Algorithms for Imaging Inverse Problems Under Sparsity and Structured Sparsity Regularization.

Mário A. T. Figueiredo and José M. Bioucas-Dias

Instituto de Telecomunicações, and
Instituto Superior Técnico, Technical University of Lisbon, **Portugal**

Email: mario.figueiredo@lx.it.pt, bioucas@lx.it.pt

***Abstract*—This paper reviews our recent work on the application of a class of techniques known as ADMM (alternating direction method of multipliers, which belongs to the family of augmented Lagrangian methods) to several imaging inverse problems under sparsity-inducing regularization. After reviewing ADMM, a formulation that allows handling problems with more than two terms is introduced; this formulation is then applied to a variety of problems, namely: standard image restoration/reconstruction from linear observations (e.g., compressive sensing, deconvolution, inpainting) with Gaussian or Poisson noise, using either analysis or synthesis regularization, and unconstrained or constrained optimization. We also show how the proposed framework can be used to address hybrid analysis-synthesis regularization. In all these cases, the proposed approach inherits the theoretic convergence guarantees of ADMM and achieve state-of-the-art speed.**

I. Introduction

Imaging inverse problems abound in the modern world. Medical imaging (CT, MRI, PET, ultrasound), remote sensing, seismography, non-destructive inspection, digital photography/video, astronomy, all involve at their computational core the solution of imaging inverse problems: they produce visual representations (images) of an underlying reality from indirect/imperfect observations. Inverse problems are typically ill-posed, this meaning that even if the observation model/operator is perfectly known, the observations do not uniquely and stably determine the solution. This difficulty is typically dealt with by seeking some sort of balance between data fidelity and adherence to a set of properties that the unknown image is known (or desired) to have. The classical way to achieve such a balance consists in formulating an optimization problem (usually convex) where the objective function includes a term encouraging the estimates to explain the observed data and another term (the regularizer or prior) penalizing solutions considered undesirable. Current state of the art regularizers encourage or enforce sparseness of the representation of the underlying image with respect to some redundant frame or dictionary, a feature known to characterize natural noiseless images (see [15] and the many references therein). This sparseness may be expressed via the analysis or synthesis formulations [16], [28], usually via the standard ℓ_1 norm or, more recently and with better performance, by taking into account dependency structures among the representation coefficients via group norms [26].

A significant fraction of the recent progress in imaging inverse problems revolved around developing fast algorithms to address the convex optimization problems referred to in the previous paragraph. Of course, the literature on this topic is too large to be comprehensively covered in this short review; instead, this paper will focus on some current state of the art algorithms for a variety of imaging inverse problems and formulations that we have recently developed [1]–[3], [6], [18], [19]. All those algorithms will be presented as instances of a common optimization framework, based on the *alternating direction method of multipliers* (ADMM) [14], [22], [23].

II. General Problem Formulation

Let $\Psi(\mathbf{y},\mathbf{x})$ be a function that measures how much a given candidate estimate $\mathbf{x}$ deviates from explaining the data $\mathbf{y}$. This function is typically derived from a model of how the observations are generated; *e.g.*, in a probabilistic formulation, this would be the negative log-likelihood ($\Psi(\mathbf{y},\mathbf{x}) = -\log p(\mathbf{y}|\mathbf{x})$), but other semantics are possible. Consider also the *regularization* function $\Phi(\mathbf{x})$ that measures how undesirable a candidate estimate $\mathbf{x}$ is; *e.g.*, in a Bayesian approach, $\Phi(\mathbf{x}) = -\log p(\mathbf{x})$ is the negative log-prior, but other formulations exist.

There are three standard ways to combine Ψ and Φ into an optimization problem[1], the solution of which strikes a balance between the two desiderata expressed by these functions:

- Tikhonov regularization: $\min_{\mathbf{x}} \Psi(\mathbf{y},\mathbf{x}) + \alpha\Phi(\mathbf{x})$;
- Morozov regularization: $\min_{\mathbf{x}} \Phi(\mathbf{x})$ s. t. $\Psi(\mathbf{y},\mathbf{x}) \leq \delta$;
- Ivanov regularization: $\min_{\mathbf{x}} \Psi(\mathbf{y},\mathbf{x})$ s. t. $\Phi(\mathbf{x}) \leq \tau$.

If both $\Psi(\mathbf{y},\mathbf{x})$ and $\Phi(\mathbf{x})$ are convex functions of $\mathbf{x}$, these formulations are equivalent, in principle, in the following sense: under some mild conditions, for any choice of the parameter defining one of the problems (α, δ, or τ), there is a choice of the other two parameters for which all the problems have a common solution [32]. However, in practice it is necessary to choose/adjust these parameters, which is sometimes more conveniently done in one formulations than the others. In this paper, we will consider only Tikhonov

[1]We are following the designations recently proposed by D. Lorenz; see http://regularize.wordpress.com/2011/05/04/ivanov-regularization/

and Morozov regularization, since most of the derivations for Morozov regularization apply with minor changes to the Ivanov counterpart.

In this paper, we will focus only on frame-based regularization, which has the following rationale: the frame[2] coefficients of natural noise-free images are *sparse*. A detailed discussion of what sparseness exactly means and how it applies to images is beyond the scope of this paper (see [15], for details and pointers to a vast literature); we simply use the classical ℓ_1 norm ($\|\mathbf{v}\|_1 = \sum_i |v_i|$) of the frame coefficients as a measure of (non)-sparseness. There are essentially two formulations of frame-based ℓ_1 regularization [16], [28]:

- In the *analysis formulation*, $\mathbf{x} \in \mathbb{R}^n$ represents[3] the image itself, and the regularizer is applied to its frame analysis coefficients, thus it has the form $\Phi(\mathbf{x}) = \|\mathbf{W}^T\mathbf{x}\|_1$. The data term has the form $\Psi(\mathbf{y},\mathbf{x}) = \Upsilon(\mathbf{y},\mathbf{x})$, where $\Upsilon(\mathbf{y},\mathbf{x})$ is a function that measures the degree of discrepancy between **image** $\mathbf{x}$ and the data $\mathbf{y}$.
- In the synthesis formulation, rather than the image itself, $\mathbf{x} \in \mathbb{R}^k$ denotes the vector of coefficients of its frame-based representation $\mathbf{Wx}$. The regularizer is thus $\Phi(\mathbf{x}) = \|\mathbf{x}\|_1$ and the data term has the form $\Psi(\mathbf{y},\mathbf{x}) = \Upsilon(\mathbf{y},\mathbf{Wx})$.

