Poisson Image Reconstruction With Hessian Schatten-Norm Regularization

Stamatios Lefkimmiatis, *Member, IEEE*, and Michael Unser, *Fellow, IEEE*

Abstract—Poisson inverse problems arise in many modern imaging applications, including biomedical and astronomical ones. The main challenge is to obtain an estimate of the underlying image from a set of measurements degraded by a linear operator and further corrupted by Poisson noise. In this paper, we propose an efficient framework for Poisson image reconstruction, under a regularization approach, which depends on matrix-valued regularization operators. In particular, the employed regularizers involve the Hessian as the regularization operator and Schatten matrix norms as the potential functions. For the solution of the problem, we propose two optimization algorithms that are specifically tailored to the Poisson nature of the noise. These algorithms are based on an augmented-Lagrangian formulation of the problem and correspond to two variants of the alternating direction method of multipliers. Further, we derive a link that relates the proximal map of an ℓ_p norm with the proximal map of a Schatten matrix norm of order p. This link plays a key role in the development of one of the proposed algorithms. Finally, we provide experimental results on natural and biological images for the task of Poisson image deblurring and demonstrate the practical relevance and effectiveness of the proposed framework.

Index Terms—Poisson noise, Hessian operator, schatten norms, eigenvalue optimization, ADMM, image reconstruction.

I. INTRODUCTION

IN this work, we focus on the recovery of an image from a set of measurements that are degraded by a linear operator and further corrupted by Poisson noise. Poisson (or shot) noise [1] is due to fluctuations in the number of detected photons and is an inherent limitation of the detection process occurring in many imaging devices such as CCD cameras. In this context, the measurements at every location of the sensor can be considered as the realization of an indirect Poisson random variable, in the sense that its mean and variance is indirectly related, through a linear transformation, to the underlying intensity of the image. Such inverse problems, which in the literature are also known as *Poisson inverse problems* [2], [3], arise in several imaging applications, including medical [4], biological [5], and astronomical [6] ones.

For most cases of practical interest, Poisson inverse problems are ill-posed; the linear operator relating the under-

lying intensities to the Poisson measurements is either ill-conditioned or singular. Thus, additional knowledge is required to obtain a meaningful image reconstruction. This knowledge is expressed in the variational framework by a penalty function that constrains the number of plausible solutions. Then, the reconstruction task is cast as the optimization of an objective function consisting of two terms: the *data fidelity* that quantifies the proximity between the reconstructed image and the measurements, and the *regularizer* that incorporates our assumptions about the underlying image. From a statistical point of view, the image reconstruction corresponds to estimating the underlying intensity from a single realization of the Poisson random variables.

A. Overview of Current Methods

Well-established reconstruction methods have been shown to provide satisfactory results in the context of linear inverse problems with Gaussian independent identically distributed (i.i.d.) noise. Unfortunately, they cannot be applied in our case because Poisson inverse problems face a noise that is signal-dependent. Despite this difficulty, the prevalence of photon-imaging applications has generated an increasing interest during the past few years and has triggered the development of several algorithms that were designed to handle the Poisson nature of the noise. Most of these methods can be interpreted as optimization techniques of an objective function. For instance, the classical Richardson-Lucy (RL) algorithm [7], [8] corresponds to a maximum-likelihood (ML) estimate. Another approach which is widely followed is to obtain the solution as a penalized ML estimate. In [9], the authors proposed an algorithm consisting of two steps. In the first step, a variance-stabilizing transform (such as Anscombe [10]) is applied to the measurements to reform the data so that the noise becomes approximately Gaussian with constant variance. Then, in the second step, the solution is obtained as the minimizer of an objective function that consists of a nonlinear data-fidelity term and a regularizer acting on the transform-domain coefficients of the underlying image. This kind of regularization is known in the literature as frame-synthesis (FS) since the regularizer is not a direct function of the image. Methods that seek the solution by penalizing coefficients that arise from multiscale nonlinear transformations of the underlying image have been proposed in [2], [11] and in references therein. For the simplest Poisson inverse problem (denoising) and despite the nonlinearity of the employed transform, it has been shown in [11] that the resulting estimate is optimal, in the minimum mean-squared error (MMSE) sense, both in the coefficient and the image domain. Finally, there also exist algorithms that obtain penalized ML solutions where the regularizers are

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

directly applied on the underlying image, *i.e.*, frame-analysis (FA) regularization framework. This includes the modified RL algorithm with total-variation (TV) regularization [12] and the methods proposed in [13]–[15], which use TV as well but adopt different optimization strategies.

B. Contributions

In this work, we consider methods specifically tailored to Poisson-distributed data and we employ a class of second-order derivative-based regularizers that act directly on the underlying image. Our proposed algorithms are based on an augmented-Lagrangian framework and follow the same principles as other existing algorithms in the literature [13], [14]. However, while the latter have been designed to work with regularizers involving scalar and vector-valued operators, ours are designed to handle matrix-valued linear operators. This generalization is important since it introduces additional difficulties which cannot be handled in a trivial way. Nevertheless, we are able to address these problems and we introduce two novel algorithms that can be considered extensions of existing techniques to the case of matrix-valued regularization operators.

The key contributions of this work are as follows.

1) We employ a family of non-quadratic regularization functionals that involves matrix-valued operators [16] for dealing with Poisson inverse problems. These functionals are based on the Schatten norms [17] of the Hessian operator, evaluated at every pixel of the image, and are shown to be appropriate candidates for modeling a wide class of natural images.
2) We propose two alternative minimization approaches that can provide a solution to the optimization problem under any Schatten-norm-based penalty function. These two algorithms are based on an augmented-Lagrangian formulation and arise as two different variants of the *alternating-direction method of multipliers* (ADMM).
3) We derive a result that relates the proximal map evaluation of a Schatten matrix norm of order p to the proximal map evaluation of an ℓ_p norm. This link plays a key role in the development of one of the proposed algorithms but can potentially have a wider range of applicability.

The rest of this paper is organized as follows: In Section II, we provide a formal description of Poisson inverse problems, we discuss regularization strategies, and we introduce the family of non-quadratic second-order functionals that we employ for reconstructing images. In Section III, we briefly review the ADMM framework and we describe our two proposed optimization algorithms. Then, in Section IV, we perform experiments on natural and biomedical images to assess the performance of our techniques for the task of Poisson image deblurring. We conclude our work in Section V.

II. REGULARIZATION OF POISSON INVERSE PROBLEMS

A. Problem Formulation

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

$$\mathbf{y} = \mathcal{T}(\mathbf{Kx}), \tag{1}$$

where $\mathbf{K} \in \mathbb{R}^{M \times N}$ is a matrix that models the spatial response of the imaging device, $\mathcal{T}$ is an operation that describes the nature of the measurement noise and how it affects the image acquisition, and $\mathbf{y} \in \mathbb{R}_+^M$, $\mathbf{x} \in \mathbb{R}_+^N$ are the vectorized versions of the observed image and the image to be estimated, respectively. The set $\mathbb{R}_+$ denotes the set of real non negative numbers. We further expand (1) by noting that the unobserved intensities $\mathbf{x}$ of the underlying image are related to other intensities $\boldsymbol{\lambda}$ (which are also unobserved) through the linear relation $\boldsymbol{\lambda} = \mathbf{Kx}$. For the elements of $\boldsymbol{\lambda}$ to be interpreted as photon intensities, we further need to assume that $\boldsymbol{\lambda} \in \mathbb{R}_+^M$. This hypothesis is consistent with the idea that $\mathbf{K}$ represents a linear degradation effect, e.g., blurring.

Under the basic photon-imaging model, the measurement noise implied by $\mathcal{T}$ in (1) is signal-dependent. In particular, the observed image $\mathbf{y}$ can be considered as a vector containing the samples of a random sequence $\mathbf{Y} = (Y_1, \ldots, Y_M)$. The M random variables Y_m are conditionally independent upon $\boldsymbol{\lambda}$, and each one follows a Poisson distribution with rate parameter λ_m, so that $Y_m \sim \text{Pois}(Y_m | \lambda_m)$. The likelihood of $\mathbf{x}$ can then be expressed as

$$p(\mathbf{Y} = \mathbf{y} | \mathbf{x}) = \prod_{m=1}^M \frac{e^{-\lambda_m} \lambda_m^{y_m}}{y_m!} = \prod_{m=1}^M \frac{e^{-[\mathbf{Kx}]_m} [\mathbf{Kx}]_m^{y_m}}{y_m!}, \tag{2}$$

where $[\cdot]_m$ indicates the mth element of the vector argument.

One way to recover $\mathbf{x}$ from the measurements $\mathbf{y}$ is by minimizing the negative log-likelihood

$$f_L(\mathbf{x}) = \sum_{m=1}^M ([\mathbf{Kx}]_m - \mathbf{y}_m \log [\mathbf{Kx}]_m) + \iota_{\mathcal{C}}(\mathbf{x}) + \text{const.}, \tag{3}$$

where $\iota_{\mathcal{C}}$ is the indicator function of the convex set $\mathcal{C} = \mathbb{R}_+^N$ that represents non-negativity constraints on the solution. The convention is that $\iota_{\mathcal{C}}(\mathbf{x})$ takes values 0 for $\mathbf{x} \in \mathcal{C}$ and ∞ otherwise. In order for (3) to be consistent with (2), we also use the convention that $0 \log(0) = 0$. The minimizer of (3) is equivalent to the ML estimate. One way to obtain it is by using RL, which is an iterative multiplicative gradient type technique [7], [8]. Unfortunately, the ill-posed nature of the problem implies that the operator $\mathbf{K}$ for the cases of practical interest is either ill-conditioned or singular, and the corresponding ML estimator has a large variance that may lead to a not informative solution. For this reason, the RL algorithm should be terminated early in practice, after just a few iterations. Otherwise, the noise gets amplified and results in an image estimate dominated by noise.

To alleviate this amplification of the noise, one can follow the alternative approach of minimizing a penalized version of the negative log-likelihood. This can be formulated by

$$f(\mathbf{x}) = f_L(\mathbf{x}) + \phi(\mathbf{x}), \tag{4}$$

where the role of the term $\phi(\mathbf{x})$ is to narrow down the set of plausible solutions by favoring those that satisfy properties related to the assumed regularity of the underlying image. The minimizer of (4) corresponds to a maximum *a posteriori* (MAP) estimate, with $\phi(\mathbf{x})$ being interpreted as the negative logarithm of the prior distribution of the underlying image $\mathbf{x}$.

The minimization of (4) can also arise if, instead of a Bayesian approach, we treat the problem in a variational framework. Indeed, we can interpret (4) as an energy functional consisting of two terms: the data fidelity $f_L(\mathbf{x})$ and the regularizer $\phi(\mathbf{x})$. Then, the estimate of the underlying image corresponds to the minimizer of this energy functional.

B. Regularization Strategy

Many regularization techniques have been proposed for solving ill-posed inverse problems. Among the most popular ones is the TV semi-norm [18] which produces good estimates for a variety of imaging applications. The main reason for its wide acceptance in the imaging community is that TV manages to provide reconstructions with well-preserved and sharp edges. However, its downside is that it introduces strong block artifacts, commonly known in the literature as the *staircase effect*. In this work, to avoid this undesirable effect, we employ instead the family of non-quadratic regularizers that we recently proposed in [16]. Our regularizers are based on the Schatten norms of the Hessian matrix computed at every pixel of the image. They can be considered as second-order extensions of TV, in the sense that they satisfy the same invariance properties while involving second-order derivatives instead of first-order ones. Their main feature is that they favor piecewise-smooth solutions, as opposed to TV which favors piecewise-constant solutions. Therefore, for a wide range of images, including those that are commonly met in biomedical applications, second-order regularization can potentially provide reconstructions which better approximate the intensity variations of the underlying image and at the same time avoid the introduction of severe artifacts [16], [19].

