\mathrm{orig}}\|_n$.

(d) $\hat{P}$ with $\|\hat{P}\|_n \approx \frac{1}{2}\|P_{\mathrm{orig}}\|_n$.

Figure 8: Solutions of the described inpainting problem $(D_{2,\lambda})$ for various values of λ.
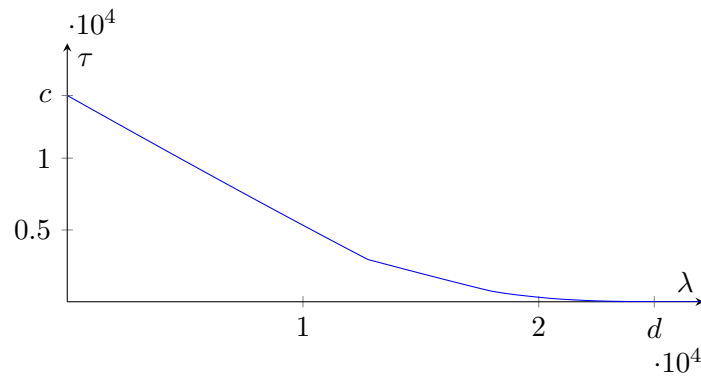

Figure 9: Function g^{-1} from Theorem 3.6 for the described tensor inpainting problem.

5 Conclusions

In this paper we have considered relations between convex problems with constrained homogeneous functions and their penalized versions together with various old and new examples from image restoration. These constrained optimization problems can be solved directly by using first order primal dual methods together with an appropriate splitting to decouple the involved variables. The models are useful if we have a priori knowledge on the image or the class of images rather than on the noise statistics. Recently bilevel optimization approaches for parameter learning in variational models based, e.g., on solving

$$\min_{\lambda} \|x(\lambda) - b\|_2^2 \quad \text{subject to} \quad x(\lambda) = \operatorname*{argmin}_{x} \left\{ \frac{1}{p} \sum_{k=1}^{c} \lambda_k \|K_k x\|_p^p + \frac{1}{2} \|x - b\|_2^2 \right\}$$

were proposed in [36, 46, 51]. Here b is a collection of images from a certain class.

If the noise level can be estimated it would be also interesting to consider constrained problems where we change the roles of Φ and Ψ, i.e., the data misfit term is in the constraint. For the continuous L_2 - TV model the relations between these models were examined in [11, Theorem 2.1]. One can attempt to find a minimizer of such a constrained functional by solving the corresponding penalized problem or a sequence of penalized problems. This goes back to the Morozov discrepancy principle [43]. For image restoration tasks such approaches were recently successfully applied in connection with first order primal dual methods in [45, 64]. For 'sparsity' models the reader may again consider [61, 40]. Further modifications for spatially adapted regularization parameter selection as in [15, 22] are also of interest. However, one of the main topics is still the estimation of the noise statistics.

A Appendix: Algorithms

In this section we describe the convex optimization algorithms used in Section 4 to find the minimizers of the various functionals, namely the forward-backward splitting algorithm (FBS) and the alternating direction method of multipliers (ADMM).

A.1 Forward-Backward Splitting

Let $f_1, f_2 \in \Gamma_0(\mathbb{R}^n)$. Assume that there exists $\beta > 0$ such that f_1 is $1/\beta$-Lipschitz differentiable on $\mathbb{R}^n$. We are looking for solutions of

$$\hat{x} \in \operatorname*{argmin}_{x \in \mathbb{R}^n} \{f_1(x) + f_2(x)\}. \tag{39}$$

For $f \in \Gamma_0(\mathbb{R}^n)$ the *proximity operator* is given by

$$\operatorname{prox}_f(x) := \operatorname*{argmin}_{y \in \mathbb{R}^n} \{\frac{1}{2}\|x - y\|_2^2 + f(y)\}.$$

The iterates of the following algorithm converge to a solution of (39), see, e.g., [1, 27].