If $\mathbf{W}$ is an orthonormal basis, the synthesis and analysis formulations are equivalent; however, for redundant frames, the two formulations yield different results [16]. Hybrid analysis-synthesis formulations are also possible, as proposed in [19] (see Section VI).

III. The Alternating Direction Method of Multipliers (ADMM)

A. The Standard ADMM

Consider an unconstrained problem of the form

$$\min_{\mathbf{z}\in\mathbb{R}^d} \; f_1(\mathbf{z}) + f_2(\mathbf{G}\,\mathbf{z}), \tag{1}$$

where $f_1 : \mathbb{R}^d \to \bar{\mathbb{R}}$, $f_2 : \mathbb{R}^p \to \bar{\mathbb{R}}$, and $\mathbf{G} \in \mathbb{R}^{p\times d}$. The ADMM for this problem is defined as follows:

Algorithm *ADMM*
1. Set $k = 0$, choose $\mu > 0$, $\mathbf{u}_0$, and $\mathbf{d}_0$.
2. **repeat**
3. $\quad\mathbf{z}_{k+1} \in \arg\min_{\mathbf{z}} f_1(\mathbf{z}) + \frac{\mu}{2}\|\mathbf{G}\,\mathbf{z} - \mathbf{u}_k - \mathbf{d}_k\|_2^2$
4. $\quad\mathbf{u}_{k+1} \in \arg\min_{\mathbf{u}} f_2(\mathbf{u}) + \frac{\mu}{2}\|\mathbf{G}\,\mathbf{z}_{k+1} - \mathbf{u} - \mathbf{d}_k\|_2^2$
5. $\quad\mathbf{d}_{k+1} \leftarrow \mathbf{d}_k - (\mathbf{G}\,\mathbf{z}_{k+1} - \mathbf{u}_{k+1})$
6. $\quad k \leftarrow k+1$
7. **until** stopping criterion is satisfied.

[2]In a vector space, say $\mathbb{R}^n$, a frame is a collection of vectors $\{\mathbf{w}_1, ..., \mathbf{w}_k\}$ satisfying $A\|\mathbf{x}\|^2 \le \sum_j |\langle \mathbf{x}, \mathbf{w}_j\rangle|^2 \le B\|\mathbf{x}\|^2$, for some $0 < A \le B < \infty$. If $A = B$, the frame is called tight, and if $A = B = 1$ it's called a Parseval frame (we will only use Parseval frames). Collecting the frame vectors into a matrix $\mathbf{W} \in \mathbb{R}^{n\times k}$, we have $\mathbf{W}\mathbf{W}^T = \mathbf{I}$, with $\mathbf{W}^T\mathbf{W} = \mathbf{I}$ also holding only if the frame is an orthonormal basis (of course with $k = n$). If $n \ge k$, the frame is called redundant. For a Parseval frame, $\mathbf{W}$ is called the synthesis matrix and $\mathbf{W}^T \in \mathbb{R}^{k\times n}$ is the analysis matrix.

[3]As is commonly done, $\mathbf{x}$ is the vector representation of an image, obtained by a stacking its pixels in lexicographical order.

Convergence of (a generalized version of) ADMM was shown by Eckstein and Bertsekas (in [14]):

Theorem 1: Consider problem (1), where $\mathbf{G} \in \mathbb{R}^{p\times d}$ has full column rank and $f_1 : \mathbb{R}^d \to \bar{\mathbb{R}}$ and $f_2 : \mathbb{R}^p \to \bar{\mathbb{R}}$ are closed, proper, and convex. Consider arbitrary $\mu > 0$, $\mathbf{u}_0, \mathbf{d}_0 \in \mathbb{R}^p$. Let $\eta_k \ge 0,\; k = 0,1,...,$ and $\rho_k \ge 0,\; k = 0,1,...,$ be two sequences such that $\sum_{k=0}^{\infty} \eta_k < \infty$ and $\sum_{k=0}^{\infty} \rho_k < \infty$. Consider three sequences $\mathbf{z}_k \in \mathbb{R}^d$, $\mathbf{u}_k \in \mathbb{R}^p$, and $\mathbf{d}_k \in \mathbb{R}^p$, for $k = 0,1,...,$ satisfying

$$\begin{aligned}
\left\|\mathbf{z}_{k+1} - \arg\min_{\mathbf{z}} f_1(\mathbf{z}) + \frac{\mu}{2}\|\mathbf{G}\mathbf{z} - \mathbf{u}_k - \mathbf{d}_k\|_2^2\right\| &\le \eta_k \\
\left\|\mathbf{u}_{k+1} - \arg\min_{\mathbf{u}} f_2(\mathbf{u}) + \frac{\mu}{2}\|\mathbf{G}\mathbf{z}_{k+1} - \mathbf{u} - \mathbf{d}_k\|_2^2\right\| &\le \rho_k \\
\mathbf{d}_{k+1} = \mathbf{d}_k - (\mathbf{G}\,\mathbf{z}_{k+1} - \mathbf{u}_{k+1}).&
\end{aligned}$$

Then, if (1) has a solution, say $\mathbf{z}^*$, the sequence $\{\mathbf{z}_k\}$ converges to $\mathbf{z}^*$. If (1) does not have a solution, then at least one of the sequences $\{\mathbf{u}_k\}$ or $\{\mathbf{d}_k\}$ diverges.

Notice that the sequences $\mathbf{z}_k$, $\mathbf{u}_k$ and $\mathbf{d}_k$ defined in the ADMM algorithm satisfy the conditions in the theorem with $\eta_k = \rho_k = 0$. However, the theorem shows that even if the minimizations in lines 3–4 of ADMM are inexactly solved, convergence still holds if the error sequences are absolutely summable. This fact is quite relevant in designing instances of ADMM, when these minimizations lack closed form solutions. For recent and comprehensive reviews of ADMM and its relationship with Bregman methods [33], see [7], [17].