C. Notations and Definitions

In the sequel capital slanted bold-faced letters will be used to refer to multidimensional arrays, while capital bold-faced letters will be reserved for matrices. The set of unitary matrices is denoted 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. Further, the set of positive semidefinite diagonal matrices is denoted as $\mathbb{D}^{n_1 \times n_2} = \left\{ \mathbf{X} \in \mathbb{R}_+^{n_1 \times n_2} : \mathbf{X}(i, j) = 0 \ \forall i \neq j \right\}$.

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 ($\mathcal{S}_p$ norm) of $\mathbf{X}$ is defined as

$$\|\mathbf{X}\|_{\mathcal{S}_p} = \left(\sum_{k=1}^{\min(n_1, n_2)} \sigma_k^p \right)^{\frac{1}{p}}, \tag{5}$$

where $p \geq 1$ and σ_k is the kth singular value of $\mathbf{X}$.

Definition 1 implies that the $\mathcal{S}_p$ norm of a matrix $\mathbf{X}$ corresponds to the ℓ_p norm of its singular-values vector $\boldsymbol{\sigma} \in \mathbb{R}_+^{\min(n_1, n_2)}$. This further means that all Schatten norms are unitarily invariant, which is a highly desirable property that we are going to exploit next. Moreover, we note that the family of $\mathcal{S}_p$ norms includes three of the most popular matrix norms,

namely, the nuclear/trace norm ($p = 1$), the Frobenius norm ($p = 2$), and the spectral/operator norm ($p = \infty$).

D. Discrete Hessian Operator and Its Adjoint

Hereafter, we use $\mathcal{H}$ to refer to the discrete version of the Hessian operator. To simplify our analysis, we assume that the image intensities on an $N_x \times N_y$ grid are rasterized in a vector $\mathbf{x}$ of dimension $N = N_x \cdot N_y$ so that the pixel at coordinates (i, j) maps to the nth entry of $\mathbf{x}$ with $n = j N_x + (i + 1)$. In this case, the discrete Hessian operator is a mapping $\mathcal{H} : \mathbb{R}^N \mapsto \mathcal{K}$, where $\mathcal{K} = \mathbb{R}^{N \times 2 \times 2}$. For $\mathbf{x} \in \mathbb{R}^N$, $\mathcal{H}\mathbf{x}$ is given as

$$[\mathcal{H}\mathbf{x}]_n = \begin{bmatrix} [\Delta_{r_1 r_1} \mathbf{x}]_n & [\Delta_{r_1 r_2} \mathbf{x}]_n \\ [\Delta_{r_1 r_2} \mathbf{x}]_n & [\Delta_{r_2 r_2} \mathbf{x}]_n \end{bmatrix}, \tag{6}$$

where $n = 1, \ldots, N$ and $\Delta_{r_1 r_1}$, $\Delta_{r_2 r_2}$, and $\Delta_{r_1 r_2}$ denote the finite-difference operators that compute the discrete approximations of the second-order partial derivatives along the two dimensions of the image. In Section IV, where we report reconstruction results employing the Hessian, we use a forward finite-difference approximation scheme. Then the second-order derivative operators $\Delta_{r_i r_j}$, are defined as in [16].

We equip the space $\mathcal{K}$ with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{K}}$ and norm $\|\cdot\|_{\mathcal{K}}$. To define them, let $X, Y \in \mathcal{K}$, with $X_n, Y_n \in \mathbb{R}^{2 \times 2} \ \forall \ n = 1, \ldots, N$. Then, we have

$$\langle X, Y \rangle_{\mathcal{K}} = \sum_{n=1}^{N} \mathrm{tr} \left(Y_n^T X_n \right) \tag{7}$$

and

$$\|X\|_{\mathcal{K}} = \sqrt{\langle X, X \rangle_{\mathcal{K}}} = \sqrt{\sum_{n=1}^{N} \|X_n\|_F^2}, \tag{8}$$

where $\mathrm{tr}(\cdot)$ is the trace of the input matrix, $(\cdot)^T$ denotes the transpose operation, and $\|\cdot\|_F$ is the Frobenius matrix norm. For the Euclidean space $\mathbb{R}^N$, we denote its inner product and norm by $\langle \cdot, \cdot \rangle_2$ and $\|\cdot\|_2$, respectively.

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

$$\langle Y, \mathcal{H}\mathbf{x} \rangle_{\mathcal{K}} = \langle \mathcal{H}^* Y, \mathbf{x} \rangle_2 . \tag{9}$$

Note that this definition of the adjoint operator is a generalization of the Hermitian transpose. Based on the relation of the inner products in (9), for any $Y \in \mathcal{K}$ it holds that [16]

$$[\mathcal{H}^* Y]_n = \left[\Delta_{r_1 r_1}^* Y^{(1,1)} \right]_n + \left[\Delta_{r_1 r_2}^* \left(Y^{(1,2)} + Y^{(2,1)} \right) \right]_n + \left[\Delta_{r_2 r_2}^* Y^{(2,2)} \right]_n, \tag{10}$$

where $\Delta_{r_1 r_1}^*$, $\Delta_{r_1 r_2}^*$, $\Delta_{r_2 r_2}^*$ are the adjoint finite-difference operators and $Y^{(i,j)}$ represents an N-size column-vector that consists of the (i, j) entries of the matrices Y_n, with $n = 1, \ldots, N$.

E. Hessian Schatten-Norm Regularization

Next, we proceed with the definition of the regularization family that we employ to solve Poisson inverse problems. The discrete Hessian Schatten-norm regularizers are defined as [16]

$$\mathcal{HS}_p(\mathbf{x}) = \|\mathcal{H}\mathbf{x}\|_{1,p} = \sum_{n=1}^{N} \|[\mathcal{H}\mathbf{x}]_n\|_{\mathcal{S}_p}, \ \forall p \geq 1. \quad (11)$$

Based on their definition, we observe that a close relation exists between them and the sparsity-promoting group norms, commonly met in the context of compressive sensing (see [20], for instance). However, a significant difference is that in (11) the mixed norm is a vector-matrix norm rather than a vector-vector norm. Therefore, the sparsity is enforced on the singular values of the Hessian matrix rather than directly on its elements.

Since the Hessian of $\mathbf{x}$ at each pixel is a 2×2 symmetric matrix, the SVD in the definition of the Schatten norm reduces to the spectral decomposition and the singular values correspond to the absolute eigenvalues, which can be computed analytically. This makes the computation of the Hessian Schatten norms very efficient but also provides us the means to interpret them. If we consider the intensity map of the (continuous) image as a 3-D differentiable surface, then the two eigenvalues of the Hessian at coordinates (i, j) correspond to the principal curvatures at that point. These can be used to measure how this surface bends by different amounts in different directions at that point. Therefore, the proposed potential functions can be interpreted as scalar measurements of the curvature at a local surface patch. Note that the $\mathcal{S}_2$ 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. Since our regularizers combine these scalar curvature measurements over the whole image domain, we can infer that they provide a measure of the total curvature of the image intensity map. Consequently, as we already mentioned above, our regularizers promote piecewise-smooth reconstruction solutions.

III. POISSON IMAGE RESTORATION

A. Penalized ML Formulation

Based on the forward model (1) for the Poisson measurements and under Hessian Schatten-norm regularization (11), a penalized ML estimate $\hat{\mathbf{x}}$ of the underlying image is obtained as the minimizer

$$\hat{\mathbf{x}} = \underset{\mathbf{x} \in \mathbb{R}^N}{\arg \min} \ \left(f(\mathbf{x}) = \sum_{m=1}^{M} ([\mathbf{K}\mathbf{x}]_m - \mathbf{y}_m \log [\mathbf{K}\mathbf{x}]_m) \right.$$
$$\left. + \tau \|\mathcal{H}\mathbf{x}\|_{1,p} + \iota_{\mathcal{C}}(\mathbf{x}) \right), \quad (12)$$

where $\tau \geq 0$ is the *regularization parameter* which balances the influence of the data fidelity and the penalty term.

Next, we design different methods that efficiently deal with Problem (12). The proposed methods rely on the ADMM optimization framework and will be derived as solvers of special instances of the generic optimization task

$$\hat{\mathbf{x}} = \underset{\mathbf{x} \in \mathbb{R}^N}{\arg \min} \left(\sum_{k=1}^{K} f_k(\mathbf{A}_k \mathbf{x}) \right), \quad (13)$$

where $\mathbf{A}_k$ is a linear operator that performs the mapping $\mathbf{A}_k : \mathbb{R}^N \mapsto \mathcal{X}_k$. We further assume that each function $f_k : \mathcal{X}_k \mapsto \mathbb{R}$ in (13) is closed, proper, and convex. Our motivation for first studying (13) instead of directly dealing with (12) is that this provides a clearer insight and better justifies the splitting strategies that we pursue in Sections III-C and III-D.

B. Generic ADMM Optimization Framework

In this section, we review the ADMM optimization strategy that can be utilized to solve the generic problem (13). The form of (13) is difficult to work with, due to the coupling that exists among the functions f_k. To decouple them, we introduce the auxiliary variables $\mathbf{z}_k = \mathbf{A}_k \mathbf{x} \in \mathcal{X}_k$ and express (13) in the equivalent constrained form

$$\underset{\substack{\mathbf{x} \in \mathbb{R}^N \\ \mathbf{z} = \mathbf{A}\mathbf{x} \in \mathcal{X}}}{\min} f(\mathbf{z}) = \sum_{k=1}^{K} f_k(\mathbf{z}_k), \quad (14)$$

where $\mathbf{z} = \mathbf{A}\mathbf{x}$ denotes the mapping of $\mathbf{x}$ to the multilinear space $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_K$. In this case $\mathbf{z}$ and $\mathbf{A}$ are formed as concatenated versions of all the auxiliary variables $\mathbf{z}_k$ and the linear operators $\mathbf{A}_k$, respectively:

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_K \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \mathbf{x} \\ \vdots \\ \mathbf{A}_K \mathbf{x} \end{bmatrix} = \mathbf{A}\mathbf{x}. \quad (15)$$

Since (14) corresponds to a constrained minimization problem, we solve it by forming the augmented Lagrangian and employing the ADMM algorithm (for comprehensive and detailed reviews of this optimization strategy, see [21], [22]). Note that (14) is a special instance of the functions that can be minimized using ADMM. In particular, ADMM solves problems of the general form

$$\underset{\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} = \mathbf{c}}{\min} g(\mathbf{x}) + f(\mathbf{z}), \quad (16)$$

where $\mathbf{x} \in \mathbb{R}^N$, $\mathbf{z} \in \mathbb{R}^M$, $\mathbf{A} \in \mathbb{R}^{L \times N}$, $\mathbf{B} \in \mathbb{R}^{L \times M}$, and $\mathbf{c} \in \mathbb{R}^L$. Now, since in our case $\mathbf{A}$ corresponds to a set of linear operators we can always express the mapping $\mathbf{z} = \mathbf{A}\mathbf{x} \in \mathcal{X}$ as a matrix-vector multiplication with $\mathbf{z}$ being properly re-ordered as a vector and $\mathbf{A}$ as a matrix. Then, by choosing $\mathbf{B}$ in (16) to be a negative block-identity matrix, $\mathbf{c}$ a zero vector, and $g(\mathbf{x}) = 0$, we end up with the exact form of Problem (14). In the sequel we shall retain the operator-based notation since it is better suited to our needs.