Algorithm (FBS)
Initialization: $x^{(0)} \in \mathbb{R}^n, \eta \in (0, 2\beta)$.
For $k = 0, \ldots$ repeat until a stopping criterion is reached

$$x^{(k+1)} := \operatorname{prox}_{\eta f_2}(I - \eta \nabla f_1)(x^{(k)}).$$

A.2 Alternating Direction Method of Multipliers

We briefly introduce the alternating direction method of multipliers (ADMM) for problems of the form

$$\underset{x\in\mathbb{R}^n,y\in\mathbb{R}^m}{\operatorname{argmin}} \{f_1(x) + f_2(y)\} \text{ subject to } Kx = y \tag{40}$$

with $f_1 \in \Gamma_0(\mathbb{R}^n)$, $f_2 \in \Gamma_0(\mathbb{R}^m)$ and $K \in \mathbb{R}^{m,n}$. With the augmented Lagrangian

$$\mathcal{L}_\gamma(x,y,p) \quad := \quad f_1(x) + f_2(y) + \langle p, Kx - y \rangle + \frac{\gamma}{2}\|Kx - y\|_2^2, \quad \gamma > 0$$

the ADMM algorithm reads

Algorithm (ADMM)
Initialization: $y^{(0)}$, $p^{(0)} \in \mathbb{R}^m$ and $\gamma > 0$.
For $k = 0, \dots$ repeat until a convergence criterion is reached:

$$x^{(k+1)} = \underset{x\in\mathbb{R}^n}{\operatorname{argmin}} \mathcal{L}_\gamma(x, y^{(k)}, p^{(k)})$$
$$y^{(k+1)} = \underset{y\in\mathbb{R}^m}{\operatorname{argmin}} \mathcal{L}_\gamma(x^{(k+1)}, y, p^{(k)})$$
$$p^{(k+1)} = p^{(k)} + \gamma(Kx^{(k+1)} - y^{(k+1)}).$$

Replacing p by $q := p/\gamma$ one obtains a *scaled* ADMM version, see [6].

Algorithm (Scaled ADMM)
Initialization: $y^{(0)}$, $b^{(0)} \in \mathbb{R}^m$ and $\gamma > 0$.
For $k = 0, \dots$ repeat until a convergence criterion is reached:

$$x^{(k+1)} = \underset{x\in\mathbb{R}^n}{\operatorname{argmin}} \left\{f_1(x) + \frac{\gamma}{2}\|q^{(k)} + Kx - y^{(k)}\|_2^2\right\}$$
$$y^{(k+1)} = \underset{y\in\mathbb{R}^m}{\operatorname{argmin}} \left\{f_2(y) + \frac{\gamma}{2}\|q^{(k)} + Kx^{(k+1)} - y\|_2^2\right\}$$
$$q^{(k+1)} = q^{(k)} + Kx^{(k+1)} - y^{(k+1)}$$

Note that for our kind of problems, the ADMM coincides with the alternating split Bregman method [33] and the Douglas-Rachford splitting method applied to the dual problem, see, e.g., [6, 25, 26, 30, 52]. The values $p^{(k)}$ converge to a solution of the dual problem of (40). Convergence of $x^{(k)}$ to a solution of (40) is ensured if there is a unique minimizer in the first step of the algorithm, see [52], which is the case in all our examples. Note that a useful adaptation strategy for choosing γ is contained in [6].
We have applied the splitting idea from the PIDSplit algorithm in [29, 53] to the setting

$$(P_{1,\tau}) \quad \underset{x\in\mathbb{R}^n}{\operatorname{argmin}} \{\langle 1_n, Hx - b\log(Hx)\rangle \quad \text{subject to} \quad \||Lx|\|_1 \leq \tau\},$$

which can be rewritten as

$$\underset{x\in\mathbb{R}^n,y\in\mathbb{R}^{n+m}}{\operatorname{argmin}} \Big\{ \underbrace{\langle 0, x\rangle}_{=:f_1(x)} + \underbrace{\langle 1, y_1 - b\log(y_1)\rangle + \iota_{B_{1,\tau}}(y_2)}_{:=f_2(y_1,y_2)} \text{ subject to } \underbrace{\begin{pmatrix} H \\ L \end{pmatrix}}_{K} x = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \Big\}. \tag{41}$$

Application of the scaled ADMM results in the following algorithm.

Scaled ADMM for (41)

Initialization: $q_1^{(0)} = q_2^{(0)} = 0$, $y_1^{(0)} = Hb$, $y_2^{(0)} = Lb$.