B. ADMM for More than Two Functions

Consider a generalization of (1): instead of two functions, we have J (closed, proper, and convex) functions, *i.e.*,

$$\min_{\mathbf{z}\in\mathbb{R}^d} \sum_{j=1}^{J} g_j(\mathbf{H}^{(j)}\,\mathbf{z}), \tag{2}$$

where $g_j : \mathbb{R}^{p_j} \to \bar{\mathbb{R}}$, and $\mathbf{H}^{(j)} \in \mathbb{R}^{p_j\times d}$ are arbitrary matrices. The minimization problem (2) can be written as (1) using the following correspondences: $f_1 = 0$,

$$\mathbf{G} = \left[(\mathbf{H}^{(1)})^T \;\cdots\; (\mathbf{H}^{(J)})^T\right]^T \in \mathbb{R}^{p\times d}, \tag{3}$$

where $p = p_1 + \cdots + p_J$, and $f_2 : \mathbb{R}^{p\times d} \to \bar{\mathbb{R}}$ is given by

$$f_2(\mathbf{u}) = \sum_{j=1}^{J} g_j(\mathbf{u}^{(j)}), \tag{4}$$

where $\mathbf{u}^{(j)} \in \mathbb{R}^{p_j}$ and $\mathbf{u} = [(\mathbf{u}^{(1)})^T, \ldots, (\mathbf{u}^{(J)})^T]^T \in \mathbb{R}^p$. In applying ADMM to the resulting problem, it is convenient to define the following partitions (where $\mathbf{d}_k^{(j)}$, $\mathbf{u}_k^{(j)} \in \mathbb{R}^{p_j}$):

$$\mathbf{d}_k = \left[(\mathbf{d}_k^{(1)})^T \cdots (\mathbf{d}_k^{(J)})^T\right]^T, \;\; \mathbf{u}_k = \left[(\mathbf{u}_k^{(1)})^T \cdots (\mathbf{u}_k^{(J)})^T\right].$$

The fact that $f_1 = 0$ turns step 3 of ADMM into a quadratic problem, which has a unique solution if $\mathbf{G}$ has full column rank. Given the block structure of $\mathbf{G}$ in (3), the corresponding solution is (with $\boldsymbol{\zeta}_k^{(j)} = \mathbf{u}_k^{(j)} + \mathbf{d}_k^{(j)}$):

$$\arg\min_{\mathbf{z}} \left\|\mathbf{G}\,\mathbf{z} - \boldsymbol{\zeta}_k\right\|_2^2 = \left[\sum_{j=1}^{J} (\mathbf{H}^{(j)})^T \mathbf{H}^{(j)}\right]^{-1} \sum_{j=1}^{J} \big(\mathbf{H}^{(j)}\big)^T \boldsymbol{\zeta}_k^{(j)}. \tag{5}$$

Furthermore, our particular way of mapping problem (2) into problem (1) allows decoupling the minimization in Step 4 of ADMM into a set of J independent ones:

$$\mathbf{u}_{k+1}^{(j)} \leftarrow \arg\min_{\mathbf{v}\in\mathbb{R}^{p_j}} \; g_j(\mathbf{v}) + \frac{\mu}{2}\left\|\mathbf{v} - \mathbf{s}_k^{(j)}\right\|_2^2, \tag{6}$$

for $j = 1, ..., J$, where $\mathbf{s}_k^{(j)} = \mathbf{H}^{(j)}\mathbf{z}_{k+1} - \mathbf{d}_k^{(j)}$.

The minimization problem in the right hand side of (6) defines the so-called *Moreau proximity operator* of g_j/μ (denoted as $\text{prox}_{g_j/\mu}$) [10], applied to $\mathbf{s}_k^{(j)}$, thus

$$\mathbf{u}_{k+1}^{(j)} \leftarrow \text{prox}_{g_j/\mu}(\mathbf{s}_k^{(j)}) \equiv \arg\min_{\mathbf{x}} \frac{\mu}{2}\left\|\mathbf{x} - \mathbf{s}_k^{(j)}\right\|_2^2 + g_j(\mathbf{x}). \tag{7}$$

For several functions, the corresponding Moreau proximity operators can be computed exactly in closed form [11]. A notable case (the only one used in this paper) is the ℓ_1 norm ($\|\mathbf{x}\|_1 = \sum_i |x_i|$), for which the corresponding proximity operator is the well-known soft threshold: $\text{prox}_{\|\cdot\|\gamma}(\mathbf{v}) = \text{soft}(\mathbf{v},\gamma) = \text{sign}(\mathbf{v}) \odot \max\{|\mathbf{v}| - \gamma, 0\}$, where $\text{sign}(\cdot)$ is the component-wise application of the sign function, $\odot$ is the component-wise product, $|\mathbf{v}|$ denotes the vector of absolute values of the elements of $\mathbf{v}$, and the maximum is computed in a component-wise fashion.

The computational bottleneck of this instance of ADMM is the matrix inversion in (5); below, we will see how this inversion can be very efficiently computed in a variety of problems of interest.

