Hessian Schatten-Norm Regularization for Linear Inverse Problems

Stamatios Lefkimmiatis, *Member, IEEE*, John Paul Ward, and Michael Unser, *Fellow, IEEE*

***Abstract*— We introduce a novel family of invariant, convex, and non-quadratic functionals that we employ to derive regularized solutions of ill-posed linear inverse imaging problems. The proposed regularizers involve the Schatten norms of the Hessian matrix, which are computed at every pixel of the image. They can be viewed as second-order extensions of the popular total-variation (TV) semi-norm since they satisfy the same invariance properties. Meanwhile, by taking advantage of second-order derivatives, they avoid the staircase effect, a common artifact of TV-based reconstructions, and perform well for a wide range of applications. To solve the corresponding optimization problems, we propose an algorithm that is based on a primal-dual formulation. A fundamental ingredient of this algorithm is the projection of matrices onto Schatten norm balls of arbitrary radius. This operation is performed efficiently based on a direct link we provide between vector projections onto ℓ_q norm balls and matrix projections onto Schatten norm balls. Finally, we demonstrate the effectiveness of the proposed methods through experimental results on several inverse imaging problems with real and simulated data.**

***Index Terms*— Eigenvalue optimization, Hessian operator, image reconstruction, matrix projections, Schatten norms.**

I. Introduction

LINEAR inverse problems arise in a host of imaging applications, ranging from microscopy and medical imaging to remote sensing and astronomical imaging [1]. The task is to reconstruct the underlying image from a series of degraded measurements. These problems are often formulated within a variational framework, where image reconstruction can be cast as the minimization of an energy functional subject to some penalty. The role of the penalty is significant, since it imposes certain constraints on the solution and considerably affects the quality of the reconstruction.

The importance of choosing an appropriate penalty has initiated the development of regularization functionals that can effectively model certain properties of natural images. A popular regularization criterion is the total-variation (TV) semi-norm [2] which has been successfully applied to several imaging problems such as image denoising, restoration [3], [4], inpainting [5], zooming [6], and MRI reconstruction [7]. TV owes its success to its ability to preserve the edges of the underlying image well. Its downside, however, is that it introduces blocking artifacts (a.k.a. *staircase effect*) [8]. The reason is that TV favors vanishing first-order derivatives. Thus, it tends to result in piecewise-constant solutions even when the underlying images are not necessarily piecewise constant. This tendency is responsible for oversharpening the contrast along image contours and can be a serious drawback in many applications.

A common workaround to prevent the oversharpening of regions with smooth intensity transitions is to replace TV by functionals that involve higher-order differential operators, because higher-order derivatives can potentially restore a wider class of images. Often, moving from piecewise-constant to piecewise-linear reconstructions offers a satisfactory improvement in the fitting of smooth intensity changes, so that most of the published functionals involve second-order differentials. Such regularizers have been considered, mostly for image denoising, either combined with TV [8]–[12] or in a standalone way [13]–[17]. These recent advances motivate us to investigate a class of regularizers that depend on matrix norms of the Hessian. These regularizers enjoy most of the favorable properties of TV; namely, convexity, contrast, rotation, translation, and scale invariance (up to a multiplicative constant), while they avoid the staircase effect by not penalizing first-order polynomials.

The key contributions of this work are as follows:

1) The identification of a novel family of invariant functionals that involve Schatten norms of the Hessian matrix, computed at every pixel of the image. These are used in a variational framework to derive regularized solutions of ill-posed linear inverse imaging problems. Our functionals capture curvature information related to the image intensity and lead to reconstructions that avoid the staircase effect.
2) A general first-order algorithm for solving the resulting constrained optimization problems under any choice of Schatten norm. The proposed algorithm relies on our derivation of a primal-dual formulation and copes with the non-smooth nature and the high dimensionality of the problem.
3) A direct link between matrix projections onto Schatten norm balls and vector projections onto ℓ_q norm balls. This link enables us to design an efficient method for performing matrix projections. Although it is a funda-

Manuscript received April 27, 2012; revised December 7, 2012; accepted December 17, 2012. Date of publication January 4, 2013; date of current version March 14, 2013. This work was supported in part by the Hasler Foundation, the ERC Grant ERC-2010-AdG 267439-FUN-SP, and the Indo-Swiss Joint Research Program. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Alessandro Foi.

The authors are with the Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne, Lausanne CH-1015, Switzerland (e-mail: stamatis.lefkimmiatis@epfl.ch; john.ward@epfl.ch; michael.unser@epfl.ch).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2013.2237919

1057-7149/$31.00 © 2013 IEEE

mental component of our optimization algorithm, our result is not specific to the Hessian and can potentially have a wider applicability.

The rest of the paper is organized as follows: In Section II, we discuss regularization functionals that are commonly used in imaging problems. Then, by focusing on invariance principles we derive our novel family of non-quadratic second-order functionals. In Section III, we present the discrete formulation of the problem and we describe the proposed optimization algorithm. In Section IV, we assess the performance of our approach for several linear inverse imaging problems with experiments on standard test images and real biomedical images. We conclude our work in Section V. Proofs of mathematical statements are given in Appendices.

II. Derivative-Based Regularization

The most commonly-used regularizers can be expressed as

$$R(f) = \int_{\Omega} \Phi(\mathbf{D} f(\mathbf{r}))\, \mathrm{d}\mathbf{r}, \tag{1}$$

where f is an image, $\Omega \subset \mathbb{R}^2$, $\mathbf{D}$ is the *regularization operator* (scalar or multi-component) acting on the image, and $\Phi(\cdot)$ is a potential function. Typical choices for $\mathbf{D}$ are differential operators such as the Laplacian (scalar operator) or the gradient (vectorial operator), while the potential function Φ usually involves a norm distance. For many years, the preferred choice for the potential function has been the squared Euclidean norm, because of its mathematical tractability and computational simplicity. However, it is now widely documented that non-quadratic potential functions can lead to improved results; they can be designed to be less sensitive to outliers and therefore provide better edge reconstruction. A typical example is TV, which for smooth images corresponds to the L_1 norm of the magnitude of the gradient.

Our present goal is to introduce new regularization functionals of the form of (1) which amounts to specifying some suitable linear operator $\mathbf{D}$, and potential function Φ. To do so, certain requirements should be fulfilled. In particular, following the example of TV, we restrict ourselves to regularization operators that commute with translation and scaling, and potential functions that preserve these properties while introducing additional rotation invariance. Our motivation for enforcing these invariances is that, similarly to what is the case in many physical systems, one should opt for reconstruction algorithms that lead to solutions which are not affected by transformations of the coordinate system. An additional desirable requirement is that the regularizers should be convex to ensure that if a minimum exists, then this is a global one. Furthermore, convexity permits the design of efficient minimization techniques.

A. Gradient Norm Regularization

We would like our regularization operator to be translation and scale invariant. Therefore, a reasonable choice for $\mathbf{D}$ is some form of derivative operator. Based on this, we first characterize the complete class of gradient-based regularizers satisfying all the required invariances. This is accomplished by Theorem 1 which specifies the valid form for the potential functions Φ. The proof of this theorem is given in Appendix I.

Theorem 1: Let $R(f)$ be of the form (1), where $\mathbf{D}$ is the gradient operator and f is continuously differentiable. $R(f)$ is a translation-, rotation-, and scale-invariant functional, if and only if the potential function $\Phi : \mathbb{R}^2 \mapsto \mathbb{R}$ is of the form: $\Phi(\nabla f(\mathbf{r})) = c\,|\nabla f(\mathbf{r})|^{\nu}$, where $\nu \in \mathbb{R}$ and c is an arbitrary constant.

As a direct consequence of Theorem 1, we see that the following gradient-based regularizers are the only choice of regularization satisfying the required invariance properties; ignoring the multiplicative constant c of the potential function, which can be absorbed by the regularization parameter, we get

$$R(f) = \int_{\Omega} |\nabla f(\mathbf{r})|^{\nu}\, \mathrm{d}\mathbf{r}. \tag{2}$$

Since we are also interested in convex regularization functionals, we shall focus on cases where $\nu \geq 1$ in (2). A popular instance of convex functionals arises if we choose $\nu = 1$, which corresponds to TV. This regularizer enjoys an additional property, that of contrast covariance.

B. Hessian Schatten-Norm Regularization

As already mentioned, the use of TV, which is the best representative of the gradient-based regularization family, suffers from certain drawbacks. Therefore, for the reasons specified in the introduction, we are interested in differential operators of higher-order and in particular of the second order. In N-dimensions, the complete spectrum of second-order derivatives is embodied in the Hessian operator,

$$Hf(\mathbf{r}) = \begin{bmatrix} f_{r_1 r_1}(\mathbf{r}) & f_{r_1 r_2}(\mathbf{r}) \\ f_{r_2 r_1}(\mathbf{r}) & f_{r_2 r_2}(\mathbf{r}) \end{bmatrix}, \tag{3}$$

where $f_{r_i r_j}(\mathbf{r}) = \frac{\partial^2}{\partial r_i\, \partial r_j} f(\mathbf{r})$. Indeed, with the aid of the Hessian we can compute any second-order derivative of $f(\mathbf{r})$ as $\mathrm{D}^2_{\theta,\phi} f(\mathbf{r}) = \mathbf{u}_\theta^T Hf(\mathbf{r})\, \mathbf{v}_\phi$, where $\mathbf{u}_\theta = (\cos\theta, \sin\theta)$ and $\mathbf{v}_\phi = (\cos\phi, \sin\phi)$ are unit-norm vectors specifying the directions of differentiation and $(\cdot)^T$ is the transpose operation.

Having specified the regularization operator $\mathbf{D}$, the next step is to investigate which class of potential functions Φ leads to translation-, rotation-, and scale-invariant second-order regularizers. Next, we provide Theorem 2 which completely characterizes the form of Φ, under these prerequisites. Before presenting this result, we first give the general definition of a Schatten matrix norm [18] that will be used in the sequel, and introduce some of the adopted notation. We denote the set of unitary matrices as $\mathbb{U}^n = \left\{\mathbf{X} \in \mathbb{C}^{n\times n} : \mathbf{X}^{-1} = \mathbf{X}^H\right\}$, where $\mathbb{C}$ is the set of complex numbers and $(\cdot)^H$ is the Hermitian transpose. We also denote the set of positive semidefinite diagonal matrices as $\mathbb{D}^{n_1\times n_2} = \left\{\mathbf{X} \in \mathbb{R}_+^{n_1\times n_2} : \mathbf{X}(i,j) = 0\ \forall\, i \neq j\right\}$, where $\mathbb{R}_+$ is the set of real non-negative numbers.

Definition 1 (Schatten Norms): Let $\mathbf{X} \in \mathbb{C}^{n_1\times n_2}$ be a matrix with the singular-value decomposition (SVD) $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, where $\mathbf{U} \in \mathbb{U}^{n_1}$ and $\mathbf{V} \in \mathbb{U}^{n_2}$ consist of the singular vectors of $\mathbf{X}$, and $\mathbf{\Sigma} \in \mathbb{D}^{n_1\times n_2}$ consists of the singular values of $\mathbf{X}$. The Schatten norm of order p (S_p norm) of $\mathbf{X}$ is

defined as

$$\|\mathbf{X}\|_{S_p} = \left(\sum_{k=1}^{\min(n_1,n_2)} \boldsymbol{\sigma}_k^p (\mathbf{X}) \right)^{\frac{1}{p}}, \quad (4)$$

where $p \geq 1$, and $\boldsymbol{\sigma}_k (\mathbf{X})$ is the k-th singular value of $\mathbf{X}$, which corresponds to the (k, k) entry of $\boldsymbol{\Sigma}$.

Definition 1 implies that the S_p norm of a matrix $\mathbf{X}$ corresponds to the ℓ_p norm of its singular-values vector $\boldsymbol{\sigma} (\mathbf{X}) \in \mathbb{R}_+^{\min(n_1,n_2)}$. This further means that all Schatten norms are unitarily invariant. Moreover, we note that the family of S_p norms includes three of the most popular matrix norms, i.e., the nuclear/trace norm ($p = 1$), the Frobenius norm ($p = 2$) and the spectral/operator norm ($p = \infty$).

Theorem 2: Let $R(f)$ be of the form (1), where $\mathbf{D}$ is the Hessian operator and f is twice continuously differentiable. $R(f)$ is a translation-, rotation-, and scale-invariant functional, if and only if the potential function $\Phi : \mathbb{R}^{2\times 2} \mapsto \mathbb{R}$ is of the form: $\Phi (Hf (\mathbf{r})) = \Phi_0 \left(\boldsymbol{\lambda}_f (\mathbf{r}) / \|Hf (\mathbf{r})\|_{S_p} \right) \|Hf (\mathbf{r})\|_{S_p}^{\nu}$, where $\nu \in \mathbb{R}$ and Φ_0 is a zero-degree homogeneous function of the Hessian eigenvalues $\boldsymbol{\lambda}_f (\mathbf{r})$.

The proof of Theorem 2 is given in Appendix I. Now, according to it, the admissible second-order regularizers, with respect to the invariance properties of the coordinate system, are those depending on the Schatten norms of the Hessian. If we set $\Phi_0 = 1$, we obtain the following regularization family

$$R(f) = \int_{\Omega} \|Hf (\mathbf{r})\|_{S_p}^{\nu} \, \mathrm{d}\mathbf{r}, \forall p \geq 1 \text{ and } \nu \in \mathbb{R}. \quad (5)$$

To further ensure convexity we need to impose that $\nu \geq 1$. Finally, for our regularizers to also enjoy the contrast-covariance property (similar to TV), we focus on the case where $\nu = 1$. Consequently, we define our proposed family of non-quadratic second-order regularization functionals as

$$R(f) = \int_{\Omega} \|Hf (\mathbf{r})\|_{S_p} \, \mathrm{d}\mathbf{r}, \forall p \geq 1. \quad (6)$$

The introduced functionals, depending on the Hessian, lead to piecewise-linear reconstructions. These reconstructions can better approximate the intensity variations observed in natural images than the piecewise-constant reconstructions provided by TV. Thus, they are able to avoid the staircase effect. Moreover, since the Hessian of f at coordinates $\mathbf{r}$ is a 2×2 symmetric matrix, the SVD in the Schatten norm definition reduces to the spectral decomposition and the singular values correspond to the absolute eigenvalues, which can be computed analytically. Now, if we consider the intensity map of the image as a 3-D differentiable surface, then the two eigenvalues of the Hessian at coordinates $\mathbf{r}$ correspond to the principal curvatures. They can be used to measure how this surface bends by different amounts in different directions at that point. Therefore, the proposed potential functions, which depend upon those, can be interpreted as scalar measurements of the curvature at a local surface patch. For example, the S_2 norm (Frobenius norm) of the Hessian is a scalar curvature index, commonly used in differential geometry, which quantifies lack of flatness of the surface at a specific point. Therefore, we can safely state that the proposed regularizers incorporate curvature information about the image intensity.