For $k = 0, \ldots$ repeat until a stopping criterion is reached

$$
\begin{aligned}
x^{(k+1)} &= \operatorname*{argmin}_{x \in \mathbb{R}^n} \left\{ \|q_1^{(k)} + Hx - y_1^{(k)}\|_2^2 + \|q_2^{(k)} + Lx - y_2^{(k)}\|_2^2 \right\}, \\
y_1^{(k+1)} &= \operatorname*{argmin}_{y_1 \in \mathbb{R}^n} \left\{ \langle 1, y_1 - b\log(y_1) \rangle + \frac{\gamma}{2} \|q_1^{(k)} + Hx^{(k+1)} - y_1\|_2^2 \right\}, \\
y_2^{(k+1)} &= \operatorname*{argmin}_{y_2 \in \mathbb{R}^{2n}} \left\{ \iota_{B_{1,\tau}}(y_2) + \frac{\gamma}{2} \|q_2^{(k)} + Lx^{(k+1)} - y_2\|_2^2 \right\}, \\
q_1^{(k+1)} &= q_1^{(k)} + Hx^{(k+1)} - y_1^{(k+1)}, \\
q_2^{(k+1)} &= q_2^{(k)} + Lx^{(k+1)} - y_2^{(k+1)}.
\end{aligned}
$$

We have that $y_2^{(k+1)} = P_{B_{1,\tau}}\left(q_2^{(k)} + Lx^{(k+1)}\right)$. The other steps of the algorithm can be performed as in [53, 56].

To see how (a part of) the dual variable $\hat{q}$ computed in the scaled ADMM algorithm is related to the dual variable $\hat{p}$ in the Fenchel relation in Lemma 2.6 and in the computation of λ we need the following remark.

Remark A.1. *An alternative formulation of the primal and dual problems in Lemma 2.6 can be given by the help of the Lagrangian. We rewrite* (P) *as*

$$
\operatorname*{argmin}_{x,y} \{ \Phi(x) + \Psi(\mu y) \quad subject\ to \quad Lx = y \}
$$

which has the Lagrangian $\mathcal{L}(x, y, p) = \Phi(x) + \Psi(\mu y) + \langle p, Lx - y \rangle$. *Then the primal and dual problems can be written as*

$$
(P) \quad \operatorname*{argmin}_{x \in \mathbb{R}^d} \min_{y \in \mathbb{R}^m} \max_{p \in \mathbb{R}^m} \left\{ \Phi(x) + \Psi(\mu y) + \langle p, Lx - y \rangle \right\},
$$

$$
(D) \quad \operatorname*{argmax}_{p \in \mathbb{R}^m} \min_{x \in \mathbb{R}^d, y \in \mathbb{R}^m} \left\{ \Phi(x) + \Psi(\mu y) + \langle p, Lx - y \rangle \right\},
$$

see [4, 5]. Since $\min_{y \in \mathbb{R}^m} \{ \Psi(\mu y) - \langle p, y \rangle \} = -\Psi^*(p/\mu)$ *the dual and the primal problems can be rewritten as*

$$
(P) \quad \operatorname*{argmin}_{x \in \mathbb{R}^d} \max_{p \in \mathbb{R}^m} \left\{ \Phi(x) - \Psi^*\left(\frac{p}{\mu}\right) + \langle p, Lx \rangle \right\},
$$

$$
(D) \quad \operatorname*{argmax}_{p \in \mathbb{R}^m} \min_{x \in \mathbb{R}^d} \left\{ \Phi(x) - \Psi^*\left(\frac{p}{\mu}\right) + \langle p, Lx \rangle \right\}.
$$

To obtain the curve g *from Corollary 4.8 we have to compute*

$$
\lambda = \| |\hat{p}| \|_\infty = \gamma \| |\hat{q}_2| \|_\infty
$$

which can be seen as follows: By the above considerations problem $(P_{1,\tau})$ *can be written as*

$$
\operatorname*{argmin}_{x \in \mathbb{R}^n} \max_{p \in \mathbb{R}^m} \left\{ \langle 1, Hx - b\log(Hx) \rangle + \langle p, Lx \rangle - \iota_{B_{1,\tau}}^*(p) \right\}.
$$

On the other hand, the saddle point formulation of (41) *reads*

$$\operatorname*{argmin}_{x\in\mathbb{R}^n} \max_{p_1,p_2} \left\{ -\psi_1^*(p_1) - \iota_{B_{1,\tau}}^*(p_2) + \langle p_1, Hx \rangle + \langle p_2, Lx \rangle \right\}$$

$$= \operatorname*{argmin}_{x\in\mathbb{R}^n} \max_{p_2\in\mathbb{R}^m} \left\{ \psi_1(Hx) - \iota_{B_{1,\tau}}^*(p_2) + \langle p_2, Lx \rangle \right\}$$

with $\psi_1(x) := \langle 1, x - b\log(x) \rangle$.