IV. Linear Observations with Gaussian Noise

A. Observation Model

Arguably, the most classical imaging inverse problem involves linear observations with additive white Gaussian noise; formally, the observed data $\mathbf{y}$ is modeled as

$$\mathbf{y} \sim \mathcal{N}(\mathbf{B}\mathbf{x}, \mathbf{I}) \tag{8}$$

where $\mathbf{B}$ is the matrix representation of the direct operator and $\mathcal{N}(\boldsymbol{\mu}, \mathbf{M})$ denotes a Gaussian distribution of mean vector $\boldsymbol{\mu}$ and covariance $\mathbf{M}$ (there is no loss of generality in assuming unit variance, since we assume it is known). In the case of image deconvolution, under periodic boundary conditions, $\mathbf{B}$ is a block-circulant matrix. Matrix $\mathbf{B}$ can also represent other linear operators, such as tomographic (Radon) projections or the loss of image pixels (inpainting). Given (8), the natural choice for the data term is the negative log-likelihood:

$$\Upsilon(\mathbf{y},\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{B}\mathbf{x}\|_2^2. \tag{9}$$

B. Tikhonov Analysis Regularization

The analysis formulation of Tikhonov regularization yields the following unconstrained optimization problem:

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{B}\mathbf{x}\|_2^2 + \alpha\|\mathbf{W}^T\mathbf{x}\|_1. \tag{10}$$

Problem (10) has the canonical form (2), with $J = 2$, $g_1(\mathbf{u}) = \frac{1}{2}\|\mathbf{y} - \mathbf{u}\|_2^2$, $g_2(\mathbf{u}) = \alpha\|\mathbf{u}\|_1$, $\mathbf{H}^{(1)} = \mathbf{B}$, and $\mathbf{H}^{(2)} = \mathbf{W}^T$. To implement the ADMM instance introduced in Subsection III-B, the necessary building blocks are the proximity operators of g_1 and g_2 in (7) and the matrix inversion appearing in (5). The proximity operators in this case are simple:

$$\begin{aligned} \text{prox}_{g_1/\mu}(\mathbf{s}) &= \arg\min_{\mathbf{x}} \frac{\mu}{2}\|\mathbf{s} - \mathbf{x}\|_2^2 + \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 \\ &= \frac{\mathbf{y} + \mu\mathbf{s}}{1+\mu}; \qquad (11) \\ \text{prox}_{g_2/\mu}(\mathbf{s}) &= \text{soft}(\mathbf{s}, 1/\mu). \qquad (12) \end{aligned}$$

In this case, the matrix inversion in (5) has the form

$$\left[\mathbf{B}^T\mathbf{B} + \mathbf{W}\mathbf{W}^T\right]^{-1} = \left[\mathbf{B}^T\mathbf{B} + \mathbf{I}\right]^{-1}, \tag{13}$$

because $\mathbf{W}$ is the synthesis matrix of a Parseval frame. The cost of computing this inverse depends critically on the structure of $\mathbf{B}$; in the following paragraphs, we will show how in a variety of problems of interest, this inversion can be computed with low cost. The algorithm also involves matrix-vector products involving $\mathbf{W}$ and $\mathbf{W}^T$, that is, frame synthesis and analysis operations; we only consider frames for which fast $O(n\log n)$ implementations of these operations exist [25]. Examples of such frames include undecimated wavelets, complex wavelets, curvelets, and shearlets. The resulting algorithm was proposed in [1] and was termed SALSA (split augmented Lagrangian shrinkage algorithm). Finally, notice that convergence of SALSA is guaranteed by Theorem 1, since matrix $\mathbf{G}$ (see (3)), which in this case is equal to $\mathbf{G} = [\mathbf{B}^T\ \mathbf{W}]^T$, has full column rank, as $\mathbf{W}^T$ is the analysis matrix of a Parseval frame.

1) Periodic Deconvolution: If $\mathbf{B}$ is a block-circulant matrix with circulant blocks, representing a periodic convolution, it can be factorized as

$$\mathbf{B} = \mathbf{U}^H\mathbf{D}\mathbf{U}, \tag{14}$$

where $\mathbf{U}$ is the matrix that represents the 2D discrete Fourier transform (DFT), $\mathbf{U}^H = \mathbf{U}^{-1}$ is its inverse[4] ($\mathbf{U}$ is unitary, *i.e.*, $\mathbf{U}\mathbf{U}^H = \mathbf{U}^H\mathbf{U} = \mathbf{I}$), and $\mathbf{D}$ is a diagonal matrix with the DFT coefficients of the convolution kernel represented by $\mathbf{B}$. Thus (with $\mathbf{B}^T = \mathbf{B}^H$, since $\mathbf{B}$ is a real matrix)

$$\left[\mathbf{B}^T\mathbf{B} + \mathbf{I}\right]^{-1} = \mathbf{U}^H\left[|\mathbf{D}|^2 + \mathbf{I}\right]^{-1}\mathbf{U}, \tag{15}$$

where $|\mathbf{D}|^2$ is the matrix with the squared absolute values of the entries of $\mathbf{D}$. Since $|\mathbf{D}|^2 + 2\,\mathbf{I}$ is diagonal, its inversion has $O(n)$ cost. Products by $\mathbf{U}$ and $\mathbf{U}^H$ can be carried out with $O(n\log n)$ cost using the FFT algorithm.

[4] The notation $\mathbf{A}^H$ denotes the conjugate transpose of matrix $\mathbf{A}$.

2) Image Inpainting: In image inpainting problems, the observed image $\mathbf{y}$ results from the loss of some elements of $\mathbf{x}$; the corresponding $m \times n$ (with $m < n$) binary matrix $\mathbf{B}$ is thus a subset of the rows of an $n \times n$ identity matrix. In this case, $\mathbf{B}^T\mathbf{B}$ is a diagonal matrix with ones and zeros in the diagonal (with the zeros corresponding to the missing elements and the ones to the observed elements). Consequently, $\mathbf{B}^T\mathbf{B} + \mathbf{I}$ is a diagonal matrix and its inversion is very inexpensive: $O(n)$.

3) Compressive Fourier Imaging: The final case considered is that of partial Fourier observations, which is used to model magnetic resonance imaging (MRI) [24], and has been the focus of much recent interest due to its connection to compressed sensing [8], [13]. In this case $\mathbf{B} = \mathbf{C}\mathbf{U}$, where $\mathbf{C}$ is an $m \times n$ (with $m < n$) binary matrix, similar to the observation matrix in the inpainting problem, and $\mathbf{U}$ is the DFT matrix. In this case,

$$\begin{aligned} \left[\mathbf{B}^T\mathbf{B} + \mathbf{I}\right]^{-1} &= \left[\mathbf{U}^H\mathbf{C}^T\mathbf{C}\mathbf{U} + \mathbf{I}\right]^{-1} \\ &= \mathbf{I} - \mathbf{U}^H\mathbf{C}^T\left[\mathbf{C}\,\mathbf{U}\,\mathbf{U}^H\mathbf{C}^T + \mathbf{I}\right]^{-1}\mathbf{C}\,\mathbf{U} \\ &= \mathbf{I} - \frac{1}{2}\,\mathbf{U}^H\mathbf{C}^T\mathbf{C}\,\mathbf{U}, \end{aligned} \tag{16}$$

where the second equality results from the application of the famous Sherman-Morrison-Woodbury (SMW) matrix inversion formula, and the third one from the fact that $\mathbf{U}\mathbf{U}^H = \mathbf{I}$ and $\mathbf{C}\mathbf{C}^T = \mathbf{I}$. Again, the cost of computing and applying this matrix is dominated by the $O(n \log n)$ cost of the FFT implementations of the products by $\mathbf{U}$ and $\mathbf{U}^H$.