Finally, we note that the regularizers obtained for two choices of $p = 2, \infty$, coincide with functionals we considered in our previous work in [19], where we followed another path for extending TV based on rotational averages of directional derivatives. To the best of our knowledge, the Hessian Schatten norms for $p \neq 2, \infty$ have not been considered before in the context of inverse problems.

III. Variational Image Reconstruction

From now on, we focus on the discrete formulation of image reconstruction. Hereafter, to avoid any confusion between the continuous and the discrete domains we will use bold-faced symbols to refer to the discrete manipulation of the problem.

A. Discrete Problem Formulation

Our approach for reconstructing the underlying image from the measurements is based on the linear observation model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}, \quad (7)$$

where $\mathbf{A} \in \mathbb{R}^{M\times N}$ is a matrix that models the spatial response of the imaging device, while $\mathbf{y} \in \mathbb{R}^M$ and $\mathbf{x} \in \mathbb{R}^N$ are the vectorized versions of the observed image and the image to be estimated, respectively. Apart from the effect of the operator $\mathbf{A}$ acting on the underlying image, another perturbation is the measurement noise, which is intrinsic in the detection process. This degradation factor is represented in our observation model by $\mathbf{w}$ that we, here on, will assume to be i.i.d Gaussian noise with variance σ_w^2.

The recovery of $\mathbf{x}$ from the measurements $\mathbf{y}$ belongs to the category of linear inverse problems. Usually, for the cases of practical interest, it is *ill-posed* [20]: the operator $\mathbf{A}$ is either ill-conditioned or singular. This is dealt with in the variational framework by forming an objective function

$$\varphi (\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \tau \psi (\mathbf{x}), \quad (8)$$

whose role is to quantify the quality of a given estimate. The first term, also known as *data fidelity*, corresponds to the negative Gaussian log-likelihood and measures how well a candidate estimate explains the observed data. The second term (*regularization*) encodes our beliefs about certain characteristics of the underlying image. Its role is to narrow down the set of plausible solutions by penalizing those that do not satisfy the assumed properties. The parameter $\tau \geq 0$ provides a balance between the contribution of the two terms. The image reconstruction problem is then cast as the minimization of (8) and leads to a penalized least-squares solution.

B. Discrete Hessian Operator and Basic Notations

In this work we focus on the class of Hessian Schatten-norm regularizers presented in (6). In the sequel, we use $\boldsymbol{H}$ to refer to the discrete version of the Hessian operator. To simplify our analysis, we assume that the image intensities on a $N_x \times N_y$ grid are rasterized in a vector $\mathbf{x}$ of size $N = N_x \cdot N_y$, so that the pixel at coordinates (i, j) maps to the nth entry of $\mathbf{x}$ with $n = jN_x + (i+1)$. In this case, the discrete Hessian operator is

a mapping $\boldsymbol{H}: \mathbb{R}^N \mapsto X$, where $X = \mathbb{R}^{N\times 2\times 2}$. For $\mathbf{x} \in \mathbb{R}^N$, $\boldsymbol{H}\mathbf{x}$ is given as

$$[\boldsymbol{H}\mathbf{x}]_n = \begin{bmatrix} \left[\Delta_{r_1r_1}\mathbf{x}\right]_n & \left[\Delta_{r_1r_2}\mathbf{x}\right]_n \\ \left[\Delta_{r_1r_2}\mathbf{x}\right]_n & \left[\Delta_{r_2r_2}\mathbf{x}\right]_n \end{bmatrix}, \tag{9}$$

where $[\cdot]_n$ denotes the nth element of the argument, $n = 1, \ldots, N$, and $\Delta_{r_1r_1}$, $\Delta_{r_2r_2}$, and $\Delta_{r_1r_2}$ denote the forward finite-difference operators [21] that approximate the second-order partial derivatives along the two dimensions of the image. If we assume Neumann boundary conditions and use the standard representation of the image rather than the vectorized one, these operators are defined as

$$\left[\Delta_{r_1r_1}\mathbf{x}\right]_{i,j} = \begin{cases} x_{i+2,j} - 2x_{i+1,j} + x_{i,j}, & 1 \le i \le N_x - 2, \\ x_{N_x-1,j} - x_{N_x,j}, & i \ge N_x - 1, \end{cases} \tag{10a}$$

$$\left[\Delta_{r_2r_2}\mathbf{x}\right]_{i,j} = \begin{cases} x_{i,j+2} - 2x_{i,j+1} + x_{i,j}, & 1 \le j \le N_y - 2, \\ x_{i,N_y-1} - x_{i,N_y}, & j \ge N_y - 1, \end{cases} \tag{10b}$$

$$\left[\Delta_{r_1r_2}\mathbf{x}\right]_{i,j} = \begin{cases} x_{i+1,j+1} - x_{i+1,j} - x_{i,j+1} + x_{i,j}, \\ \quad 1 \le i \le N_x - 1 \text{ and } 1 \le j \le N_y - 1, \\ 0, \quad \text{otherwise}. \end{cases} \tag{10c}$$

We equip the space X with the inner product $\langle \cdot\,,\, \cdot \rangle_X$ and norm $\|\cdot\|_X$. To define them, let $\mathbf{X}, \mathbf{Y} \in X$, with $\mathbf{X}_n, \mathbf{Y}_n \in \mathbb{R}^{2\times 2}\ \forall\ n = 1, \ldots, N$. Then we have

$$\langle \mathbf{X}\,,\, \mathbf{Y} \rangle_X = \sum_{n=1}^{N} \operatorname{tr}\left(\mathbf{Y}_n^T \mathbf{X}_n\right) \tag{11}$$

and

$$\|\mathbf{X}\|_X = \sqrt{\langle \mathbf{X}\,,\, \mathbf{X} \rangle_X}\,, \tag{12}$$

where $\operatorname{tr}(\cdot)$ is the trace operator. For the Euclidean space $\mathbb{R}^N$ we use the standard definition of the inner product and of the norm. We denote them by $\langle \cdot\,,\, \cdot \rangle_2$ and $\|\cdot\|_2$, respectively.

The adjoint of $\boldsymbol{H}$ is the discrete operator $\boldsymbol{H}^*: X \mapsto \mathbb{R}^N$ such that

$$\langle \mathbf{Y}\,,\, \boldsymbol{H}\mathbf{x} \rangle_X = \left\langle \boldsymbol{H}^*\mathbf{Y}\,,\, \mathbf{x} \right\rangle_2 . \tag{13}$$

This definition of the adjoint operator is a generalization of the Hermitian transpose for matrices. Based on the relation of the inner products in (13), we show in Appendix II that for any $\mathbf{Y} \in X$, it holds that

$$\begin{aligned} \left[\boldsymbol{H}^*\mathbf{Y}\right]_n = {} & \left[\Delta_{r_1r_1}^*\mathbf{Y}^{(1,1)}\right]_n + \left[\Delta_{r_2r_2}^*\mathbf{Y}^{(2,2)}\right]_n \\ & + \left[\Delta_{r_1r_2}^*\left(\mathbf{Y}^{(1,2)} + \mathbf{Y}^{(2,1)}\right)\right]_n , \end{aligned} \tag{14}$$

where $\mathbf{Y}_n^{(i,j)}$ is the (i, j) entry of the 2×2 matrix $\mathbf{Y}_n$ and $\Delta_{r_1r_1}^*$, $\Delta_{r_1r_2}^*$, $\Delta_{r_2r_2}^*$ are the adjoint operators that correspond to backward difference operators with Neumann boundary conditions.

C. Majorization-Minimization Algorithm

Next, we present a general method to compute the minimizer of the functional in (8), under any Hessian-based S_p norm regularizer. Since these regularizers are non-smooth, our algorithm is based on a majorization-minimization (MM) approach (cf. [22], [23], [24] for instance). Under this framework, instead of directly minimizing (8), we find the solution via the successive minimization of a sequence of surrogate functions that upper bound the initial objective function [25]. Our motivation for taking this path is that each of the surrogate functions is simpler to minimize, and we can rely on a gradient scheme that efficiently copes with the large dimensionality of the problem.

To obtain the surrogate functions, we upper bound the data term of our objective function using the following majorizer [22], [26]

$$g(\mathbf{x}, \mathbf{x}_0) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + d(\mathbf{x}, \mathbf{x}_0)\,, \tag{15}$$

where $d(\mathbf{x}, \mathbf{x}_0) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T\left[\alpha\mathbf{I} - \mathbf{A}^T\mathbf{A}\right](\mathbf{x} - \mathbf{x}_0)$ is a function that measures the distance between $\mathbf{x}$ and $\mathbf{x}_0$. To come up with a valid majorizer we need to ensure that $d(\mathbf{x}, \mathbf{x}_0) \ge 0$, $\forall \mathbf{x}$, with equality if and only if $\mathbf{x} = \mathbf{x}_0$. This prerequisite is true if $\alpha\mathbf{I} - \mathbf{A}^T\mathbf{A}$ is positive definite, which implies that $\alpha > \left\|\mathbf{A}^T\mathbf{A}\right\|$. The upper-bounded version of the overall objective (8) can be written as

$$\tilde{\varphi}(\mathbf{x}, \mathbf{x}_0) = \frac{\alpha}{2}\|\mathbf{x} - \mathbf{z}\|_2^2 + \tau\,\psi(\mathbf{x}) + c\,, \tag{16}$$

where c is a constant and $\mathbf{z} = \mathbf{x}_0 + \alpha^{-1}\mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x}_0)$. Then, the next step is to iteratively minimize (16) w.r.t $\mathbf{x}$, setting $\mathbf{x}_0$ to the previous iteration's solution. As we see, in (16) there is no coupling between $\mathbf{x}$ and the operator $\mathbf{A}$ anymore, which turns the minimization task into a much simpler one. In fact, the minimizer of (16) can also be interpreted as the solution of a denoising problem with $\mathbf{z}$ being the noisy measurements.

D. Proximal Map Evaluation and Matrix Projections

The MM formulation of our problem relies on the solution of a simpler problem of the form

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x} - \mathbf{z}\|_2^2 + \tau\,\psi(\mathbf{x}) + \iota_C(\mathbf{x})\,, \tag{17}$$

where ι_C is the indicator function of a convex set C that represents additional constraints on the solution, such as positivity or box constraints. The convention is that $\iota_C(\mathbf{x})$ takes the value 0 for $\mathbf{x} \in C$ and ∞ otherwise. If $\vartheta(\mathbf{x}) = \tau\,\psi(\mathbf{x}) + \iota_C(\mathbf{x})$ is a proper, closed, convex function, then the solution of (17) is unique and corresponds to the value of the *Moreau proximity operator* [27], defined as

$$\operatorname{prox}_{\vartheta}(\mathbf{z}) = \arg\min_{\mathbf{x}\in\mathbb{R}^N} \frac{1}{2}\|\mathbf{x} - \mathbf{z}\|_2^2 + \vartheta(\mathbf{x})\,. \tag{18}$$

The proximal map of $\vartheta(\mathbf{x})$ cannot always be obtained in closed-form, and this is also the case for the regularizers under study. For this reason, we next present a primal-dual approach that results in a novel numerical algorithm, which can efficiently compute the solution.

A fundamental ingredient of our proposed algorithm is the orthogonal projection of matrices onto S_q norm balls. This projection can be performed efficiently based on the following proposition, which provides a direct link between vector projections onto ℓ_q norm balls and matrix projections onto S_q

norm balls. This result is new, to the best of our knowledge, and its proof is provided in Appendix II. A relevant result that can be considered as a converse statement of Proposition 1 can be found in [28, Theorem A.2].

Proposition 1 (Schatten Norm Projections): Let $\mathbf{Y}\in\mathbb{C}^{n_1\times n_2}$ with SVD decomposition $\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, where $\mathbf{U} \in \mathbb{U}^{n_1}$, $\mathbf{V} \in \mathbb{U}^{n_2}$ and $\mathbf{\Sigma} \in \mathbb{D}^{n_1\times n_2}$. The orthogonal projection of $\mathbf{Y}$ onto the set $B_{S_q} = \left\{\mathbf{X} \in \mathbb{C}^{n_1\times n_2} : \|\mathbf{X}\|_{S_q} \leq \rho\right\}$ is given by

$$P_{B_{S_q}}(\mathbf{Y}) = \mathbf{U}\text{diag}\left(P_{B_q}(\boldsymbol{\sigma}(\mathbf{Y}))\right)\mathbf{V}^H,$$

where diag $(\cdot)$ is the operator that maps a vector to a diagonal matrix and P_{B_q} is the orthogonal projection onto the ℓ_q norm ball $B_q = \left\{\mathbf{v} \in \mathbb{R}_+^{\min(n_1,n_2)} : \|\mathbf{v}\|_q \leq \rho\right\}$ of radius ρ.

Based on Proposition 1, we design an algorithm for the orthogonal projection of a matrix $\mathbf{Y}$ onto the convex set B_{S_q}. Our algorithm consists of three steps: (a) decompose $\mathbf{Y}$ in its singular vectors and singular values by means of the SVD; (b) project its singular values onto the corresponding ℓ_q norm ball B_q; and (c) obtain the projected matrix via singular value reconstruction (SVR) using the projected singular values and the original singular vectors.