References

[1] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* Springer, New York, 2011.

[2] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing*, 18(11):2419–2434, 2009.

[3] M. Bertero and P. Boccacci. *Introduction to Inverse Problems in Imaging.* IOP Publishing, Bristol, 1998.

[4] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar. *Convex Analysis and Optimization.* Athena Scientific, Belmont, Massachusetts, 2003.

[5] J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*, volume 7 of *Springer Series in Operations Research.* Springer, Berlin, 2000.

[6] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3:1–122, 2011.

[7] K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM Journal on Imaging Science*, 3(3):492–526, 2010.

[8] J.-F. Cai, E. J. Candés, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956 – 1982, 2010.

[9] J.-F. Cai and S. Osher. Fast singular value thresholding without singular value decomposition. *UCLA CAM Report*, 10-24, 2010.

[10] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, (20):89–97, 2004.

[11] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76:167–188, 1997.

[12] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[13] T. F. Chan, S. Esedoglu, and F. Park. A fourth order dual method for staircase reduction in texture extraction and image restoration problems. CAM Report 05-28, UCLA, 2005.

[14] T. F. Chan and J. Shen. *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastical Methods*. SIAM, Philadelphia, 2005.

[15] D.-Q. Chen and L.-Z. Cheng. Spatially adapted regularization parameter selection based on the local discrepancy function for Poissonian image deblurring. *Inverse Problems*, 28(1):015004, 2012.

[16] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20:33–61, 1998.

[17] G. Cherchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu. A proximal approach for constrained cosparse modelling. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, 2012.

[18] R. Ciak. Homogeneous penalizers and constraints in convex problems. *Preprint University of Kaiserslautern*, 2012.

[19] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering, (H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Editors)*, pages 185–212, New York, 2011. Springer-Verlag.

[20] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling and Simulation*, 4:1168–1200, 2005.

[21] I. Daubechies, M. Fornasier, and I. Loris. Accelerated projected gradient methods for for linear inverse problems with sparsity constraints. *The Journal of Fourier Analysis and Applications*, 14(5-6):764–792, 2008.

[22] Y. Dong, M. Hintermüller, and M. M. Rincon-Camacho. Automated regularization parameter selection in multi-scale total variation models for image restoration. *Journal of Mathematical Imaging and Vision*, 40(1):82–104, 2011.

[23] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52:1289–1306, 2006.

[24] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra. Efficient projections onto the l1-ball for learning in high dimensions. In *ICML '08 Proceedings of the 25th International Conference on Machine Learning*, ACM New York, 2008.

[25] J. Eckstein and D. B. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55:293–318, 1992.

[26] E. Esser. Applications of Lagrangian based alternating direction methods and connections to split Bregman. CAM Report 09-31, UCLA, 2009.

[27] F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*, volume I+II. Springer, New York, 2003.

[28] J. Fadili and G. Peyré. Total variation projection with first order schemes. *IEEE Transactions on Image Processing*, 20(3):657–669, 2011.

[29] M. A. T. Figueiredo and J. M. Bioucas-Dias. Deconvolution of Poissonian images using variable splitting and augmented Lagrangian optimization. In *IEEE Workshop on Statistical Signal Processing*, Cardiff, 2009.

[30] D. Gabay. Applications of the method of multipliers to variational inequalities. In M. Fortin and R. Glowinski, editors, *Augmented Lagrangian Methods: Applications to the Solution of Boundary Value Problems*, chapter IX, pages 299–340. North-Holland, Amsterdam, 1983.

[31] G. Gilboa, J. Darbon, S. Osher, and T. Chan. Nonlocal convex functionals for image regularization. *UCLA CAM Report*, 06-57, 2006.

[32] G. Gilboa and S. Osher. Nonlocal operators with applications to image processing. *Multiscale Modeling & Simulation*, 7(3):1005–1028, 2008.