C. Tikhonov Synthesis Regularization

The synthesis formulation of Tikhonov regularization yields the following unconstrained optimization problem:

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{B}\mathbf{W}\mathbf{x}\|_2^2 + \alpha\|\mathbf{x}\|_1. \tag{17}$$

The standard approach for solving (17) is the so-called iterative shrinkage/thresholding (IST) algorithm [10], [12], [20]. However, IST is known to be quite slow, specially when $\mathbf{B}$ is poorly conditioned, a fact that has stimulated much research aimed at developing faster variants [4], [5], [32].

Problem (17) has the canonical form (2), with $J = 2$, $g_1(\mathbf{u}) = \frac{1}{2}\|\mathbf{y} - \mathbf{u}\|_2^2$, $g_2(\mathbf{u}) = \alpha\|\mathbf{u}\|_1$, $\mathbf{H}^{(1)} = \mathbf{B}\mathbf{W}$, and $\mathbf{H}^{(2)} = \mathbf{I}$. The resulting ADMM algorithm is also termed SALSA [1]. Notice that convergence of SALSA in this case is also guaranteed by Theorem 1; matrix $\mathbf{G}$ (see (3)), in this case is equal to $\mathbf{G} = [(\mathbf{B}\mathbf{W})^T \;\; \mathbf{I}]^T$, which has full column rank regardless of $\mathbf{B}\mathbf{W}$

The proximity operators of these g_1 and g_2 are as in (11) and (12); the matrix inversion in (5) has the form

$$\begin{aligned} \left[\mathbf{W}^T\mathbf{B}^T\mathbf{B}\mathbf{W} + \mathbf{I}\right]^{-1} &= \mathbf{I} - \mathbf{W}^T\mathbf{B}^T\left[\mathbf{B}\mathbf{W}\mathbf{W}^T\mathbf{B}^T + \mathbf{I}\right]^{-1}\mathbf{B}\mathbf{W} \\ &= \mathbf{I} - \mathbf{W}^T\mathbf{B}^T\left[\mathbf{B}\mathbf{B}^T + \mathbf{I}\right]^{-1}\mathbf{B}\mathbf{W}, \end{aligned} \tag{18}$$

where the first inequality results from the application of the SMW matrix inversion formula and the second one from the fact that $\mathbf{W}$ contains a Parseval frame, thus $\mathbf{W}\mathbf{W}^T = \mathbf{I}$. We are thus left with the problem of inverting matrix $\mathbf{B}\mathbf{B}^T + \mathbf{I}$, which again depends of the particular problem at hand.

1) Periodic Deconvolution: If $\mathbf{B}$ represents a periodic convolution, $\mathbf{B} = \mathbf{U}^H\mathbf{D}\mathbf{U}$,

$$\left[\mathbf{B}\mathbf{B}^T + \mathbf{I}\right]^{-1} = \mathbf{U}^H\left[|\mathbf{D}|^2 + \mathbf{I}\right]^{-1}\mathbf{U}, \tag{19}$$

exactly as in (15). Inserting this equality in (18) yields

$$\left[\mathbf{W}^T\mathbf{B}^T\mathbf{B}\mathbf{W} + \mathbf{I}\right]^{-1} = \mathbf{I} - \mathbf{W}^T\mathbf{U}^H\mathbf{D}\left[|\mathbf{D}|^2 + \mathbf{I}\right]^{-1}\mathbf{D}\mathbf{U}\mathbf{W}. \tag{20}$$

Since matrix $\mathbf{D}\left[|\mathbf{D}|^2 + \mathbf{I}\right]^{-1}\mathbf{D}$ is diagonal, the cost of matrix-vector products by the matrix in (20) is $O(n \log n)$, corresponding to FFT implementations of the products by $\mathbf{U}$ and $\mathbf{U}^H$ and of the fast frame analysis ($\mathbf{W}^T$) and synthesis ($\mathbf{W}$).

2) Image Inpainting: In the image inpainting problem, $\mathbf{B}\mathbf{B}^T = \mathbf{I}$, thus $\left(\mathbf{B}\mathbf{B}^T + \mathbf{I}\right)^{-1} = \frac{1}{2}\mathbf{I}$. Inserting this equality into (18), we obtain

$$\left[\mathbf{W}^T\mathbf{B}^T\mathbf{B}\mathbf{W} + \mathbf{I}\right]^{-1} = \;=\; \mathbf{I} - \frac{1}{2}\mathbf{W}^T\mathbf{B}^T\mathbf{B}\mathbf{W}, \tag{21}$$

Since matrix $\mathbf{B}^T\mathbf{B}$ is diagonal, the cost of products by the matrix in (21) is $O(n \log n)$, corresponding to fast frame analysis ($\mathbf{W}^T$) and synthesis ($\mathbf{W}$) operations.