We next describe all of the steps leading to the proposed algorithm that solves the problem

$$\arg\min_{\mathbf{x}\in C} \frac{1}{2}\|\mathbf{x}-\mathbf{z}\|_2^2 + \tau\,\|\boldsymbol{H}\mathbf{x}\|_{1,p} \;\; \forall p \geq 1\,. \tag{19}$$

With $\|\boldsymbol{H}\mathbf{x}\|_{1,p}$ we denote the discrete version of our proposed regularization family (6), where $\|\cdot\|_{1,p}$ stands for the mixed ℓ_1-S_p norm, which for an argument $\mathbf{\Psi} = \left[\mathbf{\Psi}_1^T, \mathbf{\Psi}_2^T, \ldots, \mathbf{\Psi}_N^T\right]^T \in X$ is defined as

$$\|\mathbf{\Psi}\|_{1,p} = \sum_{n=1}^{N} \|\mathbf{\Psi}_n\|_{S_p}\,, \forall p \geq 1\,. \tag{20}$$

The discrete form of our regularizers highlights their relation to the sparsity-promoting group norms, which are commonly met in the context of compressive sensing (see [29], for instance). However, a significant difference is that in our case the mixed norm is a vector-matrix norm rather than a vector-vector norm. Therefore, while the machinery we are using shares some similarities with the one employed in the group vector-norm case, there are important differences, with the most pronounced being the projection step.

Since the operator of our choice is the Hessian, which produces 2×2 symmetric matrices at every coordinate of $\mathbf{x}$, $\mathbf{\Psi}_n \in \mathbb{S}^2$ in (20), where $\mathbb{S}^2 = \left\{\mathbf{X} \in \mathbb{R}^{2\times 2} : \mathbf{X}^T = \mathbf{X}\right\}$. However, for reasons of completeness, in the following lemma, where we derive the dual of the ℓ_1-S_p norm, we consider the more general case $\mathbf{\Psi}_n \in \mathbb{C}^{n_1\times n_2}$. The proof of Lemma 1 is provided in Appendix II and follows a similar line of thought with the one presented in [30, Lemma 1]. The latter is about the dual norm of a mixed ℓ_1-ℓ_p vector norm.

Lemma 1: Let $p \geq 1$, and let q be the conjugate exponent of p, i.e., $\frac{1}{p} + \frac{1}{q} = 1$. Then, the mixed norm $\|\cdot\|_{\infty,q}$ is dual to the mixed norm $\|\cdot\|_{1,p}$.

Using Lemma 1 and noting that the dual of the dual norm is the original norm [31], we write (20) in the equivalent form

$$\|\mathbf{\Psi}\|_{1,p} = \max_{\mathbf{\Omega}\in B_{\infty,q}} \langle \mathbf{\Omega}\,,\, \mathbf{\Psi}\rangle_X\,, \tag{21}$$

where $B_{\infty,q}$ denotes the ℓ_∞-S_q unit-norm ball, defined as

$$B_{\infty,q} = \Big\{\mathbf{\Omega} = \left[\mathbf{\Omega}_1^T, \mathbf{\Omega}_2^T, \ldots, \mathbf{\Omega}_N^T\right]^T \in X : \|\mathbf{\Omega}_n\|_{S_q} \leq 1, \forall n = 1,\ldots,N\Big\}. \tag{22}$$

This alternative definition of the mixed ℓ_1-S_p norm allow us to express it in terms of an inner product that involves the dual variable $\mathbf{\Omega}$ and the unit-norm ball $B_{\infty,q}$. Moreover, from (22) it is straightforward to see that the orthogonal projection onto $B_{\infty,q}$ is obtained by projecting separately each submatrix $\mathbf{\Omega}_n$ onto a unit-norm S_q ball (B_{S_q}).

Using (21) we re-write (19) as

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in C} \frac{1}{2}\|\mathbf{x}-\mathbf{z}\|_2^2 + \tau \max_{\mathbf{\Omega}\in B_{\infty,q}} \langle \mathbf{\Omega}\,,\, \boldsymbol{H}\mathbf{x}\rangle_X\,. \tag{23}$$

This formulation naturally leads us to the following minimax problem

$$\min_{\mathbf{x}\in C}\max_{\mathbf{\Omega}\in B_{\infty,q}} L(\mathbf{x},\mathbf{\Omega})\,, \tag{24}$$

where

$$L(\mathbf{x},\mathbf{\Omega}) = \frac{1}{2}\|\mathbf{x}-\mathbf{z}\|_2^2 + \tau\left\langle \boldsymbol{H}^*\mathbf{\Omega}\,,\,\mathbf{x}\right\rangle_2\,. \tag{25}$$

Since the function $L(\mathbf{x},\mathbf{\Omega})$ is strictly convex in $\mathbf{x}$ and concave in $\mathbf{\Omega}$, we have the guarantee that a saddle-value is attained [31], and, thus, the order of the minimum and the maximum in (24) does not affect the solution. This means that there exists a saddle-point $\left(\hat{\mathbf{x}},\hat{\mathbf{\Omega}}\right)$ that leads to a common value when the minimum and the maximum are interchanged, i.e.,

$$\min_{\mathbf{x}\in C}\max_{\mathbf{\Omega}\in B_{\infty,q}} L(\mathbf{x},\mathbf{\Omega}) = L\left(\hat{\mathbf{x}},\hat{\mathbf{\Omega}}\right) = \max_{\mathbf{\Omega}\in B_{\infty,q}}\min_{\mathbf{x}\in C} L(\mathbf{x},\mathbf{\Omega})\,. \tag{26}$$

Based on this observation, we can now define the primal and dual problems by identifying the primal and dual objective functions, respectively. The l.h.s of (26) corresponds to the minimization of the primal objective function $\varrho(\mathbf{x})$, and the r.h.s to the maximization of the dual objective function $s(\mathbf{\Omega})$,

$$\varrho(\mathbf{x}) = \max_{\mathbf{\Omega}\in B_{\infty,q}} L(\mathbf{x},\mathbf{\Omega}) = \frac{1}{2}\|\mathbf{x}-\mathbf{z}\|_2^2 + \tau\,\|\boldsymbol{H}\mathbf{x}\|_{1,p}\,, \tag{27}$$

$$\begin{aligned} s(\mathbf{\Omega}) &= \min_{\mathbf{x}\in C} L(\mathbf{x},\mathbf{\Omega}) \\ &= \frac{1}{2}\|P_C(\mathbf{v})-\mathbf{v}\|_2^2 + \frac{1}{2}\|\mathbf{z}\|_2^2 - \frac{1}{2}\|\mathbf{v}\|_2^2\,, \end{aligned} \tag{28}$$

where P_C is the orthogonal projection onto the convex set C and $\mathbf{v} = \mathbf{z} - \tau\boldsymbol{H}^*\mathbf{\Omega}$. Therefore, (26) indicates that we can obtain the minimizer $\hat{\mathbf{x}}$ of $\varrho(\mathbf{x})$ from the maximizer $\hat{\mathbf{\Omega}}$ of $s(\mathbf{\Omega})$ through the relation

$$\hat{\mathbf{x}} = P_C\left(\mathbf{z} - \tau\boldsymbol{H}^*\hat{\mathbf{\Omega}}\right)\,. \tag{29}$$

This last relation is important, since in contrast to the primal problem (19), which is not continuously differentiable, the dual one involves the smooth function $s(\mathbf{\Omega})$. We can therefore

solve it by exploiting its gradient. Indeed, using the property that the gradient of a function $h(\mathbf{x}) = \|\mathbf{x} - P_C(\mathbf{x})\|_2^2$ is well defined and is equal to $\nabla h(\mathbf{x}) = 2(\mathbf{x} - P_C(\mathbf{x}))$ [4, Lemma 4.1], we compute the gradient of $s(\boldsymbol{\Omega})$ as

$$\nabla s(\boldsymbol{\Omega}) = \tau \boldsymbol{H} P_C\left(\mathbf{z} - \tau \boldsymbol{H}^* \boldsymbol{\Omega}\right). \tag{30}$$

Therefore, the solution of our primal problem (19) is obtained in two steps: (a) we find the maximizer of the dual objective function (28) as described next, and (b) we obtain the solution through (29).

E. Maximization of the Dual Objective

At this point, a main issue we need to deal with, is that the Hessian operator $\boldsymbol{H}$ does not have an empty null space and, thus, a stable inverse does not exist. Consequently, we cannot opt for a closed-form solution for the maximizer $\hat{\boldsymbol{\Omega}}$ of $s(\boldsymbol{\Omega})$. This means that we have to resort to a numerical iterative scheme. In this work, we employ Nesterov's iterative method [32] for smooth functions. This is a gradient-based scheme that exhibits convergence rates of one order higher than the standard gradient-ascent method. To ensure convergence of the algorithm, we need to choose an appropriate step-size. Since our dual objective is smooth with Lipschitz continuous gradient, we can use a constant step-size, thus, avoid a line search at every iteration. An appropriate step-size is equal to the inverse of the Lipschitz constant of $\nabla s(\boldsymbol{\Omega})$. We derive an upper bound of this Lipschitz constant in the following proposition, whose proof is given in Appendix II.

Proposition 2: Let $L(s)$ denote the Lipschitz constant of $\nabla s(\boldsymbol{\Omega})$ of the dual objective function defined in (28). Then, it holds that

$$L(s) \leq 64\tau^2. \tag{31}$$

From (26) and (28) it is clear that the maximizer of our dual objective can be derived by solving the constrained maximization problem

$$\hat{\boldsymbol{\Omega}} = \underset{\boldsymbol{\Omega} \in B_{\infty,q}}{\arg\max} \frac{1}{2}\|P_C(\mathbf{v}) - \mathbf{v}\|_2^2 - \frac{1}{2}\|\mathbf{v}\|_2^2, \tag{32}$$

with $\mathbf{v} = \mathbf{z} - \tau \boldsymbol{H}^* \boldsymbol{\Omega}$. A necessary step towards this direction is to compute the projection onto the set $B_{\infty,q}$, defined in (22). This operation is accomplished by projecting independently each of the N components $\boldsymbol{\Omega}_n$ of $\boldsymbol{\Omega}$ onto the set $B_{S_q} = \left\{\mathbf{X} \in \mathbb{S}^2 : \|\mathbf{X}\|_{S_q} \leq 1\right\}$. This projection is performed efficiently following the three steps of the algorithm we proposed in Section III-D, which is based on our Proposition 1.

Steps (a) and (c) are fairly easy to implement. Specifically, since the matrices of interest are 2×2 symmetric, we compute the SVD and the SVR steps in closed-form. Then, the most cumbersome part of our algorithm is the ℓ_q norm projection of the singular values, which for general values of q does not exist in closed form. Fortunately, this operation is still feasible thanks to the recently developed ℓ_q norm projection algorithm [30]. This projection method is based on an efficient proximity algorithm for ℓ_q norms [33]. Moreover, in Section III-F we report three cases of S_q norms, $q = 1, 2, \infty$, whose projection can be evaluated in closed-form.

F. Closed Form of S_q-Norm Projections for $q = 1, 2, \infty$.

From Proposition 1, we know that the matrix projection onto the B_{S_2} unit-norm ball is associated with the projection of the singular values of the matrix onto the B_2 ball. The latter is computed by normalizing the elements of the corresponding vector by their Euclidean norm. Therefore, we have that

$$P_{B_{S_2}}(\boldsymbol{\Omega}_n) = \begin{cases} \frac{\boldsymbol{\Omega}_n}{\|\boldsymbol{\Omega}_n\|_F} & , \text{if } \|\boldsymbol{\Omega}_n\|_F > 1 \\ \boldsymbol{\Omega}_n & , \text{if } \|\boldsymbol{\Omega}_n\|_F \leq 1. \end{cases} \tag{33}$$

This situation is advantageous since it allows us to avoid both the SVD and the SVR steps. Consequently, this drastically reduces the complexity of computing the projection. To compute the projection onto B_{S_∞}, we use that the projection onto the B_∞ unit-norm ball corresponds to setting the elements that have an absolute value greater than one to one, and adding back their original sign. Therefore, we readily get

$$P_{B_{S_\infty}}(\boldsymbol{\Omega}_n) = \mathbf{U}\text{diag}\left(\min\left(\boldsymbol{\sigma}(\boldsymbol{\Omega}_n), \mathbf{1}\right)\right)\mathbf{V}^H, \tag{34}$$

where $\mathbf{1}$ is a vector with all elements set to one and the min operator is applied component-wise. Note that, this result is directly related to the singular value thresholding (SVT) method [34], [35] developed in the field of matrix rank minimization. The derivation of SVT in [34], [35] is technical. It relies on the characterization of the subgradient of the nuclear norm [36]. By contrast, in our case the result comes out naturally as an immediate consequence of Proposition 1 and of the duality between the spectral and nuclear matrix norms. Finally, the projection of a matrix onto the B_{S_1} unit-norm ball is related to the projection of its singular values onto the B_1 unit-norm ball. The latter is computed by the soft-thresholding operator $S_\gamma(\boldsymbol{\sigma}(\boldsymbol{\Omega}_n)) = \max(\boldsymbol{\sigma}(\boldsymbol{\Omega}_n) - \gamma, 0)$ [37], where the max operator is applied component-wise. Therefore, based on Proposition 1, we have that

$$P_{B_{S_1}}(\boldsymbol{\Omega}_n) = \mathbf{U}\text{diag}\left(S_\gamma(\boldsymbol{\sigma}(\boldsymbol{\Omega}_n))\right)\mathbf{V}^H. \tag{35}$$

This last projection cannot in general be computed in closed form. The reason is that the threshold γ is not known in advance and needs to be estimated. This can be accomplised using one of the existing methods available in the literature [38]–[41]. Fortunately, in our case, the singular vectors are of low dimensionality, $\boldsymbol{\sigma}(\boldsymbol{\Omega}_n) \in \mathbb{R}_+^2$ and, thus, we derive γ analytically, as

$$\gamma = \begin{cases} 0 & , \text{if } \boldsymbol{\sigma}_1(\boldsymbol{\Omega}_n) \leq 1 - \boldsymbol{\sigma}_2(\boldsymbol{\Omega}_n), \\ \frac{\boldsymbol{\sigma}_1(\boldsymbol{\Omega}_n) + \boldsymbol{\sigma}_2(\boldsymbol{\Omega}_n) - 1}{2} & , \text{if } 1 - \boldsymbol{\sigma}_2(\boldsymbol{\Omega}_n) < \boldsymbol{\sigma}_1(\boldsymbol{\Omega}_n) \leq 1 + \boldsymbol{\sigma}_2(\boldsymbol{\Omega}_n) \\ \boldsymbol{\sigma}_1(\boldsymbol{\Omega}_n) - 1 & , \text{if } \boldsymbol{\sigma}_1(\boldsymbol{\Omega}_n) > 1 + \boldsymbol{\sigma}_2(\boldsymbol{\Omega}_n), \end{cases} \tag{36}$$

where the singular values are sorted in a decreasing order, i.e., $\boldsymbol{\sigma}_1(\boldsymbol{\Omega}_n) \geq \boldsymbol{\sigma}_2(\boldsymbol{\Omega}_n)$.