The augmented Lagrangian for Problem (14) is

$$\mathcal{L}_\alpha(\mathbf{x}, \mathbf{z}, \boldsymbol{\eta}) = f(\mathbf{z}) + \langle \boldsymbol{\eta}, \mathbf{A}\mathbf{x} - \mathbf{z} \rangle_{\mathcal{X}} + \frac{\alpha}{2} \|\mathbf{A}\mathbf{x} - \mathbf{z}\|_{\mathcal{X}}^2, \quad (17)$$

where $\boldsymbol{\eta} \in \mathcal{X}$ corresponds to the concatenation of the K dual variables (Lagrange multipliers) $\boldsymbol{\eta}_k \in \mathcal{X}_k$, $\alpha > 0$ is a penalty parameter, and $\langle \cdot, \cdot \rangle_{\mathcal{X}}$ and $\|\cdot\|_{\mathcal{X}}$ denote the inner product and

the norm of the linear space $\mathcal{X}$, respectively. The ADMM iterations that solve (14) are given by

$$\mathbf{z}^{t+1} = \underset{\mathbf{z} \in \mathcal{X}}{\arg\min} \, f(\mathbf{z}) + \frac{\alpha}{2} \left\| \mathbf{z} - \left(\mathbf{A}\mathbf{x}^t + \mathbf{s}^t \right) \right\|_{\mathcal{X}}^2 \quad (18a)$$

$$\mathbf{x}^{t+1} = \underset{\mathbf{x} \in \mathbb{R}^n}{\arg\min} \, \left\| \mathbf{A}\mathbf{x} - \left(\mathbf{z}^{t+1} - \mathbf{s}^t \right) \right\|_{\mathcal{X}}^2 \quad (18b)$$

$$\mathbf{s}^{t+1} = \mathbf{s}^t + \mathbf{A}\mathbf{x}^{t+1} - \mathbf{z}^{t+1} \quad (18c)$$

where $\mathbf{s} = \boldsymbol{\eta}/\alpha$ is the scaled version of the dual variable. Usually in the literature, the $\mathbf{z}$- and $\mathbf{x}$-update steps appear in reverse order. In this work, however, we choose to perform the $\mathbf{z}$-update first so as to simplify the initialization process of our algorithms. Nevertheless, changing the order of the update steps does not affect the convergence of the resulting ADMM algorithm [22]. Furthermore, we observe that the argument in the $\mathbf{z}$-update of the ADMM is decoupled in K independent components. Indeed, it holds that

$$f(\mathbf{z}) + \frac{\alpha}{2} \left\| \mathbf{z} - \left(\mathbf{A}\mathbf{x}^t + \mathbf{s}^t \right) \right\|_{\mathcal{X}}^2$$
$$= \sum_{k=1}^{K} f_k(\mathbf{z}_k) + \frac{\alpha}{2} \left\| \mathbf{z}_k - \left(\mathbf{A}_k\mathbf{x}^t + \mathbf{s}_k^t \right) \right\|_{\mathcal{X}_k}^2, \quad (19)$$

and we can thus compute each $\mathbf{z}_k^{t+1}$ in (18a) independently as

$$\mathbf{z}_k^{t+1} = \text{prox}_{f_k/\alpha}\left(\mathbf{u}_k^t \right), \, \forall k = 1, \ldots, K, \quad (20)$$

where $\mathbf{u}_k^t = \mathbf{A}_k\mathbf{x}^t + \mathbf{s}_k^t \in \mathcal{X}_k$. There,

$$\text{prox}_f(\mathbf{y}) = \underset{\mathbf{x} \in \mathcal{X}}{\arg\min} \, \frac{1}{2} \left\| \mathbf{x} - \mathbf{y} \right\|_{\mathcal{X}}^2 + f(\mathbf{x}) \quad (21)$$

is the *Moreau* proximity operator (or proximal map [23]) of a function $f(\mathbf{x})$ with $\mathbf{dom} f = \mathcal{X}$, evaluated at $\mathbf{y}$. If f is a closed, proper, convex function, then the solution (21) is unique. The $\mathbf{x}$-update of ADMM corresponds to solving a quadratic minimization problem. In this case, if at least one of the operators $\mathbf{A}_k$ has an empty non-trivial nullspace[1], then the solution is unique and given by

$$\mathbf{x}^{t+1} = \left(\mathbf{A}^* \mathbf{A} \right)^{-1} \mathbf{A}^* \left(\mathbf{z}^{t+1} - \mathbf{s}^t \right)$$
$$= \left(\sum_{k=1}^{K} \mathbf{A}_k^* \mathbf{A}_k \right)^{-1} \sum_{k=1}^{K} \mathbf{A}_k^* \left(\mathbf{z}_k^{t+1} - \mathbf{s}_k^t \right), \quad (22)$$

where $\mathbf{A}_k^*$ is the adjoint operator of $\mathbf{A}_k$ that provides the mapping $\mathbf{A}_k^* : \mathcal{X}_k \mapsto \mathbb{R}^N$. Similarly to the $\mathbf{z}$-update, the $\mathbf{s}$-update in (18c) is performed in a decoupled manner as

$$\mathbf{s}_k^{t+1} = \mathbf{s}_k^t + \mathbf{A}_k\mathbf{x}^{t+1} - \mathbf{z}_k^{t+1}, \, \forall k = 1, \ldots, K, \quad (23)$$

with $\mathbf{s}_k \in \mathcal{X}_k$.

The minimization approach we considered above is quite general and apparently depends on the splitting strategy (*i.e.*, the choice of $\mathbf{A}$). This permits us to consider several possibilities for the form of $\mathbf{A}$. They will result in different variants of the ADMM algorithm. We pursue this option in Sections III-C and III-D where we design two novel algorithms that can find a numerical solution to Problem (12).

[1] A typical example of a linear operator with an empty non-trivial nullspace is the identity operator.

C. Splitting of the Data Fidelity Term

One way to decouple the objective function $f(\mathbf{x})$ is to formulate the unconstrained minimization problem (12) as

$$\min_{\substack{\mathbf{x} \in \mathbb{R}^N \\ \mathbf{z}_k = \mathbf{A}_k\mathbf{x} \in \mathcal{X}_k, k=1,2}} f_1(\mathbf{z}_1) + f_2(\mathbf{z}_2) \quad (24)$$

where $\mathbf{A}_1 = \mathbf{K}$, $\mathbf{A}_2 = \mathbf{I}$, and

$$f_1(\mathbf{z}_1) = \sum_{m=1}^{M} \left([\mathbf{z}_1]_m - \mathbf{y}_m \log [\mathbf{z}_1]_m \right)$$
$$f_2(\mathbf{z}_2) = \tau \left\| \mathcal{H}\mathbf{z}_2 \right\|_{1,p} + \iota_{\mathcal{C}}(\mathbf{z}_2). \quad (25)$$

What remains now is to specify the form of the ADMM update steps described in (18).

To perform the $\mathbf{z}$-update of (18a) we need to compute the proximal maps of the functions defined in (25). The proximal map $\text{prox}_{f_1/\alpha}\left(\mathbf{u}_1^t \right)$, with $\mathbf{u}_1^t = \mathbf{K}\mathbf{x}^t + \mathbf{s}_1^t$, is separable and is computed component-wise. The solution for each component of the auxiliary variable $\mathbf{z}_1$ is obtained by solving a minimization problem of the form

$$\underset{z \in \mathbb{R}_+}{\arg\min} \, \frac{1}{2}(z - u)^2 + \frac{1}{\alpha}(z - y \log(z)). \quad (26)$$

Since the function under minimization is smooth, the solution corresponds to the positive root of a second-order polynomial. Based on this fact, it is simple to show that

$$\mathbf{z}_1^{t+1} = \frac{1}{2}\left(\mathbf{u}_1^t - \frac{1}{\alpha} + \sqrt{\left(\mathbf{u}_1^t - \frac{1}{\alpha} \right)^2 + \frac{4\mathbf{y}}{\alpha}} \right), \quad (27)$$

where all the operations are performed component-wise.

To evaluate the proximal map $\text{prox}_{f_2/\alpha}\left(\mathbf{u}_2^t \right)$, with $\mathbf{u}_2^t = \mathbf{x}^t + \mathbf{s}_2^t$, we employ the iterative algorithm we recently introduced in [16]. This method finds a numerical solution to the minimization problem

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\arg\min} \, \frac{1}{2} \left\| \mathbf{x} - \mathbf{y} \right\|_2^2 + \tau \left\| \mathcal{H}\mathbf{x} \right\|_{1,p} + \iota_{\mathcal{D}}(\mathbf{x}) \quad \forall p \geq 1, \quad (28)$$

where $\mathcal{D}$ is a convex set (in our case $\mathcal{D} \equiv \mathcal{C}$) and $\iota_{\mathcal{D}}$ represents additional constraints on the solution. As in the case of the numerical methods developed for computing the proximal map of other functionals, such as TV, our algorithm for the Hessian Schatten-norm regularizers is only meant to run for a finite number of iterations. Therefore, the practical result will be an approximation of the solution. However, this is not going to jeopardize the convergence of the overall algorithm, thanks to a theorem provided by Eckstein and Bertsekas [24]. According to this theorem, even an inexact optimization of the internal ADMM sub-problems is sufficient to guarantee convergence, under the requirement that the approximation error of the sub-problems is absolute-summable. To that end, similar to the strategy adopted for the TV case in [13], we initialize the internal variables of the algorithm we designed in [16] with those obtained in the previous ADMM iteration.

The $\mathbf{x}$-update of the algorithm depends on the solution of the quadratic minimization problem that corresponds to solving

the following system of linear equations:

$$\mathbf{x}^{t+1} = \left(\mathbf{K}^T\mathbf{K} + \mathbf{I}\right)^{-1} \left(\mathbf{K}^T \left(\mathbf{z}_1^{t+1} - \mathbf{s}_1^t\right) + \left(\mathbf{z}_2^{t+1} - \mathbf{s}_2^t\right)\right). \tag{29}$$

In practice, due to its large size, the $\left(\mathbf{K}^T\mathbf{K} + \mathbf{I}\right)^{-1}$ operator cannot be computed. However, if we assume periodic boundary conditions for the image, then a blurring operation corresponds to circular convolution and the operator $\mathbf{K}^T\mathbf{K} + \mathbf{I}$ is diagonalized by the fast Fourier transform (FFT) [25]. In this case, the $\mathbf{x}$-update (29) is computed very efficiently in a single step without having to resort to an iterative method. The same holds if we assume Neumann (mirror) boundary conditions, where the operator is now diagonalizable by the discrete cosine transform (DCT) [25]. Moreover, we note that the matrix inversion can also be computed very efficiently for a number of other types of linear degradation operators (see [26]). Finally, for the case of inverse problems where the degradation operator $\mathbf{K}$ does not exhibit any special structure, we can find the solution of (29) employing the conjugate-gradient (CG) method [27]. An approach that we have found to work well in practice and does not introduce a significant computational overhead is to employ CG with a "warm-start" strategy. This strategy amounts to choosing the starting point of CG to be the solution of the quadratic sub-problem of the previous ADMM iteration. We have observed that by following this simple strategy and by just running a few CG iterations we can efficiently minimize the overall objective function without compromising the convergence of our algorithm (results are not reported here). In fact, we manage to obtain almost exactly the same solution and at the same rate of convergence with the solution one gets by computing the $\mathbf{x}$-update in closed-form. Similar findings to ours have also been recently reported in [28], where the authors employ an ADMM strategy for differential phase contrast X-ray tomogram reconstruction. Whenever this is possible, the number of CG iterations can be further reduced by employing a preconditioned conjugate gradient algorithm.

A summary of the overall proposed numerical algorithm, which we will refer to as HSPIRAL1 (Hessian Schatten-norm Poisson Image Reconstruction by Augmented Lagrangian), is provided in Algorithm 1.

D. Splitting of the Data Fidelity and Penalty Terms

Another way to decouple the objective function $f(\mathbf{x})$ is to formulate the unconstrained minimization problem (12) as

$$\min_{\substack{\mathbf{x}\in\mathbb{R}^N \\ \mathbf{z}_k = \mathbf{A}_k\mathbf{x}\in\mathcal{X}_k, k=1,\dots,3}} f_1(\mathbf{z}_1) + f_2(\mathbf{z}_2) + f_3(\mathbf{z}_3) \tag{30}$$

where $\mathbf{A}_1 = \mathbf{K}$, $\mathbf{A}_2 = \mathcal{H}$, $\mathbf{A}_3 = \mathbf{I}$, and

$$f_1(\mathbf{z}_1) = \sum_{m=1}^{M} \left([\mathbf{z}_1]_m - \mathbf{y}_m \log [\mathbf{z}_1]_m\right)$$
$$f_2(\mathbf{z}_2) = \tau \|\mathbf{z}_2\|_{1,p} \tag{31}$$
$$f_3(\mathbf{z}_3) = \iota_{\mathcal{C}}(\mathbf{z}_3).$$

The operator $\mathbf{A}_1$ and the function f_1 in (31) are the same as those in our previous splitting approach. Therefore, the

---

Algorithm 1 Hessian Schatten-norm Poisson Image Reconstruction by Augmented Lagrangian (HSPIRAL1).