3) Compressive Fourier Imaging: In the case partial Fourier observations, $\mathbf{B} = \mathbf{C}\mathbf{U}$, where, as above, $\mathbf{U}$ is the DFT matrix and $\mathbf{C}$ contains a subset of the rows of an identity. In this case,

$$\left[\mathbf{B}\mathbf{B}^H + \mathbf{I}\right]^{-1} = \left[\mathbf{C}\mathbf{U}\mathbf{U}^H\mathbf{C}^T + \mathbf{I}\right]^{-1} = \frac{1}{2}\mathbf{I}, \tag{22}$$

again because $\mathbf{U}\mathbf{U}^H = \mathbf{I}$ and $\mathbf{C}\mathbf{C}^T = \mathbf{I}$. Inserting this equality and $\mathbf{B} = \mathbf{C}\mathbf{U}$ into (18) yields

$$\left[\mathbf{W}^T\mathbf{B}^T\mathbf{B}\mathbf{W} + \mathbf{I}\right]^{-1} = \mathbf{I} - \frac{1}{2}\,\mathbf{W}^T\mathbf{U}^H\mathbf{C}^T\mathbf{C}\,\mathbf{U}\,\mathbf{W}. \tag{23}$$

Since $\mathbf{C}^T\mathbf{C}$ is diagonal, the cost of products by the matrix in (23) is $O(n \log n)$, corresponding to fast frame analysis ($\mathbf{W}^T$) and synthesis ($\mathbf{W}$) operations and the FFT implementations of the products by $\mathbf{U}$ and $\mathbf{U}^H$.

D. Morozov Analysis Regularization

The analysis formulation of Morozov regularization yields the following constrained optimization problem:

$$\min_{\mathbf{x}} \|\mathbf{W}^T\mathbf{x}\|_1 \;\text{ subject to }\; \frac{1}{2}\|\mathbf{y} - \mathbf{B}\mathbf{x}\|_2^2 \leq 2\,\delta. \tag{24}$$

Problem (24) can be rewritten as

$$\min_{\mathbf{x}} \|\mathbf{W}^T\mathbf{x}\|_1 + \iota_{\mathcal{B}_{2\,\delta}(\mathbf{y})}(\mathbf{B}\mathbf{x}), \tag{25}$$

where $\iota_{\mathcal{S}}(\mathbf{x})$ is the indicator function of set $\mathcal{S}$, defined as

$$\iota_{\mathcal{S}}(\mathbf{x}) = \begin{cases} 0 & \Leftarrow \quad \mathbf{x} \in \mathcal{S} \\ \infty & \Leftarrow \quad \mathbf{x} \notin \mathcal{S}, \end{cases}$$

and $\mathcal{B}_{2\,\delta}(\mathbf{y})$ is a ball of radius 2δ centered at $\mathbf{y}$.

Clearly, problem (25) has the canonical form (2), with $J = 2$, $g_1(\mathbf{u}) = \iota_{\mathcal{B}_{2\,\delta}(\mathbf{y})}(\mathbf{u})$, $g_2(\mathbf{u}) = \|\mathbf{u}\|_1$, $\mathbf{H}^{(1)} = \mathbf{B}$, and $\mathbf{H}^{(2)} = \mathbf{W}^T$. The proximity operator of this g_1 is

$$\begin{aligned}\text{prox}_{g_1/\mu}(\mathbf{s}) &= \arg\min_{\mathbf{x}} \frac{\mu}{2}\|\mathbf{s}-\mathbf{x}\|_2^2 + \iota_{\mathcal{B}_{2\,\delta}(\mathbf{y})}(\mathbf{x}) \\ &= \mathbb{P}_{\mathcal{B}_{2\,\delta}(\mathbf{y})}(\mathbf{s}),\end{aligned} \tag{26}$$

where $\mathbb{P}_{\mathcal{S}}$ denotes an Euclidean projection on a set $\mathcal{S}$. As above, $\text{prox}_{g_2/\mu}(\mathbf{s}) = \text{soft}(\mathbf{s}, 1/\mu)$.

The matrix inversion in (5) has the exact same form as in (13), and all the derivations (for the analysis and synthesis formulations of periodic deconvolution, inpainting, and compressive Fourier imaging) carried out for the Tikhonov regularization also apply in this case. The resulting algorithm was proposed in [2] and was termed CSALSA (constrained split augmented Lagrangian shrinkage algorithm). Convergence of CSALSA results from the same arguments used to show convergence of SALSA. Finally, notice that the relationship between the Morozov analysis and synthesis formulations is exactly the same as that between the Tikhonov counterparts (the only difference is the replacement of the linear proximity operator (11) by the projection (26)), so we will abstain from studying it in detail here.

V. Poissonian Observations

A. Observation Model

Another widely used observation model in imaging is

$$\mathbf{y} \sim \mathcal{P}(\mathbf{B}\mathbf{x}) \tag{27}$$

where $\mathbf{B}$ is the matrix representation of the linear observation model and $\mathcal{P}(\boldsymbol{\lambda})$ denotes the distribution of a Poisson process of intensity vector $\boldsymbol{\lambda}$. Poissonian models are highly relevant in fields such as astronomical [30], biomedical [27], [31], and photographic imaging [21]. Given (27), the natural choice for the data term is the negative log-likelihood,

$$\Upsilon(\mathbf{y},\mathbf{x}) = \sum_i \xi((\mathbf{B}\mathbf{x})_i, y_i), \tag{28}$$

where

$$\xi(z, y) = z + \iota_{\mathbb{R}_+}(z) - y\,\log(z_+), \tag{29}$$

where $z_+ = \max\{0, z\}$ and $0\log(0) \equiv 0$ (see [18] for a detailed justification of this expression).

B. Tikhonov Analysis and Synthesis Regularization

The analysis formulation of Tikhonov regularization yields the following unconstrained optimization problem:

$$\min_{\mathbf{x}} \sum_i \xi((\mathbf{B}\mathbf{x})_i, y_i) + \alpha\|\mathbf{W}^T\mathbf{x}\|_1 + \iota_{\mathbb{R}_+^n}(\mathbf{x}), \tag{30}$$