G. Numerical Algorithm

Equipped with all the necessary ingredients, we conclude with a summarized description of the complete optimization algorithm. Our method consists of two components that interact. The first component is responsible for the majorization of the objective function, as we described in Sec. III-C,

Algorithm 1 Image Reconstruction Algorithm Under Hessian-Based $\ell_1 - S_p$ Norm Regularization

Input: $\mathbf{y}$, $\mathbf{A}$, $\tau > 0$, $p \geq 1$, $\alpha > \left\|\mathbf{A}^T\mathbf{A}\right\|$, $\mathcal{P}_{\mathcal{C}}$.

Initialization: $\mathbf{u}_1 = \mathbf{x}_0$, $t_1 = 1$, $c_1 = \varphi(\mathbf{x}_0)$.

while *stopping criterion is not satisfied* **do**

$\mathbf{s}_n \leftarrow \texttt{denoise}\left(\mathbf{u}_n + \alpha^{-1}\mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{u}_n), \frac{\tau}{\alpha}, p, \mathcal{P}_{\mathcal{C}}\right)$;

$t_{n+1} \leftarrow \frac{1+\sqrt{1+4t_n^2}}{2}$;

$c_{n+1} = \varphi(\mathbf{s}_n)$;

if $c_{n+1} > c_n$ **then**

$c_{n+1} = c_n$;

$\mathbf{x}_{n+1} \leftarrow \mathbf{x}_n$;

$\mathbf{u}_{n+1} \leftarrow \mathbf{x}_n + \frac{t_n}{t_{n+1}}(\mathbf{s}_n - \mathbf{x}_n)$;

else

$\mathbf{x}_{n+1} \leftarrow \mathbf{s}_n$;

$\mathbf{u}_{n+1} \leftarrow \mathbf{s}_n + \left(\frac{t_n - 1}{t_{n+1}}\right)(\mathbf{s}_n - \mathbf{x}_n)$;

end

$n \leftarrow n + 1$;

end

return $\mathbf{x}_n$;

while the second one undertakes the minimization of the resulting upper-bounded version. Then, the algorithm proceeds by iteratively minimizing the majorizer that is formed based on the solution of the previous iteration. Since the convergence of this scheme can be slow in practice, to speed it up we employ the FISTA algorithm [24]. This method exhibits state-of-the-art convergence rates by combining two consecutive iterates, in an optimum way. A description of our image reconstruction approach that is based on the monotone version of FISTA (MFISTA) [4] is given in Algorithm 1. The sub-routine `denoise` corresponds to the second component that finds the solution of (19). This minimizer is related to the proximal map $\text{prox}_{\tau\|\boldsymbol{H}\cdot\|_{1,p}}$ but we can also interpret it as a denoising step under Hessian-based ℓ_1-S_p norm regularization. The computation of the `denoise` sub-routine is described in Algorithm 2 and is based on the primal-dual formulation that we proposed in Secs. III-D and III-E.

Finally, regarding the computational complexity of the algorithm, it is only mildly higher than that of TV's. The extra computational cost is due to (a) the use of a tensor (Hessian) instead of a vectorial (gradient) operator and (b) the projections onto the B_{S_q} balls instead of the B_2 ball. Our projections are somewhat more expensive because of the SVD and SVR steps. However, these steps are computed in closed form. It is also worth mentioning that the proposed algorithm is highly parallelizable, since all the involved operations are performed independently for each pixel of the image.

Algorithm 2 Denoise ($\mathbf{z}$, τ, p, P_c)—Denoising Algorithm Under Hessian-Based $\ell_1 - S_p$ Norm Regularization

Input: $\mathbf{z}$, $\tau > 0$, $p \geq 1$, $\mathcal{P}_{\mathcal{C}}$.

Initialization: $\boldsymbol{\Psi}_1 = \boldsymbol{\Omega}_0 = \mathbf{0} \in \mathcal{X}$, $t_1 = 1$.

Output: $\hat{\mathrm{x}}$ – optimal solution of (19).

while *stopping criterion is not satisfied* **do**

$\boldsymbol{\Omega}_n \leftarrow \mathcal{P}_{\mathcal{B}_q}\left(\boldsymbol{\Psi}_n + \frac{1}{64\tau}\boldsymbol{\mathcal{H}}\mathcal{P}_{\mathcal{C}}(\mathbf{z} - \tau\boldsymbol{\mathcal{H}}^*\boldsymbol{\Psi}_n)\right)$;

$t_{n+1} \leftarrow \frac{1+\sqrt{1+4t_n^2}}{2}$;

$\boldsymbol{\Psi}_{n+1} \leftarrow \boldsymbol{\Omega}_n + \left(\frac{t_n - 1}{t_{n+1}}\right)(\boldsymbol{\Omega}_n - \boldsymbol{\Omega}_{n-1})$;

$n \leftarrow n + 1$;

end

return $\mathcal{P}_{\mathcal{C}}(\mathbf{z} - \tau\boldsymbol{\mathcal{H}}^*\boldsymbol{\Omega}_{n-1})$;

IV. Experimental Results

To evaluate the effectiveness of our proposed regularization framework, we report results for several linear inverse imaging problems. In particular, we consider the problems of image deblurring, sparse reconstruction from random samples, image interpolation and image zooming. For the image deblurring problem we compare our results against those obtained by using three alternative methods; namely, TV regularization, regularization with the fully redundant Haar wavelet transform, and the image deblurring version of the BM3D patch-based method [42]. In Haar's case, we use the frame analysis framework since it has been reported in the literature (c.f [43]) that the frame synthesis framework usually leads to inferior results. For the rest of the inverse problems we provide comparisons against TV and quadratic derivative-based regularizers.

A. Restoration Setting

For the image deblurring experiments, we use a set of 8 grayscale images[1] which have been normalized so that their intensities lie in the range of $[0\,,1]$.

The performance of the methods under comparison is assessed for various blurring kernels and different noise levels. In particular, in our experiments we employ three *point spread functions* (PSFs) to produce blurred versions of the images. We use a Gaussian PSF of standard deviation $\sigma_b = 4$, a moving average (uniform) PSF, and a motion-blur kernel. The first two PSFs have a support of 9×9 pixel while the third one has a support of 19×19 pixel. As an additional degradation factor we consider Gaussian noise of three noise levels corresponding to a blurred signal-to-noise-ratio (BSNR) of $\{15, 20, 25\}$ dB, respectively. The BSNR is defined as $\text{BSNR} = \text{var}(\mathbf{Ax})/\sigma_w^2$

[1]Three of these images along with the motion-blur kernel used in the experiments were obtained from http://www.wisdom.weizmann.ac.il/~levina/papers/LevinEtalCVPR09Data.rar.

TABLE I
ISNR COMPARISONS OF IMAGE RESTORATION FOR THREE BLURRING KERNELS AND THREE NOISE LEVELS

	Image	*Boat*			*Face*			*Fluor. Cells*			*Hill*			*House*			*Kids*			*Lena*			*Peppers*		
	BSNR	15 dB	20 dB	25 dB	15 dB	20 dB	25 dB	15 dB	20 dB	25 dB	15 dB	20 dB	25 dB	15 dB	20 dB	25 dB	15 dB	20 dB	25 dB	15 dB	20 dB	25 dB	15 dB	20 dB	25 dB
Gaussian Blur	Haar	2.34	2.77	3.71	2.37	2.80	3.88	2.36	2.41	3.19	2.61	2.48	2.95	2.56	3.23	4.27	3.20	3.50	4.45	2.84	2.67	3.20	2.30	3.43	5.10
	TV	2.59	2.95	3.81	3.60	3.96	4.78	2.84	2.84	3.48	2.85	2.65	3.07	2.91	3.49	4.45	4.15	4.25	4.93	3.52	3.25	3.64	3.24	4.31	5.69
	HS_∞	2.63	3.03	3.88	4.32	4.71	5.62	3.61	3.62	4.32	2.95	2.74	3.22	3.28	3.92	4.92	4.33	4.61	5.42	3.60	3.42	3.91	3.52	4.35	5.59
	HS_2	2.71	3.12	4.00	4.43	4.84	5.80	3.66	3.69	4.40	2.99	2.80	3.29	3.33	3.99	**5.00**	4.41	4.70	5.51	3.67	3.51	4.01	3.63	4.52	5.78
	HS_1	2.75	3.18	4.07	4.45	4.88	5.84	3.65	3.70	**4.41**	3.00	2.82	3.32	3.33	3.97	4.99	4.42	**4.70**	**5.53**	3.71	3.55	4.06	3.71	4.67	**5.93**
	BM3D	**2.88**	**3.37**	**4.32**	**4.75**	**5.14**	**6.08**	**3.70**	**3.75**	4.39	**3.04**	**3.00**	**3.51**	**3.51**	**4.07**	4.92	**4.51**	4.66	5.27	**3.78**	**3.70**	**4.24**	**3.79**	**4.76**	5.91
Uniform Blur	Haar	3.04	3.69	4.73	2.84	3.55	4.76	2.90	3.22	4.14	3.13	3.19	3.89	3.35	4.23	5.49	3.85	4.40	5.51	3.21	3.32	4.03	3.47	4.89	6.37
	TV	3.20	3.80	4.77	3.71	4.30	5.41	3.26	3.50	4.30	3.27	3.30	3.93	3.57	4.35	5.50	4.48	4.95	5.87	3.79	3.78	4.36	4.32	5.84	7.26
	HS_∞	3.22	3.84	4.81	4.38	4.95	6.16	3.92	4.20	5.04	3.32	3.39	4.06	3.87	4.78	6.00	4.62	5.13	6.15	3.85	3.89	4.52	4.39	5.75	7.03
	HS_2	3.32	3.95	4.96	4.48	5.08	6.35	3.98	4.27	5.12	3.38	3.46	4.15	3.94	4.87	**6.10**	4.74	5.25	6.27	3.95	4.00	4.62	4.55	5.91	7.20
	HS_1	3.36	4.02	5.05	4.51	5.13	6.41	**3.99**	4.28	**5.13**	3.40	3.50	4.18	3.95	4.88	6.09	**4.77**	**5.29**	**6.30**	4.00	4.04	4.68	4.66	**6.03**	**7.32**
	BM3D	**3.48**	**4.19**	**5.26**	**4.97**	**5.80**	**6.81**	3.97	**4.29**	5.06	**3.53**	**3.68**	**4.35**	**4.05**	**4.91**	5.97	4.70	5.03	5.91	**4.06**	**4.18**	**4.89**	**4.76**	5.70	6.78
Motion Blur	Haar	3.89	5.11	6.92	4.33	5.75	7.74	3.70	4.60	6.30	3.55	4.10	5.46	4.19	5.82	7.99	4.91	5.95	7.77	3.97	4.42	5.83	4.52	6.50	8.74
	TV	3.88	5.02	6.78	4.61	5.99	7.93	3.86	4.68	6.32	3.56	4.05	5.39	4.20	5.76	7.86	5.10	6.07	7.77	4.22	4.58	5.92	5.28	7.39	9.01
	HS_∞	3.80	5.09	6.94	5.80	7.32	9.26	4.52	5.50	7.23	3.69	4.22	5.51	4.75	6.50	8.65	5.35	6.57	8.52	4.34	4.88	6.29	5.19	7.09	8.68
	HS_2	3.94	5.26	7.12	5.92	7.48	9.44	4.62	5.63	7.36	3.79	4.34	5.63	4.86	6.64	**8.80**	5.47	6.72	**8.69**	4.48	5.03	6.45	5.38	7.31	8.92
	HS_1	4.03	5.35	7.19	5.96	7.53	9.49	4.64	5.66	**7.37**	3.84	4.39	5.64	4.86	**6.65**	8.78	5.50	**6.75**	8.67	4.56	5.11	6.51	5.52	**7.48**	**9.09**
	BM3D	**4.46**	**5.81**	**7.51**	**6.75**	**8.04**	**9.62**	**4.79**	**5.78**	7.31	**4.05**	**4.60**	**5.75**	**5.09**	6.63	8.43	**5.70**	6.66	8.27	**5.09**	**5.82**	**7.14**	**6.14**	7.46	9.00

where var $(\mathbf{Ax})$ is the variance of the blurred image and σ_w is the standard deviation of the noise.

Regarding the restoration task, for the methods that involve the minimization of an objective function this is performed under the constraint that the restored intensities must lie in the convex set $C = \{\mathbf{x} \in \mathbb{R}^N | x_n \in [0, 1] \forall n = 1, \ldots, N\}$. To accomplish that, we use the corresponding projection operation, P_C, which for a vector $\mathbf{x}$ amounts to setting the elements that are less than zero and greater to one, to zero and one, respectively. For the Hessian-based functionals we use the minimization method proposed in Section III, while for TV- and Haar-based ones we employ the algorithm of [4]. The latter belongs to the same category of minimization algorithms as ours with a comparable convergence behavior. The rationale for this choice is that, the quality of the restoration will not depend on the choice of the minimization strategy but rather on the choice of the regularizer. In all cases, the stopping criterion is set to either reaching a relative normed difference of 10^{-5} between two successive estimates, or a maximum of 100 MFISTA iterations. We also use 10 inner iterations for the solution of the corresponding denoising problem. Moreover, instead of using the true PSF that produces the blurred images, we use a slightly perturbed version by adding Gaussian noise of standard deviation 10^{-3}. The motivation is to test the performance of the algorithms under more realistic conditions, since, in practice the employed PSF normally contains some error and thus deviates from the true one. Finally, in all the reported experiments the quality of the reconstruction is evaluated in terms of an increase in the SNR (ISNR), measured in dB. The ISNR is defined as $\text{ISNR} = 10 \log_{10} (\text{MSE}_{\text{in}}/\text{MSE}_{\text{out}})$, where MSE_{in} and MSE_{out} are the mean-squared errors between the degraded and the original image, and the restored and the original image, respectively.