Input: $\mathbf{y}$, $\mathbf{K}$, $\tau > 0$, $\alpha > 0$, $p \geq 1$, $\mathcal{P}_{\mathcal{D}}$.

Initialization: $\mathbf{x}^0 = \mathbf{y}$, $\mathbf{s}_1^0 = \mathbf{0}$, $\mathbf{s}_2^0 = \mathbf{0}$.

Output: $\hat{\mathbf{x}}$ (optimal solution of (12)).

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

$\quad \mathbf{u}_1^t \leftarrow \mathbf{K}\mathbf{x}^t + \mathbf{s}_1^t$;

$\quad \mathbf{u}_2^t \leftarrow \mathbf{x}^t + \mathbf{s}_2^t$;

$\quad \mathbf{z}_1^{t+1} \leftarrow \frac{1}{2}\left(\mathbf{u}_1^t - \frac{1}{\alpha} + \sqrt{\left(\mathbf{u}_1^t - \frac{1}{\alpha}\right)^2 + \frac{4\mathbf{y}}{\alpha}}\right)$;

$\quad \mathbf{z}_2^{t+1} \leftarrow \arg\min_{\mathbf{z}_2\in\mathbb{R}^n} \frac{1}{2}\left\|\mathbf{z}_2 - \mathbf{u}_2^t\right\|_2^2 + \frac{\tau}{\alpha}\left\|\mathcal{H}\mathbf{z}_2\right\|_{1,p} + \iota_{\mathcal{D}}(\mathbf{z}_2)$;

$\quad \mathbf{v}^t \leftarrow \mathbf{K}^T\left(\mathbf{z}_1^{t+1} - \mathbf{s}_1^t\right) + \left(\mathbf{z}_2^{t+1} - \mathbf{s}_2^t\right)$;

$\quad \mathbf{x}^{t+1} \leftarrow \left(\mathbf{K}^T\mathbf{K} + \mathbf{I}\right)^{-1}\mathbf{v}^t$;

$\quad \mathbf{s}_1^{t+1} \leftarrow \mathbf{s}_1^t + \mathbf{K}\mathbf{x}^{t+1} - \mathbf{z}_1^{t+1}$;

$\quad \mathbf{s}_2^{t+1} \leftarrow \mathbf{s}_2^t + \mathbf{x}^{t+1} - \mathbf{z}_2^{t+1}$;

$\quad t \leftarrow t + 1$;

end

return $\mathbf{x}^t$;

---

proximal map $\text{prox}_{f_1/\alpha}\left(\mathbf{u}_1^t\right)$, with $\mathbf{u}_1^t = \mathbf{K}\mathbf{x}^t + \mathbf{s}_1^t$, is given by (27). Moreover, since f_3 is an indicator function, its proximal map corresponds to the projection onto the convex set $\mathcal{C}$. This projection can be computed as

$$\text{prox}_{f_3/\alpha}\left(\mathbf{u}_3^t\right) = \max\left(\mathbf{u}_3^t, \mathbf{0}\right), \tag{32}$$

where $\mathbf{u}_3^t = \mathbf{x}^t + \mathbf{s}_3^t$, $\mathbf{0}$ is a zero vector of the same size as $\mathbf{u}_3^t$, and the maximum operation in (32) is computed in a component-wise fashion. Lastly, to complete with the $\mathbf{z}$-updates we also need to compute the proximal map related to the function f_2. To do so, we first note that $\mathbf{A}_2$ corresponds to the Hessian operator. Thus, it holds that $\mathbf{z}_2 \in \mathcal{K}$. In addition, we can also show that $\text{prox}_{f_2/\alpha}\left(\mathbf{u}_2^t\right)$ is separable, with $\mathbf{u}_2^t = \mathcal{H}\mathbf{x}^t + \mathbf{s}_2^t \in \mathcal{K}$. Indeed, for a variable $\mathbf{Y} \in \mathcal{K}$ we have that

$$\text{prox}_{f_2/\alpha}(\mathbf{Y}) = \arg\min_{\mathbf{X}\in\mathcal{K}} \frac{1}{2}\|\mathbf{X} - \mathbf{Y}\|_{\mathcal{K}}^2 + \frac{\tau}{\alpha}\|\mathbf{X}\|_{1,p}$$

$$= \arg\min_{\mathbf{X}_n\in\mathbb{R}^{2\times 2}} \sum_{n=1}^{N} \frac{1}{2}\|\mathbf{X}_n - \mathbf{Y}_n\|_F^2 + \frac{\tau}{\alpha}\|\mathbf{X}_n\|_{\mathcal{S}_p}. \tag{33}$$

From (33), we easily verify that the minimization can be performed independently for each matrix component $\mathbf{Y}_n$ of $\mathbf{Y}$. Therefore, the computation of the proximal map of f_2 boils down to computing the proximal maps of N functions defined as the $\mathcal{S}_p$ norms of a 2×2 matrix multiplied by a constant:

$$\text{prox}_{\tau\|\cdot\|_{\mathcal{S}_p}}(\mathbf{Y}) = \arg\min_{\mathbf{X}\in\mathbb{R}^{2\times 2}} \frac{1}{2}\|\mathbf{X} - \mathbf{Y}\|_F^2 + \tau\|\mathbf{X}\|_{\mathcal{S}_p}. \tag{34}$$

This operation can be performed efficiently based on the Proposition 1 which provides a direct link between the proximal map of an ℓ_p vector-norm and the proximal map of an $\mathcal{S}_p$ matrix-norm. This result is new, to the best of our knowledge.

Proposition 1: Let $\mathbf{Y} \in \mathbb{C}^{n_1 \times n_2}$ be a matrix with the SVD decomposition $\mathbf{Y} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$, where $\mathbf{U} \in \mathbb{U}^{n_1}$, $\mathbf{V} \in \mathbb{U}^{n_2}$, and $\boldsymbol{\Sigma} \in \mathbb{D}^{n_1 \times n_2}$. If $\psi_p(\cdot) = \tau \left\| \cdot \right\|_{\mathcal{S}_p}$ and $\phi_p(\cdot) = \tau \left\| \cdot \right\|_p$, then the proximal map

$$\hat{\mathbf{X}} = \text{prox}_{\psi_p}(\mathbf{Y}) \tag{35}$$

is equal to

$$\hat{\mathbf{X}} = \mathbf{U}\text{diag}\left(\text{prox}_{\phi_p}(\boldsymbol{\sigma})\right)\mathbf{V}^H, \tag{36}$$

where $\text{diag}(\cdot)$ is the operator that transforms a vector to a diagonal matrix and $\boldsymbol{\sigma}$ denotes the vector with the singular values of $\mathbf{Y}$.

Proof: Since all Schatten norms are unitarily invariant, the proximal map (35) can be equivalently written as

$$\begin{aligned} \hat{\mathbf{X}} = \text{prox}_{\psi_p}(\mathbf{Y}) &= \arg\min_{\mathbf{X}} \frac{1}{2}\left\|\mathbf{X} - \mathbf{Y}\right\|_F^2 + \tau\left\|\mathbf{X}\right\|_{\mathcal{S}_p} \\ &= \arg\min_{\mathbf{X}} \frac{1}{2}\left\|\mathbf{U}^H\mathbf{X}\mathbf{V} - \mathbf{U}^H\mathbf{Y}\mathbf{V}\right\|_F^2 + \tau\left\|\mathbf{U}^H\mathbf{X}\mathbf{V}\right\|_{\mathcal{S}_p}. \end{aligned} \tag{37}$$

Let us now consider the matrix $\mathbf{Z} = \mathbf{U}^H\mathbf{X}\mathbf{V}$ that is associated with the solution of (37). If we substitute $\mathbf{Z}$ in (37), then we end up with the regularized problem

$$\hat{\mathbf{Z}} = \arg\min_{\mathbf{Z}} \underbrace{\frac{1}{2}\left\|\mathbf{Z} - \boldsymbol{\Sigma}\right\|_F^2 + \tau\left\|\mathbf{Z}\right\|_{\mathcal{S}_p}}_{h(\mathbf{Z};\boldsymbol{\Sigma})}. \tag{38}$$

Next, we show that the optimal solution of (38) should be a positive semidefinite diagonal matrix. Indeed, we have

$$\begin{aligned} h\left(\hat{\mathbf{Z}}; \boldsymbol{\Sigma}\right) &= \frac{1}{2}\|\hat{\mathbf{Z}}\|_F^2 + \frac{1}{2}\|\boldsymbol{\Sigma}\|_F^2 - \text{Re}\left(\text{tr}\left(\hat{\mathbf{Z}}^H\boldsymbol{\Sigma}\right)\right) + \tau\|\hat{\mathbf{Z}}\|_{\mathcal{S}_p} \\ &\geq \frac{1}{2}\|\hat{\boldsymbol{\Sigma}}\|_F^2 + \frac{1}{2}\|\boldsymbol{\Sigma}\|_F^2 - \text{tr}\left(\hat{\boldsymbol{\Sigma}}^T\boldsymbol{\Sigma}\right) + \tau\|\hat{\boldsymbol{\Sigma}}\|_{\mathcal{S}_p} \\ &= \frac{1}{2}\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_F^2 + \tau\|\hat{\boldsymbol{\Sigma}}\|_{\mathcal{S}_p} = h\left(\hat{\boldsymbol{\Sigma}}; \boldsymbol{\Sigma}\right), \end{aligned} \tag{39}$$

where $\hat{\boldsymbol{\Sigma}}$ is a diagonal matrix with the singular values of $\hat{\mathbf{Z}}$. We note that the inequality in (39) follows from

$$\text{Re}\left(\text{tr}\left(\mathbf{A}^H\mathbf{B}\right)\right) \leq \langle \boldsymbol{\sigma}_{\mathbf{A}}, \boldsymbol{\sigma}_{\mathbf{B}}\rangle_2, \tag{40}$$

where $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{n_1 \times n_2}$ and $\boldsymbol{\sigma}_{\mathbf{A}}, \boldsymbol{\sigma}_{\mathbf{B}}$ are the vectors with the singular values of $\mathbf{A}$ and $\mathbf{B}$, respectively. The inequality in (40) stems from von Neumann's trace theorem [29]. Further, the equality $\|\hat{\mathbf{Z}}\|_{\mathcal{S}_p} = \|\hat{\boldsymbol{\Sigma}}\|_{\mathcal{S}_p}, \forall p \geq 1$, is a direct consequence of the Schatten-norm definition.

Now, we argue that, since $\hat{\mathbf{Z}}$ is the minimizer of the function h, $h\left(\hat{\mathbf{Z}}; \boldsymbol{\Sigma}\right)$ should be the smallest of all values. However, in (39) we obtained a lower bound on this value, which immediately implies that $h\left(\hat{\mathbf{Z}}; \boldsymbol{\Sigma}\right) = h\left(\hat{\boldsymbol{\Sigma}}; \boldsymbol{\Sigma}\right)$ and therefore $\hat{\mathbf{Z}} = \hat{\boldsymbol{\Sigma}}$. This means that $\hat{\mathbf{Z}}$ is a positive semidefinite diagonal matrix. We then express the solution of (38) as $\hat{\mathbf{Z}} = \text{diag}(\hat{\mathbf{z}})$, where $\hat{\mathbf{z}}$ is given by

$$\hat{\mathbf{z}} = \text{prox}_{\phi_p}(\boldsymbol{\sigma}) = \arg\min_{\mathbf{z} \in \mathbb{R}^{\min(n_1, n_2)}} \frac{1}{2}\left\|\mathbf{z} - \boldsymbol{\sigma}\right\|_2^2 + \tau\left\|\mathbf{z}\right\|_p, \tag{41}$$

and $\boldsymbol{\sigma}$ is the vector formed by the diagonal entries of $\boldsymbol{\Sigma}$. Using this result and the relation between the optimal solutions of (37) and (38), we are finally led to (36). ∎

Based on Proposition 1, we design an algorithm to evaluate the proximal map of an $\mathcal{S}_p$ norm. The algorithm consists of three steps: **(a)** decompose $\mathbf{Y}$ in its singular vectors and singular values by means of the SVD; **(b)** evaluate the proximal map of the ℓ_p norm at $\boldsymbol{\sigma}$ (a vector with the singular values of $\mathbf{Y}$); and **(c)** obtain the final result via singular-value reconstruction (SVR) using the result obtained at step **(b)**.