where the indicator $\iota_{\mathbb{R}_+^n}$ is added to impose non-negativity of the solution, since the elements of $\mathbf{x}$ represent Poisson intensities. Problem (30) has the canonical form (2), with $J = 3$, $g_1(\mathbf{u}) = \sum_i \xi(u_i, y_i)$, $g_2(\mathbf{u}) = \alpha\|\mathbf{u}\|_1$, $g_3(\mathbf{u}) = \iota_{\mathbb{R}_+^n}(\mathbf{u})$, $\mathbf{H}^{(1)} = \mathbf{B}$, $\mathbf{H}^{(2)} = \mathbf{W}^T$, and $\mathbf{H}^{(3)} = \mathbf{I}$. To implement the ADMM instance introduced in Subsection III-B, the necessary building blocks are the proximity operators of g_1, g_2, and $\mathbf{g}_3$ and the matrix inversion appearing in (5). The proximity operator of g_2 is as above: $\text{prox}_{g_2/\mu}(\mathbf{s}) = \text{soft}(\mathbf{s}, 1/\mu)$. The proximity operator of g_3 is simply the projection on the first orthant:

$$\text{prox}_{g_3/\mu}(\mathbf{s}) = \max\{\mathbf{s}, 0\}. \tag{31}$$

Concerning the proximity operator of prox_{g_1}, it can be shown that it is given (component-wise) by

$$\Big(\text{prox}_{g_1/\mu}(\mathbf{s})\Big)_i = \frac{1}{2}\left(s_i - \frac{1}{\mu} + \sqrt{\left(s_i - \frac{1}{\mu}\right)^2 + \frac{4\,y_i}{\mu}}\right), \tag{32}$$

Notice that $\Big(\text{prox}_{g_1/\mu}(\mathbf{s})\Big)_i$ is always non-negative.

All the derivations made in the previous section concerning the matrix inversion in (5) apply unchanged to this case. The resulting classof algorithm was proposed in [18] and was termed PIDAL (Poisson image deconvolution via augmented Lagrangia). Finally, notice that convergence of PIDAL is guaranteed by Theorem 1, since matrix $\mathbf{G}$ (see (3)), which in this case is equal to $\mathbf{G} = [\mathbf{B}^T\ \mathbf{W}\ \mathbf{I}]^T$, has full column rank due to the presence of $\mathbf{I}$.

The ADMM algorithm for the Tikhonov synthesis regularization for linear-Poisson observations (also termed PIDAL in [18]) is obtained by using the same g_1, g_2, and g_3 functions, and $\mathbf{H}^{(1)} = \mathbf{B}\mathbf{W}$, $\mathbf{H}^{(2)} = \mathbf{I}$, and $\mathbf{H}^{(3)} = \mathbf{I}$. All the derivations made for linear-Gaussian case concerning the matrix inversion in (5) also apply unchanged to this case.

Finally, we mention that in the linear-Poisson case, the Morozov formulation is not as straightforward as in the Gaussian case. In fact, the required projection (that takes the place of (26)) doesn't have a simple closed form solution, and has to be computed numerically [9].

VI. Hybrid Analysis-Synthesis Regularization

Although some research has focused on comparing the analysis and synthesis formulations [16], [28], there is no consensus on which of the two is to be preferred for a given problem. This choice can be avoided by combining the two formulations into a hybrid synthesis-analysis criterion. Considering linear-Gaussian observations (the application to the linear-Poisson case is straightforward), one possible hybrid (Tikhonov-type) formulation is

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{B}\,\mathbf{W}_1\,\mathbf{x}\|_2^2 + \alpha\|\mathbf{x}\|_1 + \beta\|\mathbf{W}_2^T\,\mathbf{W}_1\,\mathbf{x}\|_1, \tag{33}$$

where $\mathbf{W}_1$ and $\mathbf{W}_2$ are the synthesis matrices of two Parseval frames (the same or two different ones). Clearly, problem (33) can be written in the canonical form (2), with $J = 3$, $g_1(\mathbf{u}) = \frac{1}{2}\|\mathbf{y}-\mathbf{u}\|_2^2$, $g_2(\mathbf{u}) = \alpha\|\mathbf{u}\|_1$, $g_3(\mathbf{u}) = \beta\|\mathbf{u}\|_1$, $\mathbf{H}^{(1)} = \mathbf{B}\mathbf{W}_1$, $\mathbf{H}^{(2)} = \mathbf{I}$, and $\mathbf{H}^{(3)} = \mathbf{W}_2^T\,\mathbf{W}_1$. The proximity operator of g_1 is the linear shrinkage in (11), while those of g_2 and g_3 are soft thresholds (12). The final component needed is the matrix inverse in (5); since $\mathbf{W}_2\,\mathbf{W}_2^T = \mathbf{I}$, we obtain

$$\left[\mathbf{W}_1^T(\mathbf{B}^T\mathbf{B} + \mathbf{I})\mathbf{W}_1 + \mathbf{I}\right]^{-1}. \tag{34}$$

Using again the SMW matrix inversion formula, (34) can be re-written as

$$\begin{aligned}\mathbf{I} - \mathbf{W}_1^T\left[(\mathbf{B}^T\mathbf{B}+\mathbf{I})^{-1} + \mathbf{W}_1\mathbf{W}_1^T\right]^{-1}\mathbf{W}_1 \\ = \mathbf{I} - \mathbf{W}_1^T\left[(\mathbf{B}^T\mathbf{B}+\mathbf{I})^{-1} + \mathbf{I}\right]^{-1}\mathbf{W}_1. \qquad (35)\end{aligned}$$

It is easy to show that in the three cases studied in Subsection IV-B (periodic deconvolution, inpainting, compressive Fourier imaging), the matrices being inverted in (35) are diagonal, yielding $O(n \log n)$ cost for computing and multiplying by this matrix. Finally, notice that convergence of the resulting ADMM algorithm is guaranteed by the fact that matrix $\mathbf{G} = \left[(\mathbf{B}\mathbf{W}_1)^T \;\; \mathbf{I} \;\; (\mathbf{W}_2^T\mathbf{W}_1)^T\right]^T$

VII. Conclusion

We have reviewed some of our recent work on using the alternating direction method of multipliers (ADMM) to solve a variety of convex optimization problems arising in imaging inverse problems. We presented an integrated view of several formulations for different problems, based on an instantiation of ADMM for sums of two or more functions. The matrix inversion required by the algorithm was shown to be cheaply computable in several cases of interest (periodic deconvolution, inpainting, compressive Fourier observations). We also showed that the algorithm can be seamlessly used for unconstrained (Tikhonov) or constrained (Morozov) regularization, and for analysis and synthesis formulations. The algorithms are shown to satisfy sufficient conditions for convergence. We have not presented any experimental results, since comprehensive experimental assessment of this approach can be found in [1], [2], [18]; the conclusions of those assessments is that the ADMM-based algorithms exhibits state-of-the-art speed.