B. Image Restoration on Standard Test Images

In Table I we provide comparative restoration results for all the test images and all the combinations of degradation conditions (PSF and noise level). To distinguish between the different Hessian-based regularizers, we refer to them as HS_k with k denoting the order of the Schatten norm. We report the results obtained by using Schatten norms of order one, two and infinity, which correspond to the well-known nuclear, Frobenius and spectral matrix norms, respectively. For the sake of consistency among comparisons, the reported results for each regularizer, including Haar and TV, are derived using the individualized (w.r.t. degradation conditions) regularization parameter τ, that gives the best ISNR performance. The results of the BM3D algorithm are also optimized by providing the true standard deviation of the Gaussian noise.

On average, for all the tested images and degradation conditions, the BM3D algorithm produces the best PSNR scores. However, despite its non-adaptive nature, our regularization scheme manages to provide comparable results. Regarding comparisons among the regularization techniques, the Hessian-based framework leads to improved quantitative results compared to those of Haar and TV. The best SNR improvement, on average, is achieved for the HS_1 regularizer, while comparable results are also obtained for the HS_2 regularizer. While the HS_∞ regularizer outperforms Haar and TV most of the time, the improvement is less pronounced than that of the other two regularizers. We can thus conclude, that as the order of the Schatten norm approaches one, the reconstruction results improve. This can be attributed to the fact that, in the extreme case of order infinity, the corresponding regularizer takes into account only the maximum absolute eigenvalue and thus fails to include additional information possibly provided by the second eigenvalue. Overall, the improvement in performance over Haar and TV can be quite substantial (more than 0.5 dB), which justifies Hessian-based regularization as a viable alternative approach.

Beyond the ISNR comparisons, the effectiveness of the proposed method can also be visually appreciated by inspecting the representative *Face*, *Kids* and *House* deblurring examples of Figures 1–3. From these examples we can verify our initial claims, that TV regularization leads to image reconstructions

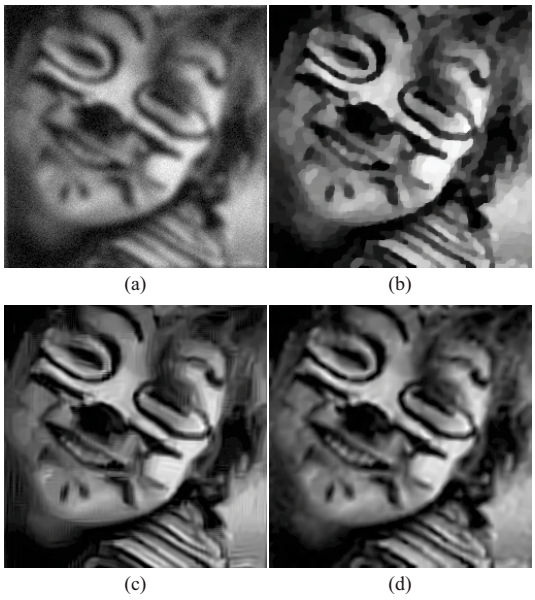

Fig. 1. Restoration of the *Face* image degraded by Gaussian blurring and Gaussian noise corresponding to a BSNR level of 15 dB. (a) Degraded image (PSNR = 21.76 dB), (b) TV result (PSNR = 25.36 dB), (c) BM3D result (PSNR = 26.51 dB), (d) HS_1 result (PSNR = 26.21 dB).

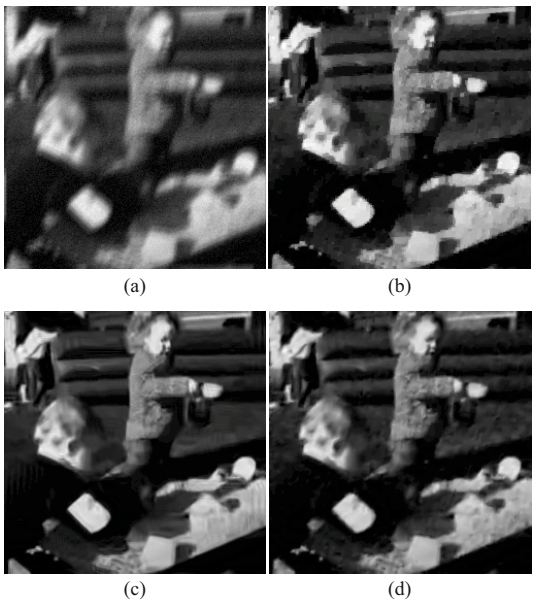

Fig. 2. Restoration of the *Kids* image degraded by motion blurring and Gaussian noise corresponding to a BSNR level of 20 dB. (a) Degraded image (PSNR = 21.84 dB), (b) Haar result (PSNR = 27.79 dB), (c) BM3D result (PSNR = 28.50 dB), (d) HS_1 result (PSNR = 28.59 dB).

that suffer from the presence of heavy block artifacts. These artifacts become more evident in regions where the image is characterized by smooth intensity transitions, and they are responsible for shuffling details of the image and broadening its fine structures. See for example the TV solution in Fig. 1, where the image has cartoon-like appearance. Similar blocking effects, which are even more pronounced, appear on the Haar-based reconstructions. On the other hand, even in cases where the presence of the noise is significant, the Hessian-based regularizers manage to avoid introducing pronounced artifacts and thus, they lead to reconstructions that are more faithful representations of the original content of the image. Comparing our results with those of BM3D, we note that even in cases where the final PSNR favors the latter reconstruction, such as in Fig. 1, our restored images have certain advantages. For example, by a careful inspection of Figs. 1 and 2, one can clearly observe the presence of ripple-like artifacts in the BM3D solutions which do not appear in the Hessian-based reconstructions.

C. Deblurring of Biomedical Images

Our interest in image restoration is mostly motivated by the problem of microscopy image deblurring. In widefield microscopy, the acquired images are degraded by out-of-focus blur due to the poor localization of the microscope's PSF. This severely reduces our ability to clearly distinguish fine specimen structures. Since a widefield microscope can be modeled in intensity as a linear space-invariant system [44], the adopted forward model in (7) is still valid and we can, thus, employ the proposed framework for the restoration of the underlying biomedical images.

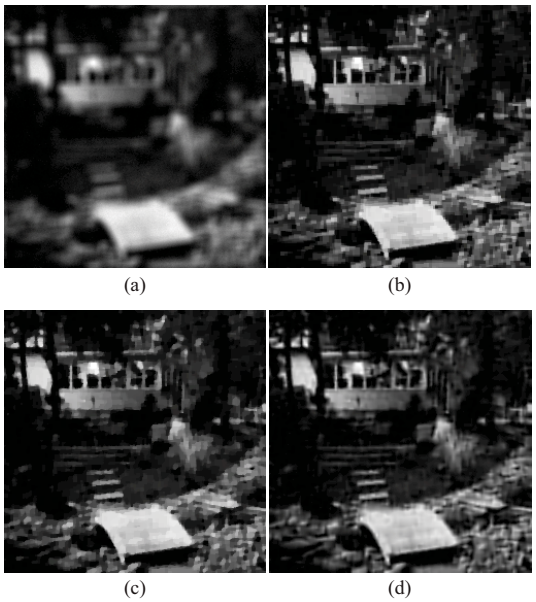

Fig. 3. Restoration of the *House* image degraded by uniform blurring and Gaussian noise corresponding to a BSNR level of 25 dB. (a) Degraded image (PSNR = 20.43 dB), (b) Haar result (PSNR = 25.92 dB), (c) TV result (PSNR = 25.93 dB), (d) HS_2 result (PSNR = 26.53 dB).

To evaluate the practical relevance of our approach, we provide deblurring experiments on two real images of fluorescence specimens. For each sample we acquired two image-stacks using a confocal microscope. This type of microscope can reject out-of-focus light using a small aperture in front of the detector and can thus avoid the blurring effect, but at the expense of more measurement noise. When the aperture is opened, the intensity of the incoming light is increased and

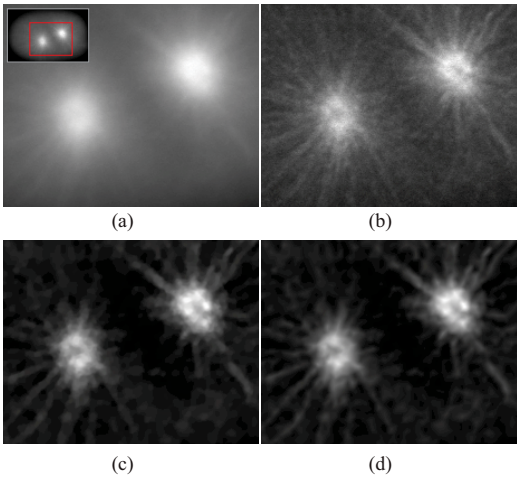

Fig. 4. Restoration results on a real *Fluorescent Cell* image of size 352 × 512. Close-up of (a) widefield image, (b) reference confocal image, (c) TV reconstruction, and (d) HS_1 reconstruction. The details of this figure are better seen in the electronic version of this paper by by zooming in on the screen.

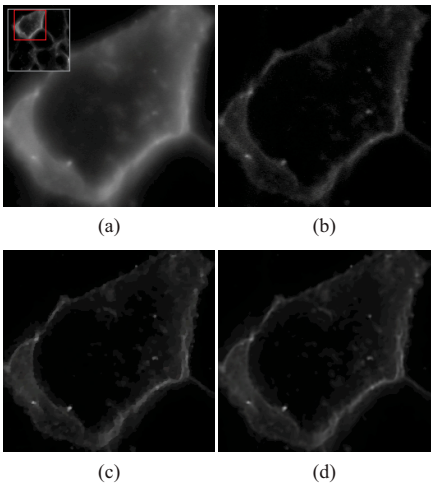

Fig. 5. Restoration results on a real *Fluorescent Cell* image of size 512 × 512. Close-up of (a) widefield image, (b) reference confocal image, (c) TV reconstruction, and (d) HS_2 reconstruction. The details of this figure are better seen in the electronic version of this paper by zooming in on the screen.

the SNR is improved, but this time the measurements include interference from adjacent out-of-focus objects. In this case the final result is blurred and it is equivalent to an image acquired by a "cheaper" widefield microscope. For more details on the image acquisition we refer to [45].

The reported results refer to the restoration of the second type of image-stacks, with the first ones serving as visual references to evaluate the quality of the reconstruction. The size of the image-stacks for the first specimen shown in Figs. 4(a) and 4(b) are $352 \times 512 \times 96$ while the size of the image stacks for the second sample shown in Figs. 5(a) and 5(b) are $512 \times 512 \times 16$. From each of the image-stacks we obtained a single image to work with, by computing the average intensity with respect to the z-axis. We did the same to obtain a 2D PSF out of a standard diffraction-limited 3D PSF model using the nominal optical parameters of the microscope (numerical aperture, wavelength, optical zoom) [44].

In Figs. 4(c) and 4(d) we present the restored images using TV and HS_1 regularization, while in Figs. 5(c) and 5(d) we provide the restored images using TV and HS_2 regularization. From these two examples, if we compare the obtained results with the confocal acquisitions, we can verify that the Hessian-based solutions are quite successful in revealing the primary features of the specimens without introducing severe artifacts, as opposed to TV which oversmooths certain features and wipes out important details of the image structure. Therefore, we conclude that our regularizers can do a better job, especially when one has to deal with images that consist mostly of ridges and filament-like structures, as is often the case in biomedical imaging.

D. Sparse Image Reconstruction

In sparse image reconstruction the observed image $\mathbf{y}$ is degraded by a masking operator which randomly sets pixel values to zero. This operator corresponds to a diagonal matrix $\mathbf{A}$ whose diagonal entries are randomly set to zero or one. We refer to this problem as sparse because in our experiments we consider masking operators that retain only $2\%, 5\%, 8\%$ and 10% of the initial pixel values. Note that this problem can be considered as compressive sensing if we assume that the image is sparse in the spatial domain.

The reported experiments are conducted on the gray-level images: Boat, Hill, Lena and Peppers. The masked images are then reconstructed using our regularizers as before plus TV and two quadratic regularizers based on the gradient and the Laplacian operators, respectively. In this setting we do not consider any presence of noise and thus, for all the regularizers under comparison, we use the same regularization parameter $\tau = 10^{-4}$. The value of τ is chosen to be small to ensure that the results will be consistent, in the sense that the reconstruction methods will not alter the unmasked pixel values. However, due to the small value of the regularization parameter, we have observed that the convergence of the minimization task for all the regularizers can be slow and thus more than 100 iterations are required. To cope with this problem, for the non-quadratic regularizers, we apply a simple continuation scheme that significantly speeds up the convergence: We start with a large value for τ and then we gradually decrease it to reach the chosen value. We observe experimentally that following this strategy and using 200 MFISTA iterations (we still solve the corresponding denoising problems using 10 iterations) we can solve the problem to high accuracy. Regarding the two quadratic regularizers, we minimize their objective functions using the conjugate gradient method [46] with a maximum of 2000 iterations.

In Table II we report the reconstruction results we obtained for all the employed regularization techniques. The quality of the reconstructions is measured in terms of PSNR. As we can observe from this table, TV does not perform well in this problem and its reconstructions fall far behind, even from the two quadratic regularization techniques. On the other hand, our Hessian-based regularizers behave much better and in all cases they lead to estimates that outperform the other methods.