As it turns out, this algorithm depends on our ability to evaluate the proximal map of the ℓ_p norm. Fortunately, this dependency does not introduces any difficulties thanks to the efficient proximity algorithm for ℓ_p norms that has been recently proposed in [30]. Moreover, in Section III-E we report three cases of $\mathcal{S}_p$ norms, for $p = 1, 2, \infty$, where their proximal map can be evaluated in closed form.

The **x**-update for the current splitting approach is obtained by solving a system of linear equations of the form

$$\begin{aligned} \mathbf{x}^{t+1} = &\left(\mathbf{K}^T\mathbf{K} + \mathcal{H}^*\mathcal{H} + \mathbf{I}\right)^{-1}\left(\mathbf{K}^T\left(\mathbf{z}_1^{t+1} - \mathbf{s}_1^t\right)\right. \\ &\left. + \mathcal{H}^*\left(\mathbf{z}_2^{t+1} - \mathbf{s}_2^t\right) + \left(\mathbf{z}_3^{t+1} - \mathbf{s}_3^t\right)\right), \end{aligned} \tag{42}$$

with $\mathcal{H}^*\mathcal{H} = \Delta_{r_1r_1}^*\Delta_{r_1r_1} + 2\Delta_{r_1r_2}^*\Delta_{r_1r_2} + \Delta_{r_2r_2}^*\Delta_{r_2r_2}$. Similarly to the **x**-update of the HSPIRAL1 algorithm that we proposed in Section III-C, we can either employ the CG method or assume periodic or mirror boundary conditions for the image to find the solution of (42). Likewise, the adequate choice for the extension of the image leads to an efficient computation without the need of an iterative method. A summary of the overall proposed numerical algorithm which we will refer to as HSPIRAL2, is provided in Algorithm 2.

A last comment on HSPIRAL1 and HSPIRAL2 is that they can be viewed as extensions of the PIDAL-TV [13] and PIDSplit+ [14] algorithms. These numerical methods perform image reconstruction under Poisson noise subject to TV regularization, while our proposed algorithms perform Poisson image reconstruction subject to Hessian Schatten-norm regularization. Therefore, while PIDAL-TV and PIDSplit+ can handle linear vector-valued regularization operators, HSPIRAL1 and HSPIRAL2 are extensions to linear matrix-valued operators. Also note that the splitting approach we use in HSPIRAL1 is novel and differs from the one in [13], reducing the number of auxiliary functions from three to two. Finally, we would like to emphasize that our proposed methods are not limited to the use of the Hessian operator but are general enough to accommodate other (real or complex) linear matrix-valued regularization operators.

E. Closed Form of $\mathcal{S}_p$-Norm Proximal Maps for $p = 1, 2, \infty$

From Proposition 1, we have that the proximal map of the $\mathcal{S}_1$ norm is linked to the proximal map of the ℓ_1 norm. The latter is computed by applying the soft-thresholding operator $S_\gamma(\boldsymbol{\sigma}) = \max(\boldsymbol{\sigma} - \gamma, 0)$ [31], where the max is applied component-wise and γ is a threshold. Therefore, we have

$$\text{prox}_{\tau\|\cdot\|_{\mathcal{S}_1}}(\mathbf{Y}) = \mathbf{U}\text{diag}\left(S_\tau(\boldsymbol{\sigma})\right)\mathbf{V}^H, \tag{43}$$

Algorithm 2 Hessian Schatten-norm Poisson Image Reconstruction by Augmented Lagrangian (HSPIRAL2).

Input: $\mathbf{y}$, $\mathbf{K}$, $\tau > 0$, $\alpha > 0$, $p \geq 1$.

Initialization: $\mathbf{x}^0 = \mathbf{y}$, $\mathbf{s}_1^0 = \mathbf{0}$, $\mathbf{s}_2^0 = \mathbf{0}$, $\mathbf{s}_3^0 = \mathbf{0}$.

Output: $\hat{\mathbf{x}}$ (optimal solution of (12)).

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

$\quad \mathbf{u}_1^t \leftarrow \mathbf{K}\mathbf{x}^t + \mathbf{s}_1^t$;

$\quad \mathbf{u}_2^t \leftarrow \mathcal{H}\mathbf{x}^t + \mathbf{s}_2^t$;

$\quad \mathbf{u}_3^t \leftarrow \mathbf{x}^t + \mathbf{s}_3^t$;

$\quad \mathbf{z}_1^{t+1} \leftarrow \frac{1}{2}\left(\mathbf{u}_1^t - \frac{1}{\alpha} + \sqrt{\left(\mathbf{u}_1^t - \frac{1}{\alpha}\right)^2 + \frac{4\mathbf{y}}{\alpha}}\right)$;

$\quad \mathbf{z}_2^{t+1} \leftarrow \underset{\mathbf{z}_2 \in \mathcal{K}}{\arg\min} \, \frac{1}{2}\left\|\mathbf{z}_2 - \mathbf{u}_2^t\right\|_{\mathcal{K}}^2 + \frac{\tau}{\alpha}\left\|\mathbf{z}_2\right\|_{1,p}$;

$\quad \mathbf{z}_3^{t+1} \leftarrow \max\left(\mathbf{u}_3^t, \mathbf{0}\right)$;

$\quad \mathbf{v}^t \leftarrow \mathbf{K}^T\left(\mathbf{z}_1^{t+1} - \mathbf{s}_1^t\right) + \mathcal{H}^*\left(\mathbf{z}_2^{t+1} - \mathbf{s}_2^t\right) + \left(\mathbf{z}_3^{t+1} - \mathbf{s}_3^t\right)$;

$\quad \mathbf{x}^{t+1} \leftarrow \left(\mathbf{K}^T\mathbf{K} + \mathcal{H}^*\mathcal{H} + \mathbf{I}\right)^{-1}\mathbf{v}^t$;

$\quad \mathbf{s}_1^{t+1} \leftarrow \mathbf{s}_1^t + \mathbf{K}\mathbf{x}^{t+1} - \mathbf{z}_1^{t+1}$;

$\quad \mathbf{s}_2^{t+1} \leftarrow \mathbf{s}_2^t + \mathcal{H}\mathbf{x}^{t+1} - \mathbf{z}_2^{t+1}$;

$\quad \mathbf{s}_3^{t+1} \leftarrow \mathbf{s}_3^t + \mathbf{x}^{t+1} - \mathbf{z}_3^{t+1}$;

$\quad t \leftarrow t + 1$;

end

return $\mathbf{x}^t$;

---

where $\boldsymbol{\sigma}$ denotes the vector with the singular values of $\mathbf{Y}$. Note that this result corresponds to the singular-value thresholding (SVT) method [32], [33], first developed in the field of low-rank matrix reconstruction. The derivation of the SVT method in [32], [33] is quite technical. It relies on the characterization of the subgradient of the nuclear norm [34]. By contrast, our result comes out naturally as an immediate consequence of Proposition 1. To obtain the proximal map of the $\mathcal{S}_2$ norm, we use Proposition 1 and the proximal map of the ℓ_2 norm [21, Section 4.2]. Then, it is simple to show that

$$\text{prox}_{\tau\|\cdot\|_{\mathcal{S}_2}}(\mathbf{Y}) = \begin{cases} \mathbf{O}, & \text{if } \|\mathbf{Y}\|_{\mathcal{S}_2} \leq \tau \\ \left(\frac{\|\mathbf{Y}\|_{\mathcal{S}_2} - \tau}{\|\mathbf{Y}\|_{\mathcal{S}_2}}\right)\mathbf{Y}, & \text{if } \|\mathbf{Y}\|_{\mathcal{S}_2} > \tau, \end{cases} \quad (44)$$

where $\mathbf{O}$ is a zero matrix of the same size as $\mathbf{Y}$. This situation is advantageous since it allows us to avoid both the SVD and the SVR steps. Consequently, this drastically reduces the complexity for the computation of the proximity operator. Finally, for the $\mathcal{S}_\infty$ norm, the result depends on the proximal map of the ℓ_∞ norm

$$\text{prox}_{\tau\|\cdot\|_\infty}(\boldsymbol{\sigma}) = \min\left(\boldsymbol{\sigma}, \tau\gamma \cdot \mathbf{1}\right), \quad (45)$$

where $\mathbf{1}$ is a vector of ones and the min operation is computed component-wise. In general, the threshold γ cannot be obtained in closed form, so we have to recourse to one of the methods available in the literature [35]–[38] to find it. However, in our case, for $\boldsymbol{\sigma} \in \mathbb{R}_+^2$, we can compute it

analytically

$$\gamma = \begin{cases} 0, & \text{if } \sigma_1 \leq 1 - \sigma_2, \\ \frac{\sigma_1 + \sigma_2 - 1}{2}, & \text{if } 1 - \sigma_2 < \sigma_1 \leq 1 + \sigma_2, \\ \sigma_1 - 1, & \text{if } \sigma_1 > 1 + \sigma_2, \end{cases} \quad (46)$$

where σ_1, σ_2 are the largest and smallest elements of $\boldsymbol{\sigma}$, respectively. Now, we express the $\mathcal{S}_\infty$ proximal map as

$$\text{prox}_{\tau\|\cdot\|_{\mathcal{S}_\infty}}(\mathbf{Y}) = \mathbf{U}\text{diag}\left(\min\left(\boldsymbol{\sigma}, \tau\gamma \cdot \mathbf{1}\right)\right)\mathbf{V}^H. \quad (47)$$

IV. EXPERIMENTAL RESULTS

Our goal in this section is to assess experimentally the quality of reconstruction obtained by using penalty terms from the Hessian Schatten-norm family. Additionally, we want to illustrate the effectiveness of the proposed optimization strategies for the problem of Poisson image restoration. Our results are compared against the ones obtained by using three alternative regularizers; namely, TV and the fully redundant Haar and Daubechies (with four vanishing moments, DB4) wavelet transforms. In the wavelet case, we focus on FA regularization to deblur images since it has been reported in the literature (c.f [13], [39]) that FS usually leads to inferior results. In particular, the regularizers we employ correspond to the ℓ_1 norm of the wavelet coefficients. If we denote the wavelet frame-analysis operator by Φ, then the regularizer of choice is $\|\mathbf{S}\Phi\mathbf{x}\|_1$, where $\mathbf{S}$ is a masking operator which zeroes out the scaling coefficients so that they do not influence the value of the penalty function. This is common practice in wavelet-based techniques since the wavelet coefficients are sparse while the scaling coefficients, which are related to the low-frequency content of the image, are not.