[33] T. Goldstein and S. Osher. The split Bregman method for l1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[34] J.-B. Hiriart-Urruty and C. Lemarechal. *Convex Analysis and Minimization Algorithms*, volume 1. Springer, Berlin, Heidelberg, 1993.

[35] K. V. Ivanov, V. V. Vasin, and V. P. Tanana. *Theory of Linear Ill-Posed Problems and its Applications*. Brill Academic Publishers, Utrecht, Boston, Koeln, Tokyo, 2002.

[36] K. Kunisch and T. Pock. A bilevel optimization approach for parameter learning in variational models. *Preprint*, 2012.

[37] S. Lefkimmiatis, A. Bourquard, and M. Unser. Hessian-Based Norm Regularization for Image Restoration with Biomedical Applications. *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 983-995, March 2012.

[38] A. S. Lewis. Convex analysis on the Hermitian matrices. *SIAM Journal on Optimization*, 6:164 – 177, 1996.

[39] J. Liu, P. Musialski, P. Wonka, and J. Ye. Tensor completion for estimating missing values in visual data. In *Proceedings of ICCV'2009*, pages 2114–2121, 2009.

[40] D. A. Lorenz. Constructing test instances for basis pursuit denoising. Technical report, TU Braunschweig, 2011. http://arxiv.org/abs/1103.2897.

[41] M. Lysaker, A. Lundervold, and X. Tai. Noise removal using fourth-order partial differential equations with applications to medical magnetic resonance images in space and time. *IEEE Transactions on Image Processing*, 12(12):1579 – 1590, 2003.

[42] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, second edition, 1999.

[43] V. A. Morozov. *Methods for Solving Incorrectly Posed Problems*. Springer, New York, 1984.

[44] Y. Nesterov. Smooth minimization of non-smooth functions. *Mathematical Programming*, 103(1):127–152, 2005.

[45] M. K. Ng, P. Weiss, and X. Yuan. Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods. *SIAM Journal on Scientific Computing*, 32(5):2710–2736, Aug. 2010.

[46] G. Peyré and J. Fadili. Learning analysis sparsity priors. In *Proceedings of Sampta'11*, 2011.

[47] K. R. Rao and P. Yip. *Discrete Cosine Transform*. Academic Press, New York, 1990.

[48] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.

[49] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317 of *A Series of Comprehensive Studies in Mathematics*. Springer, Berlin, 2 edition, 2004.

[50] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica A*, 60:259–268, 1992.

[51] K. Samuel and M. Tappen. Learning optimized map estimates in continuously-valued mrf models. In *Proceedings of CVPR'09*, pages 477–484. IEEE, 2009.

[52] S. Setzer. Operator splittings, Bregman methods and frame shrinkage in image processing. *International Journal of Computer Vision*, 92:265–280, 2011.

[53] S. Setzer, G. Steidl, and T. Teuber. Deblurring Poissonian images by split Bregman techniques. *Journal of Visual Communication and Image Representation*, 21:193–199, 2010.

[54] S. Setzer, G. Steidl, and T. Teuber. On vector and matrix median computation. *Journal of Computational and Applied Mathematics*, 236:2200–2222, 2012.

[55] G. Steidl. A note on the dual treatment of higher order regularization functionals. *Computing*, 76:135 – 148, 2006.

[56] G. Steidl and T. Teuber. Removing multiplicative noise by Douglas-Rachford splitting methods. *Journal of Mathematical Imaging and Vision*, 36(3):168–184, 2010.

[57] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288, 1994.

[58] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill–Posed Problems*. Wiley, Washington, DC, 1977.

[59] K.-C. Toh and S. Yun. An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems. *Pacific Journal of Optimization*, 6:615–640, 2010.

[60] J. Tropp. Just relax: convex programming methods for identifying sparse signals in noise. *IEEE Transactions on Information Theory*, 52:1030:1051, 2006.

[61] E. van den Berg and M. P. Friedlander. Probing the Pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 31(2):890–912, Nov. 2008.

[62] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272, 2008.

[63] G. A. Watson. Characterization of the subdifferential of some matrix norms. *Linear Algebra and its Applications*, 170:33–45, 1992.

[64] Y. Wen and R. Chan. Parameter selection for total variation based image restoration using discrepancy principle. *IEEE Transactions on Image Processing*, 21:1770–1781, 2012.