TABLE II
PSNR COMPARISONS ON SPARSE IMAGE RECONSTRUCTION FROM RANDOM SAMPLES FOR FOUR RATIOS OF OBSERVED PIXELS

	Observed Pixels %	Grad.	Lap.	TV	HS_∞	HS_2	HS_1
Boat	2%	21.29	20.83	18.52	21.54	**21.56**	21.55
	5%	22.96	22.87	21.22	23.27	23.31	**23.33**
	8%	23.92	24.03	22.43	24.27	24.33	**24.37**
	10%	24.51	24.67	23.00	24.92	25.00	**25.06**
Hill	2%	22.53	22.39	19.68	22.97	23.00	**23.01**
	5%	24.27	24.33	22.13	24.70	24.76	**24.77**
	8%	25.34	25.49	23.69	25.75	25.79	**25.80**
	10%	25.90	26.07	24.52	26.29	26.34	**26.35**
Lena	2%	22.33	22.79	18.89	23.23	23.34	**23.38**
	5%	24.60	25.42	22.42	25.56	25.70	**25.82**
	8%	25.98	26.96	24.37	26.96	27.11	**27.24**
	10%	26.63	27.64	25.28	27.62	27.77	**27.90**
Peppers	2%	18.58	18.50	15.68	19.20	19.27	**19.32**
	5%	20.73	20.93	18.31	21.53	21.64	**21.71**
	8%	21.71	22.21	19.97	22.49	22.60	**22.66**
	10%	22.41	22.90	20.92	23.18	23.32	**23.39**

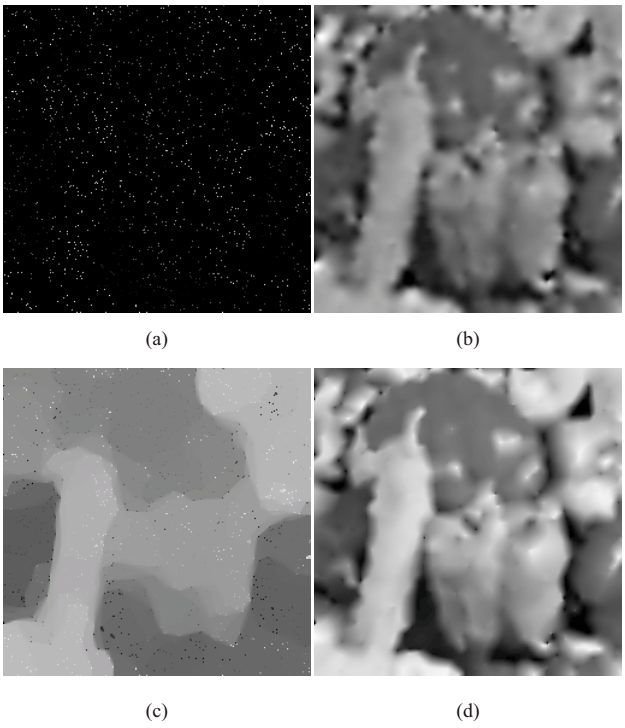

Fig. 6. Sparse reconstruction of the *Peppers* image from 2% observed pixels. (a) Masked image. (b) Laplacian-based quadratic result (PSNR = 18.50 dB). (c) TV result (PSNR = 15.68 dB). (d) HS_1 result (PSNR = 19.32 dB).

As in the image restoration case, the HS_1 regularizer leads to the best reconstructions while the HS_2 regularizer follows rather closely. In Fig. 6 we present a representative example of the reconstruction of the Peppers image from 2% observed pixels. From this example it is clear that TV cannot produce an acceptable result but instead leads to a piecewise constant solution that does not reveal any features of the image. On the other hand, both the quadratic and the proposed regularizer provide meaningful reconstructions with the latter achieving a better performance.

TABLE III
PSNR COMPARISONS ON IMAGE INTERPOLATION AND IMAGE ZOOMING FOR A $4\times$ DOWNSAMPLING FACTOR

	Images	Grad.	Lap.	TV	HS_∞	HS_2	HS_1
Interpolation	*Boat*	24.03	24.19	21.90	24.43	24.45	**24.45**
	Hill	25.48	25.62	23.31	25.83	**25.84**	25.83
	Lena	26.57	27.74	23.95	27.79	27.89	**27.92**
	Peppers	21.95	22.13	19.48	22.55	22.63	**22.68**
Zooming	*Boat*	25.95	25.89	26.00	26.01	26.08	**26.12**
	Hill	27.32	27.27	27.22	27.31	27.34	**27.35**
	Lena	29.25	29.26	29.50	29.40	29.54	**29.61**
	Peppers	24.10	24.13	24.32	24.69	24.76	**24.77**

E. Image Interpolation and Image Zooming

Image interpolation and image zooming fall into the same class of linear inverse problems. As in the sparse image reconstruction case, the degradation is due to a masking operator that zeros out some of the image pixel values. However, in these two cases the masking operator corresponds to subsampling and is highly structured, as opposed to the random masking operator. The difference between the two considered forward models is that image interpolation involves only the subsampling operator and therefore results in observed images that suffer from aliasing, while image zooming involves additionally an antialiasing operator that is applied to the underlying image before the subsampling takes place. In the last case, the matrix $\mathbf{A}$ can be expressed as $\mathbf{A} = \mathbf{SF}$ where $\mathbf{F}$ corresponds to the filtering operation and $\mathbf{S}$ to subsampling. Once more, we do not consider any presence of noise and we thus use the same regularization parameter and minimization strategy as above. The experiments we present are conducted on the same four images as in Section IV-D, using the same regularizers for a downsampling factor of 4. Finally, regarding the antialiasing filter we use a Gaussian kernel of support 9×9 and standard deviation $\sigma_b = 1.4$.

In Table III we report the obtained results and we evaluate the quality of the estimates in terms of PSNR. Regarding the interpolation problem we observe that TV, similarly to the sparse image reconstruction case, does not perform well and produces the worst scores. However, its performance gets significantly better in the image zooming case where an antialising filtering is applied. This is an indication that TV cannot perform at a satisfactory level when the operator acting on the image does not involve a mixing effect. On the other hand, the performance of the proposed regularizers is more robust to the nature of the degradation operator, and they lead to the best reconstructions. To have a visual performance assessment, we present in Figs. 7 and 8 interpolation and zooming results on the Lena and Boat

(a) (b) (c) (d) (e)

Fig. 7. Image interpolation. Close-up of (a) High-resolution image, (b) low-resolution image, (c) Laplacian-based quadratic result (PSNR=27.74 dB), (d) TV result (PSNR=23.95 dB), and (e) HS_1 result (PSNR=27.92 dB).

(a) (b) (c) (d) (e)

Fig. 8. Image zooming. Close-up of (a) High-resolution image, (b) low-resolution image, (c) gradient-based quadratic result (PSNR = 25.95 dB), (d) TV result (PSNR = 26.00 dB), and (e) HS_2 result (PSNR = 26.08 dB).

image, respectively. These results confirm our previous conclusions about the performance of TV and demonstrate the superiority of the Hessian-based regularizers over the other regularizers.

V. Conclusion

In this paper we introduced a new family of convex non-quadratic regularizers that can potentially lead to improved results in inverse imaging problems. These regularizers incorporate second-order information of the image and depend on the Schatten norms of the Hessian. We further designed an efficient and highly parrallelizable projected gradient algorithm for minimizing the corresponding objective functions. We also presented a new result that relates vector projections onto ℓ_q norm balls and matrix projections onto Schatten norm balls. This enabled us to design a matrix-projection method, which is a fundamental ingredient of our optimization algorithm.

The performance and practical relevance of the proposed regularization scheme was assessed for several linear inverse imaging problems, through comparisons on simulated and real experiments with various competing methods. The results we obtained are promising and competitive both in terms of SNR improvement and visual quality.

Appendix I

A. Proof of Theorem 1

By taking the domain Ω to be a disk, the rotation invariance of $R(f)$ implies that

$$R\left(f\left(\mathbf{R}_\theta \cdot\right)\right) = R\left(f\right) \tag{37}$$

where $\mathbf{R}_\theta$ is a rotation matrix. In particular, (37) must hold for all functions, including those of the form: $f(\mathbf{r}) = \alpha r_1 + \beta r_2$, with $\mathbf{r} \in \mathbb{R}^2$ and $\alpha, \beta \in \mathbb{R}$. Their gradient is constant and equal to $\nabla f(\mathbf{r}) = \left[\begin{smallmatrix}\alpha\\ \beta\end{smallmatrix}\right] = \mathbf{x}$. Now, using f as defined above, we write the l.h.s of (37) as

$$\begin{aligned} R\left(f\left(\mathbf{R}_\theta \cdot\right)\right) &= \int_\Omega \Phi\left(\nabla\left\{f\left(\mathbf{R}_\theta \cdot\right)\right\}(\mathbf{r})\right) \mathrm{d}\mathbf{r} \\ &= \int_\Omega \Phi\left(\mathbf{R}_\theta^T \nabla f\left(\mathbf{R}_\theta \mathbf{r}\right)\right) \mathrm{d}\mathbf{r} = \int_\Omega \Phi\left(\mathbf{R}_\theta^T \mathbf{x}\right) \mathrm{d}\mathbf{r} \\ &= \int_\Omega \Phi\left(|\mathbf{x}| \cdot \mathbf{u}_{\theta'}\right) \mathrm{d}\mathbf{r}\,, \end{aligned} \tag{38}$$

where $\mathbf{u}_{\theta'} = \left[\sin\left(\theta'\right) \sin\left(\theta' + \frac{\pi}{2}\right)\right]^T$ and $\theta' = \theta + \operatorname{sgn}(\alpha) \arccos\left(\frac{\beta}{\sqrt{\alpha^2+\beta^2}}\right)$. Setting $\theta' = \frac{\pi}{2}$ in (38) and combining the result with Property (37), we immediately get that

$$\Phi(\mathbf{x}) = \Phi\left(|\mathbf{x}|\right), \forall \mathbf{x} \in \mathbb{R}^2\,. \tag{39}$$

The scaling invariance of R can be restated as

$$\begin{aligned} R_a\left(f\left(a\cdot\right)\right) &= a^\mu R\left(f\right) \\ \int_{\Omega/a} \Phi\left(\nabla\left\{f\left(a\cdot\right)\right\}(\mathbf{r})\right) \mathrm{d}\mathbf{r} &= a^\mu \int_\Omega \Phi\left(\nabla f\left(\mathbf{r}\right)\right) \mathrm{d}\mathbf{r} \end{aligned} \tag{40}$$

for some $a > 0$, and an exponent $\mu \in \mathbb{R}$. This property must hold for all functions, including those of the form: $f(\mathbf{r}) = \alpha r_1$, with $\mathbf{r} \in \mathbb{R}^2$ and $\alpha \in \mathbb{R}$. The magnitude of their gradient is constant and equal to $|\nabla f(\mathbf{r})| = |\alpha|$. Now, using f as defined above and the result of (39), we write (40) as

$$\int_{\Omega/a} \Phi\left(a\,|\alpha|\right) \mathrm{d}\mathbf{r} = a^\mu \int_\Omega \Phi\left(|\alpha|\right) \mathrm{d}\mathbf{r}\,. \tag{41}$$

Therefore, we directly have that

$$\Phi\left(a\left|\alpha\right|\right) = a^{\nu}\Phi\left(\left|\alpha\right|\right), \forall \alpha \in \mathbb{R}, \tag{42}$$

with $\nu = \mu + 2$. Now, we define the function

$$\Phi_0(\alpha) = \frac{\Phi(|\alpha|)}{|\alpha|^{\nu}}, \forall \alpha \in \mathbb{R} \tag{43}$$

which is homogeneous of degree 0. This implies that $\Phi_0(\alpha) = c$, with c an arbitrary constant. Therefore, the potential functions Φ satisfying (42) are necessarily of the form: $\Phi(\cdot) = c\left|\cdot\right|^{\nu}$.

The inverse statement of the theorem can be verified by substitution, using the property

$$\nabla\left\{f\left(a\cdot\right)\right\}\left(\mathbf{r}\right) = a\nabla\left\{f\right\}\left(a\mathbf{r}\right), \ \forall f. \tag{44}$$

B. Proof of Theorem 2

By taking the domain $\mathbf{\Omega}$ to be a disk, the rotation invariance of $R(f)$, as defined in (37), must hold for all functions, including those of the form: $f(\mathbf{r}) = \frac{\alpha^2}{2}r_1^2 + \frac{\beta^2}{2}r_2^2 + \gamma\, r_1 r_2$, with $\mathbf{r} \in \mathbb{R}^2$ and $\alpha, \beta, \gamma \in \mathbb{R}$. Their Hessian is constant and equal to $Hf(\mathbf{r}) = \left[\begin{smallmatrix}\alpha & \gamma \\ \gamma & \beta\end{smallmatrix}\right] = \mathbf{A}$. Now, using f as defined above, we write the l.h.s of (37) as

$$\begin{aligned} R\left(f\left(\mathbf{R}_{\theta}\cdot\right)\right) &= \int_{\Omega}\Phi\left(H\left\{f\left(\mathbf{R}_{\theta}\cdot\right)\right\}\left(\mathbf{r}\right)\right)\mathrm{d}\mathbf{r} \\ &= \int_{\Omega}\Phi\left(\mathbf{R}_{\theta}^{T}Hf\left(\mathbf{R}_{\theta}\mathbf{r}\right)\mathbf{R}_{\theta}\right)\mathrm{d}\mathbf{r} \\ &= \int_{\Omega}\Phi\left(\mathbf{R}_{\theta}^{T}\mathbf{A}\mathbf{R}_{\theta}\right)\mathrm{d}\mathbf{r}\,. \end{aligned} \tag{45}$$

According to the spectral decomposition theorem, $\mathbf{A}$ being symmetric has an eigenvalue decomposition. This implies that there exists a rotation θ' such that $\mathbf{R}_{\theta'}^{T}\mathbf{A}\mathbf{R}_{\theta'} = \mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is a diagonal matrix consisting of the eigenvalues of $\mathbf{A}$. These are denoted as λ_k, where $k = 1, 2$. Based on this observation and combining it with Property (37), we immediately get that

$$\Phi\left(\mathbf{A}\right) = \Phi\left(\lambda_1, \lambda_2\right), \forall \mathbf{A} \in \mathbb{S}^2, \tag{46}$$

which implies that Φ should be a function of the Hessian eigenvalues.