A. Algorithmic Performance

We examine the convergence performance of the two proposed algorithms, HSPIRAL1 and HSPIRAL2. To do so, we consider the restoration of a degraded version of the Face image shown in Fig. 1 using the $\mathcal{HS}_2$ regularizer (11). The original image is scaled to a maximum intensity of $M = 25$ and its quality is degraded by Gaussian blurring of standard deviation $\sigma_b = 4$ and subsequent "addition" of Poisson noise. Since the two algorithms under study involve the solution of different sub-problems, their computational complexity is not the same. In particular, a main difference is that the proximal map (28) of HSPIRAL1 cannot be obtained in closed-form and needs to be computed iteratively. To estimate this proximal map we use just 5 inner iterations, which from our experience does not seem to compromise the convergence of the overall algorithm. Then, we run the two methods for the same amount of time. In both cases, we use the observed degraded image as the initial solution and set the penalty parameter as $\alpha = 60\tau/M$. In Fig. 3(a) we present for both algorithms the evolution of the objective cost function $f(\mathbf{x})$ in (12). We present in Fig. 3(b) the evolution of the normed residual error $\|\mathbf{z} - \mathbf{A}\mathbf{x}\|_{\mathcal{X}}$. This residual error indicates how efficiently the equality constraints, which involve the auxiliary variables that are used in the splitting process, are enforced. The two plots of Fig. 3 allow us to conclude that

Fig. 1.   Set of natural test images. From left to right: Boat, Face, Hill, House, Kids, Peppers, and Wall.
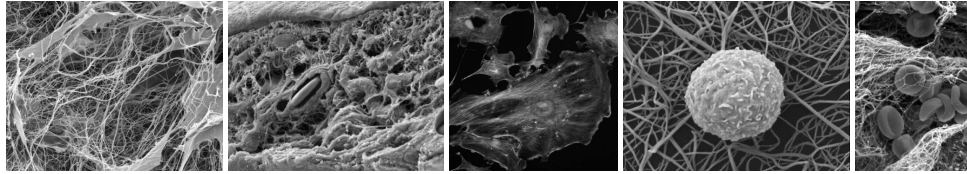

Fig. 2.   Set of biological test images. From left to right: Chick chorioallantoic membrane (CAM), CAM2, Fluorescent cell, Leukocyte (White blood cell), and Thrombus (blood clot).

HSPIRAL1 converges faster than HSPIRAL2 for the Face image. However, since the penalty parameter α in general influences significantly the speed of convergence of every ADMM-based algorithm, this result is not conclusive about the relative efficiency of the two algorithms. Still, combined with the theoretical guarantee of convergence for HSPIRAL2, this result can serve as an indication that HSPIRAL1 behaves well despite the approximate evaluation of the proximal map (28).

B. Restoration Settings

For the comparisons among the different regularization schemes we use two distinct set of images. The first set is composed of eight grayscale natural images, that is the images shown in Fig. 1 plus the standard Lena image[2]. Four of them are well-known standard test images of size 512×512 pixel (except the Peppers image which is of size 256×256) with the additional images[3] being of size 255×255. Given that our main interest leans toward biomedical imaging, we also conduct experiments on a second set composed of the five biological images shown in Fig. 2. Four of these images are part of the biomedical image database [40], they were converted to grayscale and resized. The largest resulting image has a dimension of 314×402 pixel. The last member of this set is of size 512×512.

The performance of every regularizer under comparison is assessed for various blurring kernels and different noise levels. In particular, we employ three point-spread functions (PSFs) to produce blurred versions of the natural 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. To simulate various SNR values of Poisson noise we scale the images to have a maximum intensity of (5, 25, 100, 255). Since Poisson noise is signal-dependent with local SNR $= \sqrt{\lambda_k}$, where λ_k denotes the underlying image intensity at position k, the relative amount of noise increases as the maximum
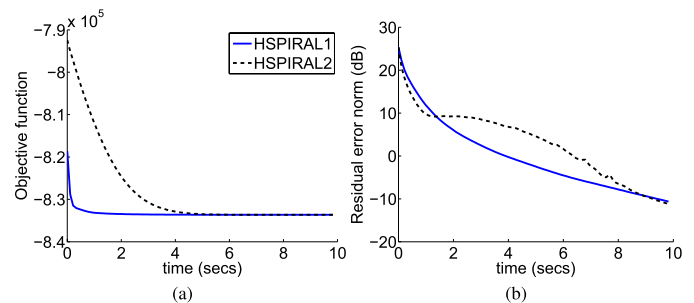

Fig. 3.   Comparisons of the speed of convergence among HSPIRAL1 and HSPIRAL2. Evolution of (a) the objective cost function $f(\mathbf{x})$ of (12) and (b) the normed residual error $\|\mathbf{z} - \mathbf{Ax}\|_{\mathcal{X}}$.

intensity of the image decreases. The mean intensity for each maximum intensity level, for all natural images, varies in the ranges of $[1.20, 2.54]$, $[5.98, 12.72]$ $[23.92, 50.97]$, $[61, 129.71]$ covering a wide gamut of noise levels. For the set of biological images we use two PSFs that can better approximate the spatial response of a microscope, namely, a Gaussian PSF of standard deviation $\sigma_b = 4$ and support size 7×7 and an airy disk PSF of support 31×31. The latter was obtained using our software[4] according to the Richards-Wolf PSF model [41]. For this image set, the Poisson noise levels are generated with the images scaled to a maximum intensity of (10, 50, 250, 500). In this case, the mean intensity for each maximum intensity level varies in the ranges of $[1.76, 4.73]$, $[8.82, 23.66]$, $[44.08, 118.31]$, $[88.17, 236.61]$.

Regarding the minimization of the objective functions, for the Hessian Schatten-norm regularizers we report the results obtained using HSPIRAL1. For the TV seminorm and the two wavelet-based penalties we use a similar optimization strategy with the difference being that, instead of the proximal map (28), we evaluate the proximal maps that are related to properly modified versions of f_2 in (25). This evaluation is performed by the algorithm proposed in [42] which is an extended version of Chambolle's dual algorithm [43] for constrained optimization. The adopted minimization approach for these three regularizers can be considered as a modification

---

[2]Due to a reviewer's request we removed the Lena image from Fig. 1.

[3]These images along with the motion-blur kernel used in the experiments can be obtained from http://www.wisdom.weizmann.ac.il/~levina/papers/LevinEtalCVPR09Data.rar.

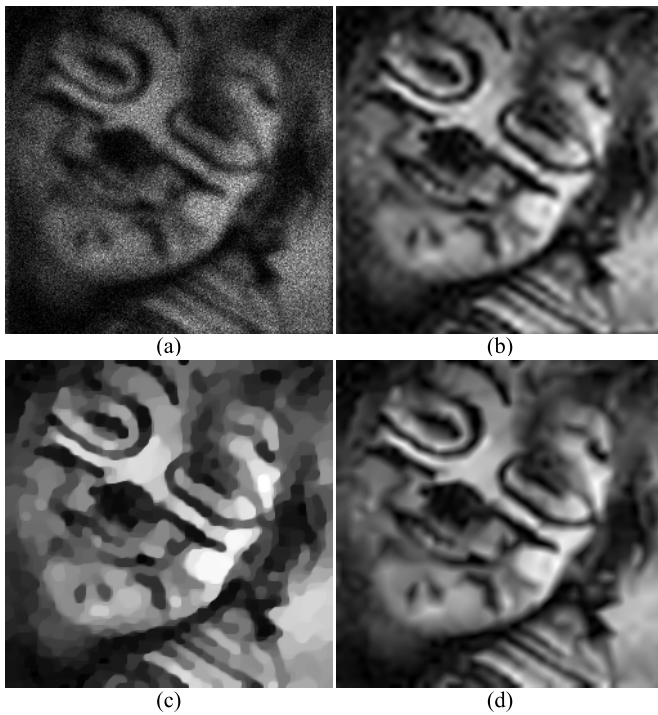[4]Software is available at http://bigwww.epfl.ch/algorithms/psfgenerator

Fig. 4. Restoration of the Face image, scaled to a maximum intensity of $M = 25$ and degraded by Gaussian blurring and Poisson noise. (a) Degraded image (PSNR = 17.01 dB), (b) DB$_4$ result (PSNR = 23.30 dB), (c) TV result (PSNR = 23.10 dB), and (d) $\mathcal{HS}_1$ result (PSNR = 23.90 dB).

Fig. 5. Restoration of the House image, scaled to a maximum intensity of $M = 100$ and degraded by motion blurring and Poisson noise. (a) Degraded image (PSNR = 19.92 dB), (b) Haar result (PSNR = 23.14 dB), (c) TV result (PSNR = 23.13 dB), and (d) $\mathcal{HS}_2$ result (PSNR = 23.71 dB).

of the PIDAL-TV algorithm [13] that reduces the required number of auxiliary splitting variables. The rationale for our choice, is that the comparison among the different regularizers is more meaningful that way, since the quality of the restoration does not depend on the choice of the minimization strategy but rather on the choice of the regularizer. After all, PIDAL-TV is ranked among the most efficient optimization techniques for Poisson inverse problems under TV regularization.

To ensure convergence of the methods, the stopping criterion is always triggered by a relative normed difference of 10^{-5} between two successive estimates, or by a maximum of 400 iterations. Further, we use 5 inner iterations for computing the proximal maps of all the different versions of f_2 in (25). Finally, the quality of the reconstructed images is assessed in terms of PSNR measured in dB and defined as $10 \log_{10} \left(M^2 / \text{MSE} \right)$, where M denotes the maximum intensity of the underlying image and MSE stands for the mean squared error between the restored and the noiseless image.

C. Poisson Deblurring of Natural Images

In Table I, we provide comparative Poisson image restoration results for the set of natural images and for many combinations of blurring and noise levels. Regarding the Hessian Schatten-norm regularizers, we report the results obtained employing Schatten norms or order one, two, and infinity. They correspond to the nuclear, Frobenius, and spectral matrix norms, respectively. The PSNR scores are pooled over 10
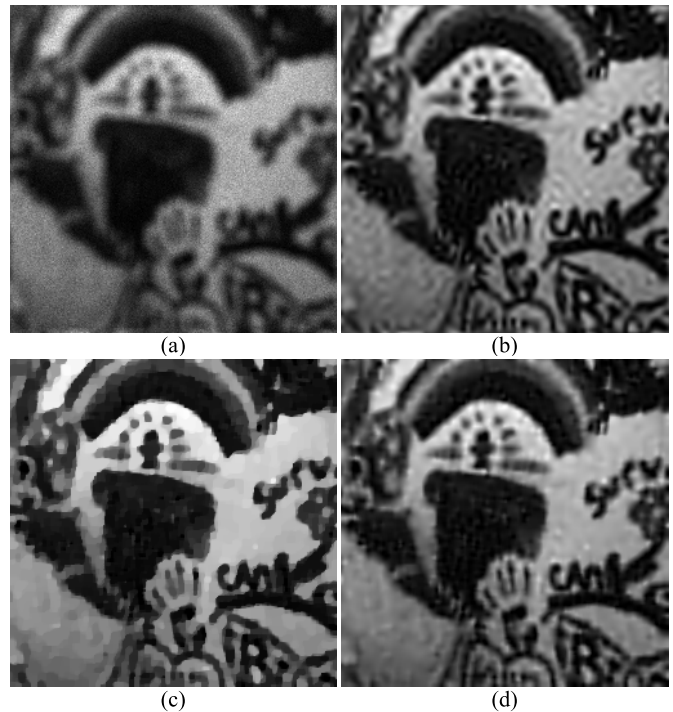
Fig. 6. Restoration of the Wall image, scaled to a maximum intensity of $M = 255$ and degraded by uniform blurring and Poisson noise. (a) Degraded image (PSNR = 19.59 dB), (b) DB$_4$ result (PSNR = 23.92 dB), (c) TV result (PSNR = 23.93 dB), and (d) $\mathcal{HS}_\infty$ result (PSNR = 24.26 dB).

independent realizations of Poisson noise. For the sake of consistency, the reported results for each regularizer under comparison, including TV, Haar, and DB$_4$, are derived using