Acknowledgment

This work was partially supported by *Fundação para a Ciência e Tecnologia* (FCT), under projects PTDC/EEA-TEL/104515/2008 and UTAustin/MAT/0009/2008.

References

[1] M. Afonso, J. Bioucas-Dias, M. Figueiredo, "Fast image recovery using variable splitting and constrained optimization", *IEEE Trans. Image Processing,* vol. 19, pp. 2345-2356, 2010.

[2] M. Afonso, J. Bioucas-Dias, M. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems", *IEEE Trans. Image Processing,*, vol. 20, pp. 681–695, 2011.

[3] M. Afonso, J. Bioucas-Dias, M. Figueiredo, "Hybrid synthesis-analysis frame-based regularization: a criterion and an algorithm", *SPARS11*, Edinburgh, UK, 2011.

[4] A. Beck, M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems", *SIAM Jour. Imaging Sciences*, vol. 2, pp. 183–202, 2009.

[5] J. Bioucas-Dias, M. Figueiredo, "A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration", *IEEE Trans. Image Processing*, vol. 16, pp. 2992-3004, 2007.

[6] J. Bioucas-Dias, M. Figueiredo, "Multiplicative noise removal using variable splitting and constrained optimization", *IEEE Trans. Image Processing,* vol. 19, pp. 1720–1730, 2010.

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

[8] E. Candès, J. Romberg, T. Tao. "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Information Theory,* vol. 52, pp. 489–509, 2006.

[9] M. Carlavan, L. Blanc–Féraud, "Two constrained formulations for deblurring Poisson noisy images", ICIP, 2011.

[10] P. Combettes, V. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM Jour. Multiscale Modeling & Simulation*, vol. 4, pp. 1168–1200, 2005.

[11] P. Combettes, J.-C. Pesquet, "Proximal thresholding algorithm for minimization over orthonormal bases," *SIAM Jour. Optimization*, vol. 18, pp.1351–1376, 2007.

[12] I. Daubechies, M. De Friese, C. De Mol. "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint." *Commu. Pure and Applied Math.*, vol. 57, pp. 1413–1457, 2004.

[13] D. Donoho. "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, 2006.

[14] J. Eckstein, D. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators", *Math. Program.,* vol. 5, pp. 293-318, 1992.

[15] M. Elad, M. Figueiredo, Y. Ma, "On the role of sparse and redundant representations in image processing", *Proc. IEEE*, vol. 98, pp. 972–982, 2010.

[16] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors", *Inverse Problems*, vol. 23, pp. 947–968, 2007.

[17] E. Esser, "Applications of Lagrangian-based alternating direction methods and connections to split Bregman", Tech. Rep. 09-31, Comp. and Applied Math., UCLA, 2009.

[18] M. Figueiredo, J. Bioucas-Dias, "Restoration of Poissonian images using alternating direction optimization", *IEEE Trans. Image Proc.,* vol. 19, pp. 3133–3145, 2010.

[19] M. Figueiredo, J. Bioucas-Dias, "An alternating direction algorithm for (overlapping) group regularization", *Proc. Signal Processing with Adaptive Sparse Structured Representations – SPARS11*, Edinburgh, 2011.

[20] M. Figueiredo, R. Nowak, "An EM algorithm for wavelet-based image restoration." *IEEE Trans. Image Processing*, vol. 12, pp. 906–916, 2003.

[21] A. Foi, S. Alenius, M. Trimeche, V. Katkovnik, K. Egiazarian, "A spatially adaptive Poissonian image deblurring, *ICIP*, Genova, 2005.

[22] D. Gabay, B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite-element approximations", *Computers and Math. with Applic.*, vol. 2, pp. 17–40, 1976.

[23] R. Glowinski, A. Marroco, "Sur l'approximation, par elements finis d'ordre un, et la resolution, par penalisation-dualité d'une classe de problemes de Dirichlet non lineares", *Revue Française d'Automatique, Informatique et Recherche Opérationelle*, vol. 9, pp. 41–76, 1975.

[24] M. Lustig, D. Donoho, J. Pauly, "Sparse MRI: the application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine,* vol. 58, pp. 1182–1195, 2007.

[25] S. Mallat, *A Wavelet Tour of Signal Processing* (3rd Edition), Academic Press, 2008.

[26] N. Rao, N. Kingsbury, S. Wright, R. Nowak, "Convex approaches to model wavelet sparsity patterns", *Proc. IEEE ICIP*, Brussels, Belgium, 2011.

[27] P. Sarder and A. Nehorai. "Deconvolution method for 3-D fluorescence microscopy images", *IEEE Signal Proc. Mag.,* vol. 23, pp. 32-45, 2006.

[28] I. Selesnick, M. Figueiredo, "Signal restoration with overcomplete wavelet transforms: comparison of analysis and synthesis priors", *Proceedings of SPIE*, vol. 7446 (Wavelets XIII), 2009.

[29] S. Setzer, G. Steidl, and T. Teuber, "Deblurring Poissonian images by split Bregman techniques", *Journal of Visual Communication and Image Representation,* 2010 (in press).

[30] J.-L. Starck, F. Murtagh, *Astronomical Image and Data Analysis,* Springer, 2006.

[31] R. Willett and R. Nowak. "Platelets: a multiscale approach for recovering edges and surfaces in photon-limited medical imaging", *IEEE Trans. Medical Imag.,* vol. 22, pp. 332–350, 2003.

[32] S. Wright, R. Nowak, M. Figueiredo, "Sparse reconstruction by separable approximation, *IEEE Trans. Signal Proc.*, vol. 57, pp. 2479-2493, 2009.

[33] W. Yin, S. Osher, D. Goldfarb, J. Darbon, "Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing", *SIAM Jour. Imaging Sciences*, vol. 1, pp. 143–168, 2008.