The scaling invariance of R can be restated as

$$\begin{aligned} R_a\left(f\left(a\cdot\right)\right) &= a^{\mu}R\left(f\right) \\ \int_{\Omega/a}\Phi\left(H\left\{f\left(a\cdot\right)\right\}\left(\mathbf{r}\right)\right)\mathrm{d}\mathbf{r} &= a^{\mu}\int_{\Omega}\Phi\left(Hf\left(\mathbf{r}\right)\right)\mathrm{d}\mathbf{r} \end{aligned} \tag{47}$$

for some $a > 0$, and an exponent $\mu \in \mathbb{R}$. This property must hold for all functions, including those of the form: $f(\mathbf{r}) = \frac{\alpha}{2}r_1^2 + \frac{\beta}{2}r_2^2$, with $\mathbf{r} \in \mathbb{R}^2$ and $\alpha, \beta \in \mathbb{R}$. Their Hessian is constant and equal to $Hf(\mathbf{r}) = \left[\begin{smallmatrix}\alpha & 0 \\ 0 & \beta\end{smallmatrix}\right]$. Now, using f as defined above and the result of (46), we write (47) as

$$\int_{\Omega/a}\Phi\left(a^2\alpha, a^2\beta\right)\mathrm{d}\mathbf{r} = a^{\mu}\int_{\Omega}\Phi\left(\alpha, \beta\right)\mathrm{d}\mathbf{r}\,. \tag{48}$$

Therefore, we have that

$$\Phi\left(a^2\alpha, a^2\beta\right) = a^{\mu+2}\Phi\left(\alpha, \beta\right), \ \forall \alpha, \beta \in \mathbb{R}\,. \tag{49}$$

Now, we define the function

$$\Phi_0(\mathbf{x}) = \frac{\Phi\left(\mathbf{x}\right)}{\left\|\mathbf{x}\right\|_p^{\nu}}, \forall \mathbf{x} \in \mathbb{R}^2, \tag{50}$$

where $p \geq 1$ and $\nu = \frac{\mu+2}{2}$. Φ_0 is homogeneous of degree 0 and thus $\Phi_0\left(\mathbf{x}\right) = \Phi_0\left(\mathbf{x}/\left\|\mathbf{x}\right\|_p\right)$. Therefore, the potential functions Φ that satisfy (49), are necessarily of the form: $\Phi\left(\mathbf{x}\right) = \Phi_0\left(\mathbf{x}/\left\|\mathbf{x}\right\|_p\right)\left\|\mathbf{x}\right\|_p^{\nu}$. Finally, since $\mathbf{x}$ represents the vector of the eigenvalues of the Hessian, its ℓ_p norm corresponds to the S_p norm of the Hessian itself.

The inverse statement of the theorem can be verified by substitution, using the property

$$H\left\{f\left(a\cdot\right)\right\}\left(\mathbf{r}\right) = a^2 H\left\{f\right\}\left(a\mathbf{r}\right), \ \forall f. \tag{51}$$

APPENDIX II

A. Adjoint of the Discrete Hessian Operator

To find the adjoint of the discrete Hessian operator, we exploit the relation of the inner products of the spaces $\mathbb{R}^N$ and X in (13). Using (11), we can equivalently write (13) as

$$\sum_{n=1}^{N}\mathrm{tr}\left(\left[\boldsymbol{H}\mathbf{x}\right]_n^T\mathbf{Y}_n\right) = \sum_{n=1}^{N}x_n\left[\boldsymbol{H}^*\mathbf{Y}\right]_n\,. \tag{52}$$

We then expand the l.h.s of (52), to obtain

$$\begin{aligned} &\sum_{n=1}^{N}\mathrm{tr}\left(\left[\boldsymbol{H}\mathbf{x}\right]_n^T\mathbf{Y}_n\right) = \sum_{n=1}^{N}\Big(\left[\boldsymbol{H}\mathbf{x}\right]_n^{(1,1)}\mathbf{Y}_n^{(1,1)} + \\ &\left[\boldsymbol{H}\mathbf{x}\right]_n^{(1,2)}\left(\mathbf{Y}_n^{(1,2)} + \mathbf{Y}_n^{(2,1)}\right) + \left[\boldsymbol{H}\mathbf{x}\right]_n^{(2,2)}\mathbf{Y}_n^{(2,2)}\Big) \\ &= \sum_{n=1}^{N}\Big(\left[\Delta_{r_1r_1}\mathbf{x}\right]_n\mathbf{Y}_n^{(1,1)} + \left[\Delta_{r_1r_2}\mathbf{x}\right]_n\left(\mathbf{Y}_n^{(1,2)} + \mathbf{Y}_n^{(2,1)}\right) \\ &+ \left[\Delta_{r_2r_2}\mathbf{x}\right]_n\mathbf{Y}_n^{(2,2)}\Big) \\ &= \sum_{n=1}^{N}x_n\Big(\left[\Delta_{r_1r_1}^*\mathbf{Y}^{(1,1)}\right]_n + \left[\Delta_{r_1r_2}^*\left(\mathbf{Y}^{(1,2)} + \mathbf{Y}^{(2,1)}\right)\right]_n \\ &+ \left[\Delta_{r_2r_2}^*\mathbf{Y}^{(2,2)}\right]_n\Big)\,. \end{aligned} \tag{53}$$

Note that $\mathbf{Y}^{(i,j)}$ corresponds to the vector that is composed of the (i, j) entries of all $\mathbf{Y}_n \in \mathbb{R}^{2\times 2}$ matrices. Now, by comparing the r.h.s of (52) to the r.h.s expansion of (53), it is straightforward to verify that the adjoint of the discrete Hessian operator is indeed computed according to (14).

B. Proof of Proposition 1

By definition, the orthogonal projection of a matrix $\mathbf{Y}$ onto the set B_{S_q} is given by

$$P_{B_{S_q}}\left(\mathbf{Y}\right) = \underset{\left\|\mathbf{X}\right\|_{S_q}\leq\rho}{\arg\min}\left\|\mathbf{X} - \mathbf{Y}\right\|_F^2\,. \tag{54}$$

Since all Schatten norms are unitarily invariant, we equivalently have

$$P_{B_{S_q}}\left(\mathbf{Y}\right) = \underset{\left\|\mathbf{U}^H\mathbf{X}\mathbf{V}\right\|_{S_q}\leq\rho}{\arg\min}\left\|\mathbf{U}^H\mathbf{X}\mathbf{V} - \mathbf{U}^H\mathbf{Y}\mathbf{V}\right\|_F^2\,. \tag{55}$$

Let us now consider the matrix $\mathbf{Z} = \mathbf{U}^H \mathbf{X} \mathbf{V}$ that is associated with the solution of (54). If we substitute $\mathbf{Z}$ in (55), then we end up with the following constrained minimization problem

$$P_{B_{S_q}}(\mathbf{\Sigma}) = \underset{\|\mathbf{Z}\|_{S_q} \leq \rho}{\arg\min} \|\mathbf{Z} - \mathbf{\Sigma}\|_F^2 , \tag{56}$$

which corresponds to the projection of the diagonal matrix $\mathbf{\Sigma}$ onto the set B_{S_q}. Now, if $P_{B_{S_q}}(\mathbf{\Sigma}) = \hat{\mathbf{Z}}$, we have

$$\left\|\hat{\mathbf{Z}} - \mathbf{\Sigma}\right\|_F^2 = \left\|\hat{\mathbf{Z}}\right\|_F^2 + \|\mathbf{\Sigma}\|_F^2 - 2\,\mathrm{Re}\left(\mathrm{tr}\left(\hat{\mathbf{Z}}^H \mathbf{\Sigma}\right)\right) \geq \left\|\hat{\mathbf{\Sigma}}\right\|_F^2 + \|\mathbf{\Sigma}\|_F^2 - 2\,\mathrm{tr}\left(\hat{\mathbf{\Sigma}}^T \mathbf{\Sigma}\right) = \left\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\|_F^2 , \tag{57}$$

where the inequality stems from von Neumann's trace theorem [47], and $\hat{\mathbf{\Sigma}}$ is a diagonal matrix with the singular values of $\hat{\mathbf{Z}}$. In addition, it holds that

$$\left\|\hat{\mathbf{\Sigma}}\right\|_{S_q} = \left\|\hat{\mathbf{Z}}\right\|_{S_q} \leq \rho . \tag{58}$$

Equations (57) and (58) immediately imply that the projection of $\mathbf{\Sigma}$ equals to $\hat{\mathbf{Z}} = \hat{\mathbf{\Sigma}}$, i.e., a positive semidefinite diagonal matrix. We can then perform this operation by projecting the vector, formed by the main diagonal of $\mathbf{\Sigma}$, onto the convex set B_q, and then by transforming the projected vector back to a diagonal matrix. Using this fact and the relations between the optimal solution of (54) and (56), we finally express the projection of the matrix $\mathbf{Y}$ onto B_{S_q} as

$$P_{B_{S_q}}(\mathbf{Y}) = \mathbf{U}\mathrm{diag}\left(P_{B_q}(\boldsymbol{\sigma}(\mathbf{Y}))\right)\mathbf{V}^H . \tag{59}$$

C. Proof of Lemma 1

First, we present a matrix inequality that involves the Schatten norms and it will be subsequently used for the proof of the lemma. Let $\mathbf{X}, \mathbf{Y} \in \mathbb{C}^{n_1 \times n_2}$. Then, the inner product of these two matrices satisfies the following inequality

$$\begin{aligned}\langle \mathbf{X}, \mathbf{Y} \rangle_{\mathbb{C}^{n_1 \times n_2}} &= \mathrm{Re}\left(\mathrm{tr}\left(\mathbf{Y}^H \mathbf{X}\right)\right) \leq \langle \boldsymbol{\sigma}(\mathbf{X}), \boldsymbol{\sigma}(\mathbf{Y}) \rangle_2 \\ &\leq \|\boldsymbol{\sigma}(\mathbf{X})\|_q \, \|\boldsymbol{\sigma}(\mathbf{Y})\|_p = \|\mathbf{X}\|_{S_q} \|\mathbf{Y}\|_{S_p} .\end{aligned} \tag{60}$$

The first inequality is due to von Neumann's trace theorem [47], while the second one due to Hölder's inequality. The last equality holds true from the definition of Schatten norms.

By definition, the dual norm of (20) is given by [31]:

$$\|\mathbf{\Omega}\|_D = \max_{\|\mathbf{\Psi}\|_{1,p} \leq 1} \langle \mathbf{\Omega}, \mathbf{\Psi} \rangle_X , \tag{61}$$

where X, instead of $\mathbb{R}^{N \times 2 \times 2}$ that is used throughout the paper, here is assumed to be the more general linear space $X = \mathbb{C}^{N \times n_1 \times n_2}$. We consider the inequality

$$\langle \mathbf{\Omega}, \mathbf{\Psi} \rangle_X = \sum_{n=1}^{N} \mathrm{Re}\left(\mathrm{tr}\left(\mathbf{\Psi}_n^H \mathbf{\Omega}_n\right)\right) \leq \sum_{n=1}^{N} \|\mathbf{\Omega}_n\|_{S_q} \|\mathbf{\Psi}_n\|_{S_p} , \tag{62}$$

which immediatelly follows from inequality (60). Now, by introducing the vectors $\boldsymbol{\omega} = \left(\|\mathbf{\Omega}_1\|_{S_q}, \|\mathbf{\Omega}_2\|_{S_q}, \ldots, \|\mathbf{\Omega}_N\|_{S_q}\right)$ and $\boldsymbol{\psi} = \left(\|\mathbf{\Psi}_1\|_{S_p}, \|\mathbf{\Psi}_2\|_{S_p}, \ldots, \|\mathbf{\Psi}_N\|_{S_p}\right)$, and applying once again Hölder's inequality, we get

$$\begin{aligned}\sum_{n=1}^{N} \|\mathbf{\Omega}_n\|_{S_q} \|\mathbf{\Psi}_n\|_{S_p} &= \langle \boldsymbol{\omega}, \boldsymbol{\psi} \rangle_2 \leq \|\boldsymbol{\omega}\|_\infty \|\boldsymbol{\psi}\|_1 \\ &= \|\mathbf{\Omega}\|_{\infty,q} \|\mathbf{\Psi}\|_{1,p} .\end{aligned} \tag{63}$$

From the definition of the dual norm (61) and the inequalities (62) and (63) we conclude that $\|\mathbf{\Omega}\|_D \leq \|\mathbf{\Omega}\|_{\infty,q}$. To prove that $\|\mathbf{\Omega}\|_D = \|\mathbf{\Omega}\|_{\infty,q}$, we next show that for each $\mathbf{\Omega}$ we can find a $\mathbf{\Psi}$ satisfying $\|\mathbf{\Psi}\|_{1,p} = 1$, and for which $\langle \mathbf{\Omega}, \mathbf{\Psi} \rangle_X = \|\mathbf{\Omega}\|_{\infty,q}$. To that end, let k be any index in the set $\left\{\arg\max_{1 \leq n \leq N} \|\mathbf{\Omega}_n\|_{S_q}\right\}$ and $\mathbf{\Omega}_k = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^H$ be the singular value decomposition of $\mathbf{\Omega}_k$. Then, we set $\mathbf{\Psi}_n = \mathbf{O}$ for all n except for $n = k$ for which we have

$$\mathbf{\Psi}_k = \mathbf{U}_k \mathbf{E} \mathbf{V}_k^H , \tag{64}$$

where

$$\mathbf{E}^{(i,j)} = \frac{\left(\mathbf{\Sigma}_k^{(i,j)}\right)^{q-1}}{\|\mathbf{\Omega}_k\|_{S_q}^{q-1}} , \tag{65}$$

and $\mathbf{\Sigma}_k^{(i,j)}$ corresponds to the (i, j)-th entry of the matrix $\mathbf{\Sigma}_k \in \mathbb{D}^{n_1 \times n_2}$. Now, we have that

$$\begin{aligned}\langle \mathbf{\Omega}, \mathbf{\Psi} \rangle_X &= \sum_{n=1}^{N} \mathrm{Re}\left(\mathrm{tr}\left(\mathbf{\Psi}_n^H \mathbf{\Omega}_n\right)\right) = \mathrm{Re}\left(\mathrm{tr}\left(\mathbf{\Psi}_k^H \mathbf{\Omega}_k\right)\right) \\ &= \mathrm{tr}\left(\mathbf{E}^H \mathbf{\Sigma}_k\right) = \frac{\sum_{i=1}^{\min(n_1,n_2)} \left(\mathbf{\Sigma}_k^{(i,i)}\right)^q}{\|\mathbf{\Omega}_k\|_{S_q}^{q-1}} \\ &= \|\mathbf{\Omega}_k\|_{S_q} = \|\mathbf{\Omega}\|_{\infty,q} .\end{aligned} \tag{66}$$

Furhermore, for the mixed norm $\|\mathbf{\Psi}\|_{1,p} = \|\mathbf{\Psi}_k\|_{S_p}$ it holds

$$\begin{aligned}\|\mathbf{\Psi}\|_{1,p} &= \left(\sum_{i=1}^{\min(n_1,n_2)} \left(\mathbf{E}^{(i,i)}\right)^p\right)^{1/p} \\ &= \frac{\left(\sum_{i=1}^{\min(n_1,n_2)} \left(\mathbf{\Sigma}_k^{(i,i)}\right)^q\right)^{\frac{q-1}{q}}}{\|\mathbf{\Omega}_k\|_{S_q}^{q-1}} = \frac{\|\mathbf{\Omega}_k\|_{S_q}^{q-1}}{\|\mathbf{\Omega}_k\|_{S_q}^{q-1}} = 1\end{aligned} \tag{67}$$

which completes the proof of the lemma.