TABLE I

PSNR COMPARISONS (POOLED OVER 10 TRIALS) ON POISSON IMAGE RESTORATION FOR THREE BLURRING KERNELS AND FOUR NOISE LEVELS

| Image / Peak		Gaussian blur: 9×9						Motion blur: 19×19						Uniform blur: 9×9					
		Haar	DB$_4$	TV	$\mathcal{HS}_\infty$	$\mathcal{HS}_2$	$\mathcal{HS}_1$	Haar	DB$_4$	TV	$\mathcal{HS}_\infty$	$\mathcal{HS}_2$	$\mathcal{HS}_1$	Haar	DB$_4$	TV	$\mathcal{HS}_\infty$	$\mathcal{HS}_2$	$\mathcal{HS}_1$
Boat	5	22.30	21.83	22.44	22.42	22.45	**22.46**	22.32	21.80	22.41	22.42	22.45	**22.45**	22.18	21.73	22.33	22.32	22.35	**22.35**
	25	23.43	23.38	23.62	23.69	23.73	**23.74**	23.70	23.36	23.79	23.77	23.82	**23.83**	23.30	23.17	23.48	23.48	23.52	**23.53**
	100	24.32	24.36	24.55	24.64	24.68	**24.69**	24.99	24.40	25.03	24.98	25.05	**25.09**	24.28	24.11	24.47	24.49	24.55	**24.58**
	255	24.94	24.91	25.18	25.23	25.29	**25.32**	26.04	25.22	26.05	25.98	26.09	**26.16**	25.06	24.80	25.25	25.27	25.35	**25.39**
Face	5	20.63	21.40	21.38	21.84	21.93	**21.96**	19.94	20.72	20.30	21.05	21.11	**21.12**	20.38	21.18	21.08	21.58	21.66	**21.69**
	25	22.26	23.21	23.17	23.72	23.81	**23.82**	21.77	22.52	22.14	22.91	22.96	**22.97**	21.90	22.87	22.75	23.36	23.44	**23.47**
	100	23.44	24.59	24.54	25.14	25.25	**25.27**	23.48	24.29	23.81	24.61	24.67	**24.70**	23.13	24.11	23.99	24.62	24.72	**24.75**
	255	24.17	25.43	25.32	25.95	26.07	**26.09**	24.85	25.93	25.13	26.10	26.20	**26.24**	23.98	25.00	24.79	25.44	25.55	**25.59**
Hill	5	23.28	22.61	23.54	23.62	23.66	**23.67**	23.18	22.48	23.42	23.57	23.60	**23.61**	23.19	22.52	23.43	23.53	23.57	**23.57**
	25	24.59	24.25	24.84	24.91	24.95	**24.95**	24.61	24.06	24.76	24.89	24.92	**24.93**	24.44	24.10	24.68	24.75	24.78	**24.78**
	100	25.44	25.16	25.69	25.78	25.81	**25.81**	25.75	24.98	25.83	25.89	25.94	**25.96**	25.38	24.94	25.59	25.61	25.65	**25.66**
	255	26.00	25.67	26.23	26.30	26.34	**26.35**	26.61	25.64	26.63	26.70	26.79	**26.83**	26.03	25.48	26.20	26.23	26.28	**26.31**
House	5	20.15	20.33	20.43	20.59	**20.61**	20.60	19.97	20.17	20.15	20.41	**20.42**	20.41	19.96	20.11	20.23	20.39	**20.41**	20.40
	25	21.44	21.77	21.78	22.04	**22.08**	22.07	21.50	21.44	21.59	21.88	21.91	**21.91**	21.24	21.45	21.55	21.77	21.81	**21.81**
	100	22.48	22.80	22.82	23.13	23.18	**23.18**	23.13	22.97	23.14	23.58	23.65	**23.68**	22.56	22.59	22.80	23.03	23.09	**23.10**
	255	23.21	23.49	23.51	23.86	23.92	**23.92**	24.58	24.60	24.57	25.18	25.29	**25.32**	23.48	23.47	23.68	23.97	24.05	**24.06**
Kids	5	21.37	21.57	22.01	22.01	**22.06**	22.05	21.09	21.30	21.54	21.78	**21.82**	21.81	21.14	21.35	21.74	21.73	**21.78**	21.77
	25	23.10	23.29	23.83	23.83	**23.89**	23.88	23.13	22.97	23.58	23.64	23.69	**23.71**	22.84	22.97	23.53	23.54	**23.60**	23.60
	100	24.36	24.56	25.11	25.24	25.31	**25.32**	24.93	24.35	25.21	25.21	25.31	**25.36**	24.25	24.22	24.87	24.92	25.00	**25.01**
	255	25.18	25.35	25.93	26.09	26.18	**26.19**	26.20	25.65	26.34	26.54	26.67	**26.74**	25.24	25.09	25.79	25.84	25.95	**25.99**
Lena	5	23.83	23.23	**24.38**	24.24	24.32	24.36	24.06	23.31	24.56	24.44	24.53	**24.59**	23.63	23.10	24.18	24.09	24.16	**24.20**
	25	25.36	25.36	25.99	25.91	25.99	**26.02**	25.83	25.53	26.35	26.24	26.33	**2.38**	25.11	25.13	25.75	25.68	25.75	**25.79**
	100	26.42	26.64	27.12	27.14	27.21	**27.24**	27.24	26.85	27.66	27.62	27.72	**27.78**	26.23	26.31	26.86	26.86	26.94	**26.98**
	255	27.07	27.32	27.76	27.85	27.92	**27.96**	28.24	27.67	28.57	28.57	28.70	**28.78**	27.01	27.02	27.60	27.64	27.73	**27.77**
Peppers	5	20.48	20.41	20.70	20.90	20.94	**20.95**	20.63	20.42	20.76	21.04	21.08	**21.08**	20.31	20.26	20.52	20.71	20.75	**20.76**
	25	21.66	21.70	21.94	22.29	22.33	**22.34**	22.12	21.75	22.28	22.80	22.88	**22.92**	21.50	21.50	21.76	22.20	22.26	**22.28**
	100	22.50	22.52	22.87	23.35	23.41	**23.44**	23.47	22.69	24.10	24.52	24.63	**24.71**	22.48	22.31	22.95	23.46	23.55	**23.61**
	255	23.07	22.97	23.63	24.18	24.26	**24.32**	24.75	23.58	25.62	25.80	25.97	**26.08**	23.28	22.89	24.12	24.41	24.54	**24.62**
Wall	5	19.52	20.00	20.09	20.25	20.31	**20.31**	19.44	19.88	19.75	20.14	20.20	**20.21**	19.24	19.69	19.75	19.90	**19.95**	19.94
	25	21.18	22.01	21.98	22.25	22.34	**22.34**	21.39	21.91	21.72	22.23	22.31	**22.34**	20.72	21.36	21.40	21.62	21.69	**21.70**
	100	22.49	23.48	23.46	23.79	23.88	**23.89**	23.32	23.70	23.60	24.11	24.22	**24.26**	22.17	22.95	22.85	23.21	23.32	**23.34**
	255	23.40	24.38	24.44	24.76	24.87	**24.88**	24.97	25.20	25.19	25.64	25.77	**25.82**	23.32	24.09	23.99	24.37	24.50	**24.54**

the individualized regularization parameter τ that gives the best PSNR performance while, similar to [13], the penalty parameter in the ADMM updates is chosen as $\alpha = 60\tau/M$. According to the authors in [13], this heuristic choice of the ADMM penalty parameter appears to lead to satisfactory results for wavelet and TV regularization. Moreover, choosing a value which is an order above or below this choice does not change significantly the convergence behavior of the algorithm.

From Table I, we observe that the Hessian-based regularization framework almost always results in the best restoration performance and leads to improved quantitative results compared to TV and to wavelet-based regularization. The $\mathcal{HS}_1$ regularizer achieves slightly better pooled scores than $\mathcal{HS}_2$. The $\mathcal{HS}_\infty$ regularizer, while not performing as well as the other two members of the family, in most of the cases still performs better than the other alternative approaches. The TV

regularizer consistently outperforms the Haar frame operator, which sometimes falls behind more than 1 dB compared to the best achieved PSNR result. However, for certain images the DB$_4$ wavelet regularization outperforms TV.

Beyond quantitative comparisons, the improved performance of the Hessian-based regularization framework can also be visually appreciated by inspecting the representative Poisson image restoration examples provided in Figs. 4–6. Even in cases where the PSNR improvement over the competing regularizers is not substantial, the $\mathcal{HS}_p$ penalty functions result in image estimates with finer feature reconstructions and avoid the presence of strong artifacts. On the other hand, as it is verified from Figs. 4–6, block artifacts are introduced by the Haar and TV regularizers, while ringing-artifacts are introduced by the DB$_4$ regularizer. These are becoming even more pronounced when the level of Poisson noise is increased.

TABLE II

PSNR COMPARISONS (POOLED OVER 10 TRIALS) ON POISSON IMAGE RESTORATION FOR TWO BLURRING KERNELS AND FOUR NOISE LEVELS

|	Image	CAM				CAM2				Fluor. Cells				Leukocyte				Thrombus			
	Peak	10	50	250	500	10	50	250	500	10	50	250	500	10	50	250	500	10	50	250	500
Gaussian blur: 7×7	Haar	18.14	19.11	20.04	20.49	19.98	21.28	22.44	22.98	26.66	28.34	29.83	30.50	20.71	22.06	23.33	23.91	20.25	21.50	22.70	23.28
	DB$_4$	18.39	19.49	20.43	20.85	20.42	21.89	23.10	23.62	27.28	29.23	30.86	31.50	21.10	22.75	24.15	24.72	20.68	22.09	23.30	23.87
	TV	18.39	19.40	20.33	20.77	20.39	21.73	22.87	23.36	27.05	28.80	30.28	30.89	21.06	22.56	23.90	24.50	20.20	21.92	23.07	23.61
	$\mathcal{HS}_\infty$	18.53	19.61	20.58	21.01	20.67	22.12	23.35	23.88	27.64	29.56	31.15	31.77	21.21	22.81	24.21	24.81	20.86	22.26	23.52	24.09
	$\mathcal{HS}_2$	18.56	19.65	20.63	21.07	20.71	**22.16**	**23.40**	**23.94**	27.70	29.61	31.21	**31.84**	21.27	22.89	24.31	24.92	20.89	**22.31**	**23.58**	**24.16**
	$\mathcal{HS}_1$	**18.57**	**19.66**	**20.64**	**21.08**	**20.72**	22.16	23.39	23.93	**27.70**	**29.62**	**31.21**	31.84	**21.29**	**22.91**	**24.33**	**24.95**	**20.90**	22.30	23.57	24.15
Airy-disk blur: 31×31	Haar	18.21	19.37	20.37	20.75	20.02	21.55	22.88	23.38	26.60	28.45	30.04	30.54	20.74	22.39	23.92	24.48	20.29	21.82	23.17	23.67
	DB$_4$	18.36	19.73	20.91	21.33	20.37	22.13	23.69	24.26	27.08	29.29	31.23	31.97	21.04	23.03	24.84	25.52	20.63	22.36	23.98	24.58
	TV	18.45	19.69	20.75	21.14	20.41	22.03	23.39	23.91	26.97	28.92	30.57	31.21	21.07	22.92	24.63	25.26	20.64	22.27	23.67	24.22
	$\mathcal{HS}_\infty$	18.55	19.87	21.01	21.43	20.67	22.40	23.93	24.49	27.48	29.63	31.51	32.25	21.18	23.12	24.90	25.56	20.85	22.58	24.16	24.78
	$\mathcal{HS}_2$	18.59	19.92	21.06	**21.48**	20.72	**22.44**	**23.98**	**24.55**	27.53	29.68	**31.57**	**32.32**	21.24	23.21	25.00	25.66	20.89	**22.63**	**24.22**	**24.84**
	$\mathcal{HS}_1$	**18.60**	19.92	**21.07**	21.47	**20.72**	22.44	23.96	24.52	**27.54**	**29.69**	31.57	32.31	**21.26**	**23.23**	**25.01**	**25.67**	**20.90**	22.62	24.21	24.82
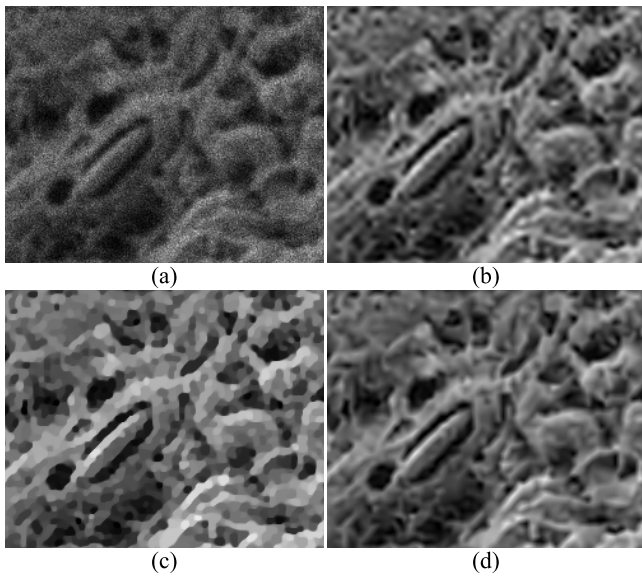
Fig. 7. Restoration of the CAM2 image, scaled to a maximum intensity of $M = 50$ and degraded by airy-disk blurring and Poisson noise. Close up of (a) Degraded image (PSNR = 18.29 dB), (b) DB$_4$ result (PSNR = 22.13 dB), (c) TV result (PSNR = 22.03 dB), and (d) $\mathcal{HS}_1$ result (PSNR = 22.44 dB).
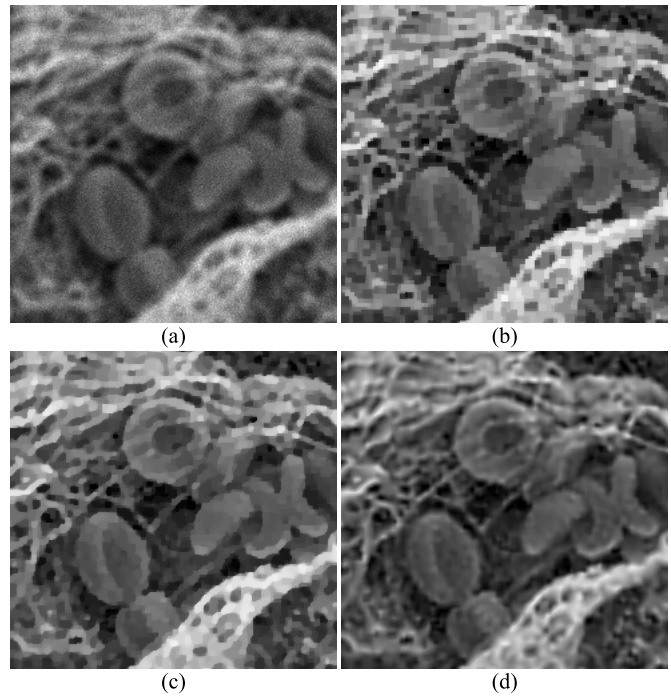