D. Proof of Proposition 2

For any pair of variables $\mathbf{\Omega}, \mathbf{\Psi} \in X$ we have

$$\begin{aligned}\|\nabla s(\mathbf{\Omega}) - \nabla s(\mathbf{\Psi})\|_X &= \|\tau \boldsymbol{H}(V(\mathbf{\Omega}) - V(\mathbf{\Psi}))\|_X \\ &\leq \tau \|\boldsymbol{H}\| \, \|V(\mathbf{\Omega}) - V(\mathbf{\Psi})\|_2 \\ &\leq \tau \|\boldsymbol{H}\| \, \left\|\tau \boldsymbol{H}^*(\mathbf{\Omega} - \mathbf{\Psi})\right\|_2 \\ &\leq \tau^2 \|\boldsymbol{H}\| \, \left\|\boldsymbol{H}^*\right\| \, \|\mathbf{\Omega} - \mathbf{\Psi}\|_X \\ &= \tau^2 \|\boldsymbol{H}\|^2 \, \|\mathbf{\Omega} - \mathbf{\Psi}\|_X ,\end{aligned} \tag{68}$$

where $V(\mathbf{\Omega}) = P_C\left(\mathbf{z} - \tau \boldsymbol{H}^* \mathbf{\Omega}\right)$. Note that, the first and third inequalities follow from the relation between the norms, defined in the spaces X and $\mathbb{R}^N$, and the induced operator

norm, i.e., $\|\boldsymbol{H}\mathbf{x}\|_X \leq \|\boldsymbol{H}\|\,\|\mathbf{x}\|_2$, while the second one holds because the projection operator P_C onto the convex set $C \subseteq \mathbb{R}^N$, is firmly nonexpansive [48, Proposition 4.8]. This means that

$$\|P_C(\mathbf{x}) - P_C(\mathbf{y})\|_2 \leq \|\mathbf{x}-\mathbf{y}\|_2 \;\forall\, \mathbf{x},\mathbf{y} \in \mathbb{R}^N. \tag{69}$$

To compute an upper bound of $\|\boldsymbol{H}\|$, we exploit that $\|\boldsymbol{H}\|^2 = \left\|\boldsymbol{H}^*\boldsymbol{H}\right\|$ [49] (a general property of bounded linear operators), and we get

$$\begin{aligned}\left\|\boldsymbol{H}^*\boldsymbol{H}\mathbf{x}\right\|_2 &= \left\|\left(\Delta^*_{r_1r_1}\Delta_{r_1r_1} + 2\Delta^*_{r_1r_2}\Delta_{r_1r_2} + \Delta^*_{r_2r_2}\Delta_{r_2r_2}\right)\mathbf{x}\right\|_2 \\ &\leq \left(\left\|\Delta_{r_1r_1}\right\|^2 + 2\left\|\Delta_{r_1r_2}\right\|^2 + \left\|\Delta_{r_2r_2}\right\|^2\right)\|\mathbf{x}\|_2 .\end{aligned} \tag{70}$$

Now, using the definitions of the second-order differential operators in (10), it is easy to show that each of $\left\|\Delta_{r_1r_1}\right\|$, $\left\|\Delta_{r_1r_2}\right\|$ and $\left\|\Delta_{r_2r_2}\right\|$ is smaller than or equal to 4. This immediately implies that $\|\boldsymbol{H}\| \leq 8$ and, hence, an upper bound of the Lipschitz constant of $\nabla s(\boldsymbol{\Omega})$ will be $L(s) \leq \tau^2\|\boldsymbol{H}\|^2 \leq 64\tau^2$.

ACKNOWLEDGMENT

The authors would like to acknowledge Jean-Charles Baritaux and Pouya Tafti for fruitful discussions. They would also like to thank the anonymous reviewers and the associate editor for their useful comments and suggestions.

REFERENCES

[1] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*. Boca Raton, FL: CRC Press, 1998.

[2] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, nos. 1–4, pp. 259–268, 1992.

[3] J. Bioucas-Dias, J. Oliveira, and M. Figueiredo, "Total variation-based image deconvolution: A majorization—minimization approach," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, vol. 2. Toulouse, France, 2006, pp. 861–864.

[4] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.

[5] T. F. Chan and J. Shen, "Mathematical models for local nontexture inpaintings," *SIAM J. Appl. Math.*, vol. 62, no. 3, pp. 1019–1043, 2002.

[6] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vis.*, vol. 20, nos. 1–2, pp. 89–97, 2004.

[7] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, 2007.

[8] T. Chan, A. Marquina, and P. Mulet, "High-order total variation-based image restoration," *Soc. Ind. Appl. Math. J. Sci. Comput.*, vol. 22, no. 2, pp. 503–516, 2000.

[9] A. Chambolle and P.-L. Lions, "Image recovery via total variation minimization and related problems," *Numer. Math.*, vol. 76, no. 2, pp. 167–188, 1997.

[10] M. Lysaker and X.-C. Tai, "Iterative image restoration combining total variation minimization and a second-order functional," *Int. J. Comput. Vis.*, vol. 66, no. 1, pp. 5–18, 2006.

[11] T. F. Chan, S. Esedoglu, and F. E. Park, "Image decomposition combining staircase reduction and texture extraction," *J. Visual Commun. Image Represent.*, vol. 18, no. 6, pp. 464–486, 2007.

[12] S. Setzer, G. Steidl, and T. Teuber, "Infimal convolution regularizations with discrete ℓ_1-type functionals," *Commun. Math. Sci.*, vol. 9, no. 3, pp. 797–827, 2011.

[13] Y.-L. You and M. Kaveh, "Fourth-order partial differential equations for noise removal," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1723–1730, Oct. 2000.

[14] M. Lysaker, A. Lundervold, and X.-C. Tai, "Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time," *IEEE Trans. Image Process.*, vol. 12, no. 12, pp. 1579–1590, Dec. 2003.

[15] G. Steidl, "A note on the dual treatment of higher-order regularization functionals," *Computing*, vol. 76, nos. 1–2, pp. 135–148, 2006.

[16] J. Yuan, C. Schnörr, and G. Steidl, "Total-variation based piecewise affine regularization," in *Proc. Sec. Int. Conf. Scale Space Methods Variat. Methods Comput. Vis.*, 2009, pp. 552–564.

[17] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM J. Imaging Sci.*, vol. 3, no. 3, pp. 492–526, 2010.

[18] R. Bhatia, *Matrix Analysis*. New York: Springer-Verlag, 1997.

[19] S. Lefkimmiatis, A. Bourquard, and M. Unser, "Hessian-based norm regularization for image restoration with biomedical applications," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 983–995, Mar. 2012.

[20] P. C. Hansen, J. G. Nagy, and D. P. O'Leary, *Deblurring Images: Matrices, Spectra, and Filtering*. Philadelphia, PA: SIAM, 2006.

[21] J. E. Denis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Philadelphia, PA: SIAM, 1983.

[22] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.

[23] M. Figueiredo, J. Bioucas-Dias, and R. Nowak, "Majorization—minimization algorithms for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2980–2991, Dec. 2007.

[24] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[25] D. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Stat.*, vol. 58, no. 1, pp. 30–37, 2004.

[26] M. Zibulevsky and M. Elad, "L1-L2 optimization in signal and image processing," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 76–88, May 2010.

[27] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, 2005.

[28] A. Lewis and J. Malick, "Alternating projections on manifolds," *Math. Oper. Res.*, vol. 33, no. 1, pp. 216–234, 2008.

[29] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with sparsity-inducing penalties," *Found. Trends Mach. Learn.*, vol. 4, no. 1, pp. 1–106, 2011.

[30] S. Suvrit, "Fast projections onto $\ell_{1,q}$-norm balls for grouped feature selection," in *Proc. Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, vol. 3. 2011, pp. 305–317.

[31] R. T. Rokcafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.

[32] Y. Nesterov, "A method for solving a convex programming problem with convergence rates $O(1/k^2)$," *Soviet Math. Dokl*, vol. 27, no. 2, pp. 372–376, 1983.

[33] J. Liu and J. Ye, "Efficient ℓ_1/ℓ_q norm regularization," 2010, Arxiv preprint arXiv: 1009.4766

[34] J.-F. Cai, E. J. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2008.

[35] S. Ma, D. Godfarb, and L. Chen, "Fixed point and Bregman iterative methods for matrix rank minimization," *Math. Program.*, vol. 128, no. 1, pp. 321–353, 2011.

[36] G. A. Watson, "Characterization of the subdifferential of some matrix norms," *Linear Algebra Appl.*, vol. 170, pp. 33–45, Jun. 1992.

[37] D. L. Donoho, "Denoising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[38] E. Candes and J. Romberg, "Practical signal recovery from random projections," *Proc. SPIE*, vol. 5674, pp. 76–86, Jan. 2005.

[39] I. Daubechies, M. Fornasier, and I. Loris, "Accelerated projected gradient method for linear inverse problems with sparsity constraints," *J. Fourier Anal. Appl.*, vol. 14, nos. 5–6, pp. 764–792, 2008.

[40] E. Van De Berg, M. Friedlander, and K. Murphy, "Group sparsity via linear-time projection," Dept. Comput. Sci., Univ. British Columbia, Vancouver, BC, Canada, Tech. Rep. TR-2008-09, 2008.

[41] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, "Efficient projections onto the ℓ_1-ball for learning in high dimensions," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 272–279.

[42] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image restoration by sparse 3D transform-domain collaborative filtering," *Proc. SPIE*, vol. 6812, p. 681207, Mar. 2008.

[43] I. Selesnick and M. Figueiredo, "Signal restoration with overcomplete wavelet transforms: Comparison of analysis and synthesis priors," *Proc. SPIE*, vol. 7446, p. 74460D, Aug. 2009.
[44] C. Vonesch, F. Aguet, J.-L. Vonesch, and M. Unser, "The colored revolution of bioimaging," *IEEE Sig. Process. Mag.*, vol. 23, no. 3, pp. 20–31, May 2006.
[45] C. Vonesch and M. Unser, "A fast thresholded Landweber algorithm for wavelet-regularized multidimensional deconvolution," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 539–549, Apr. 2008.
[46] J. R. Shewchuk. (1994). *An Introduction to the Conjugate Gradient Method without the Agonizing Pain* [Online]. Available: http://www.cs.cmu.edu/∼jrs/jrspapers.html
[47] L. Mirsky, "A trace inequality of John von Neumann," *Monatshefte Fur Mathematik*, vol. 79, no. 4, pp. 303–306, 1975.
[48] H. H. Bauschke and P. L. Combettes, "Convexity and nonexpansiveness," in *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, J. M. Borwein and P. Borwein, Eds. New York: Springer-Verlag, 2011, pp. 59–74.
[49] V. Hutson, J. S. Pym, and M. J. Cloud, *Applications of Functional Analysis and Operator Theory*, 2nd ed. New York: Elsevier, 2005.

Stamatios Lefkimmiatis (S'08–M'11) received the M.Sc. degree in computer engineering and informatics (hons.) from the University of Patras, Patras, Greece, in 2004, and the Ph.D. degree from the National Technical University of Athens, Athens, Greece, in 2009.

He has been a Research Scientist with the Biomedical Imaging Group, Swiss Federal Institute of Technology, Lausanne, Switzerland, since 2010. His current research interests include image analysis, statistical modeling, and image reconstruction using convex optimization techniques.

John Paul Ward received the B.S. degree in mathematics from the University of Georgia, Athens, and the Ph.D. degree in mathematics from Texas A&M University, College Station, in 2005 and 2010, respectively.

He is currently a Post-Doctoral Researcher with the Biomedical Imaging Group, Swiss Federal Institute of Technology, Lausanne, Switzerland.

Michael Unser (M'89–SM'94–F'99) is a Professor and Director of the École Polytechnique Fédérale de Lausanne's Biomedical Imaging Group, Lausanne, Switzerland. From 1985 to 1997, he was with the Biomedical Engineering and Instrumentation Program, National Institutes of Health, Bethesda, MD, conducting research on bioimaging and heading the Image Processing Group. He has published over 200 journal papers. His current research interests include biomedical image processing, sampling theories, multiresolution algorithms, wavelets, and the use of splines for image processing.

He was a EURASIP Fellow in 2009 and a member of the Swiss Academy of Engineering Sciences. He is the recipient of several international prizes, including three IEEE-SPS Best Paper Awards and two Technical Achievement Awards from the IEEE (2008 SPS and EMBS 2010).