Fig. 8. Restoration of the Thrombus image, scaled to a maximum intensity of $M = 250$ and degraded by Gaussian blurring and Poisson noise. Close up of (a) Degraded image (PSNR = 20.88 dB), (b) Haar result (PSNR = 22.70 dB), (c) TV result (PSNR = 23.07 dB), and (d) $\mathcal{HS}_2$ result (PSNR = 23.55 dB).

Regarding the computational cost, we report the execution time of all methods under comparison for the restoration of the Face image under Gaussian blur and at a peak intensity 25. For each regularization method we used the regularization parameter that led to the best PSNR performance. The simulations were performed on a laptop with a 2.5 GHz Intel Core i7 Processor and 8 GB memory, using a MATLAB implementation of the algorithms. The reconstruction of the image using the $\mathcal{HS}_p$ regularizers for $p = 1, 2, \infty$, took 29.2, 19.8, and 29.3 seconds, respectively. The TV reconstruction was obtained in 20.71 seconds, while the Haar and DB$_4$ reconstructions in 57.7 and 76.7 seconds, respectively. Note that while each of the HSPIRAL1 iterations is more expensive than TV's, the termination criteria are satisfied earlier, which explains why in the $\mathcal{HS}_2$ case the running time is smaller.

D. Poisson Deblurring of Biological Images

A potential and quite promising application of the photon-imaging observation model is biomicroscopy. There, the most dominant noise source is the shot noise, which obeys a Poisson law. An extra degradation factor is the out-of-focus blur due to the poor localization of the PSF of the microscope. Since the use of the proposed regularization framework seems to have a practical relevance in this field, we provide in Table II comparative restoration results for the set of biological images shown in Fig. 2. As with natural images we report the results obtained with the same $\mathcal{HS}_p$ regularizers and we compare

their performance with wavelet-based and TV regularization. Likewise, we observe once more that the Hessian-based regularizers, combined with the proposed minimization approach, consistently outperform the alternative regularization schemes. Moreover, since the biological images of Fig. 2 consist of smooth intensity transitions and filament-like features, the $\mathcal{HS}_p$ regularizers are well-tuned to these image properties and result in satisfactory reconstructions. Representative restoration examples of biological images are illustrated in Figs. 7 and 8. There, we observe that the Haar and TV penalties have the effect of shuffling the details of the images and broadening their fine structures, while the wavelet regularization with higher vanishing moments (DB_4) introduces ringing artifacts. On the other hand, the Hessian-based regularizers show a better reconstruction behavior and restore the image features more accurately.

V. Conclusion

We proposed an efficient framework for Poisson image reconstruction subject to regularization which depends upon potential functions acting on the eigenvalues of the Hessian. Our motivation for employing this class of penalty terms stems from the fact that they have recently been shown to perform well for linear inverse problems with Gaussian noise. We observe that their performance remains consistent in the case of Poisson noise. We designed two algorithms that optimize the corresponding objective functions. They are based on an augmented Lagrangian formulation and arise as two different ADMM variants. The second variant depends heavily on a result we derive in this paper. This result relates the evaluation of the proximal map of a Schatten matrix-norm of order p with the evaluation of the proximal map of the ℓ_p norm. The scope of this link, while being fundamental to the development of our second proposed algorithm, is general and potentially has a wider applicability that extends beyond this paper.

The practical relevance of the proposed regularization framework, as well as the effectiveness of our novel minimization approach, was verified through comparisons with alternative methods, including total variation, for the problem of Poisson image deblurring of natural and biological images. The results we obtained are promising and competitive. They provide an indication that this framework can be considered as a viable alternative to other existing schemes.

References

[1] H. Barett, "Objective assessment of image quality: Effects of quantum noise and object variability," *J. Opt. Soc. Amer. A*, vol. 7, no. 7, pp. 1266–1278, 1990.

[2] R. Nowak and E. Kolaczyk, "A statistical multiscale framework for Poisson inverse problems," *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 1811–1825, Aug. 2000.

[3] A. Antoniadis and J. Bigot, "Poisson inverse problems," *Ann. Stat.*, vol. 34, no. 5, pp. 2132–2158, 2006.

[4] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. 1, no. 2, pp. 113–122, Oct. 1982.

[5] P. Sarder and A. Nehorai, "Deconvolution methods for 3-D fluorescence microscopy images," *IEEE Signal Process. Mag.*, vol. 23, no. 3, pp. 32–45, May 2006.

[6] D. L. Snyder and A. M. Hammoud, "Image recovery from data acquired with a charge-coupled-device camera," *J. Opt. Soc. Amer. A*, vol. 10, no. 5, pp. 1014–1023, 1993.

[7] W. H. Richardson, "Bayesian-based iterative method of image restoration," *J. Opt. Soc. Amer. A*, vol. 62, no. 1, pp. 55–59, 1972.

[8] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *The Astronomical J.*, vol. 79, no. 6, p. 745, 1974.

[9] F. Dupe, J. Fadili, and J. Starck, "A proximal iteration for deconvolving Poisson noisy images using sparse representations," *IEEE Trans. Image Process.*, vol. 18, no. 2, pp. 310–321, Feb. 2009.

[10] F. J. Anscombe, "The transformation of Poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, no. 3, pp. 246–254, 1948.

[11] S. Lefkimmiatis, P. Maragos, and G. Papandreou, "Bayesian inference on multiscale models for Poisson intensity estimation: Applications to photon-limited image denoising," *IEEE Trans. Image Process.*, vol. 18, no. 8, pp. 1724–1741, Aug. 2009.

[12] N. Dey, L. Blanc-Feraud, C. Zimmer, P. Roux, Z. Kam, J. Olivo-Marin, and J. Zerubia, "Richardson–Lucy algorithm with total variation regularization for 3D confocal microscope deconvolution," *Microscopy Res. Tech.*, vol. 69, no. 4, pp. 260–266, 2006.

[13] M. Figueiredo and J. Bioucas-Dias, "Restoration of Poissonian images using alternating direction optimization," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3133–3145, Dec. 2010.

[14] S. Setzer, G. Steidl, and T. Teuber, "Deblurring Poissonian images by split Bregman techniques," *J. Vis. Commun. Image Represent.*, vol. 21, no. 3, pp. 193–199, 2010.

[15] Z. T. Harmany, R. F. Marcia, and R. M. Willet, "This is SPIRAL-TAP: Sparse Poisson intensity reconstruction algorithms—Theory and practice," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1084–1096, Mar. 2012.

[16] S. Lefkimmiatis, J. Ward, and M. Unser, "Hessian Schatten-norm regularization for linear inverse problems," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1873–1888, May 2013.

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

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

[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] 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.

[21] E. Esser, "Applications of Lagrangian-based alternating direction methods and connections to split Bregman," *CAM Rep.*, vol. 9, p. 31, Mar. 2009.

[22] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*. Norwell, MA, USA: Now Publishers, 2011.

[23] 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.

[24] J. Eckstein and D. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 1, pp. 293–318, 1992.

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

[26] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, Sep. 2010.

[27] 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

[28] M. Nilchian, C. Vonesch, P. Modregger, M. Stampanoni, and M. Unser, "Fast iterative reconstruction of differential phase contrast X-ray tomograms," *Opt. Exp.*, vol. 21, no. 5, pp. 5511–5528, 2013.

[29] L. Mirsky, "A trace inequality of John von Neumann," *Monatshefte fur Mathematik*, vol. 79, no. 4, pp. 303–306, 1975.

[30] J. Liu and J. Ye, "Efficient ℓ_1/ℓ_q norm regularization," arXiv:1009.4766, Sep. 2010.

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

[32] 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.

[33] 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.

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

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

[36] 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.

[37] E. Van De Berg, M. Friedlander, and K. Murphy. (2008). *Group Sparsity Via Linear-Time Projection* [Online]. Available: http://www.cs.ubc.ca/~ewout78/papers/TR-2008-09.pdf

[38] 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.

[39] I. Selesnick and M. Figueiredo, "Signal restoration with overcomplete wavelet transforms: Comparison of analysis and synthesis priors," *Proc. SPIE*, vol. 7446, p. 74460D, Sep. 2009.

[40] D. Brandner and G. Withers. (2010). *The Cell: An Image Library* [Online]. Available: http://www.cellimagelibrary.org

[41] B. Richards and E. Wolf, "Electromagnetic diffraction in optical systems. II. Structure of the image field in an aplanatic system," *Proc. R. Soc. London A*, vol. 253, no. 1274, pp. 358–379, 1959.

[42] 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.

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

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. Since 2010, he has been a Research Scientist with the Biomedical Imaging Group, Swiss Federal Institute of Technology, Lausanne, Switzerland. His current research interests include image analysis, statistical modeling, and inverse problems in imaging with focus on biomedical and computer vision applications.

Michael Unser (M'89–SM'94–F'99) is a Professor and the Director of EPFL's Biomedical Imaging Group, Lausanne, Switzerland. His main research area is biomedical image processing. He has a strong interest in sampling theories, multiresolution algorithms, wavelets, the use of splines for image processing, and, more recently, stochastic processes. He has published about 250 journal papers on those topics. From 1985 to 1997, he was with the Biomedical Engineering and Instrumentation Program, National Institutes of Health, Bethesda, MD, USA, conducting research on bioimaging and heading the Image Processing Group. He is an EURASIP fellow (2009), and a member of the Swiss Academy of Engineering Sciences. He is a recipient of several international prizes, including three IEEE-SPS Best Paper Awards and two Technical Achievement Awards from the IEEE SPS in 2008 and EMBS in 2010.