A PRIMAL-DUAL FRAMEWORK FOR MIXTURES OF REGULARIZERS

Baran Gözcü, Luca Baldassarre, Quoc Tran-Dinh, Cosimo Aprile and Volkan Cevher

LIONS - EPFL
Lausanne - Switzerland

ABSTRACT

Effectively solving many inverse problems in engineering requires to leverage all possible prior information about the structure of the signal to be estimated. This often leads to tackling constrained optimization problems with mixtures of regularizers. Providing a general purpose optimization algorithm for these cases, with both guaranteed convergence rate as well as fast implementation remains an important challenge. In this paper, we describe how a recent primal-dual algorithm for non-smooth constrained optimization can be successfully used to tackle these problems. Its simple iterations can be easily parallelized, allowing very efficient computations. Furthermore, the algorithm is guaranteed to achieve an optimal convergence rate for this class of problems. We illustrate its performance on two problems, a compressive magnetic resonance imaging application and an approach for improving the quality of analog-to-digital conversion of amplitude-modulated signals.

***Index Terms*—** Constrained Non-Smooth Optimization, Inverse Problems, Compressive Sensing, MRI, ADC.

1. INTRODUCTION

Many inverse problems in engineering, from signal denoising [1] to compressive imaging [2] and from tomographic reconstruction [3] to brain decoding [4], require solving optimization problems that are composed of multiple convex terms, usually a data fit term and one or more regularizers that favour the structures that are present in the signal to be estimated. In many cases, linear constraints that define further properties of the signal are also imposed.

In general, we are interested in solving a constrained convex optimization problem of the form:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) := \sum_{i=0}^{p} f_i(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned} \tag{1}$$

where f_0 is the data fit term, f_i, $i = 1, \ldots, p$, are possibly non-smooth regularization terms, $\mathbf{A}$ is a linear operator and $\mathbf{b}$ is an observed measurement vector. For example, in compressive magnetic resonance imaging, the data fit term is usually the square loss $\frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$, where $\mathbf{A}$ is a partial Fourier sampling operator, $\mathbf{b}$ contains the compressive samples and $\mathbf{x}$ is the image to be estimated. Frequently used regularization terms for this problem are the ℓ_1 norm computed on the wavelet coefficients of $\mathbf{x}$, i.e., $\|\mathbf{W}\mathbf{x}\|_1$, where $\mathbf{W}$ is the discrete wavelet transform, and the Total Variation norm, that is the sum of the norms of the discrete gradients computed at each pixel of the image, $\|\mathbf{x}\|_{\mathrm{TV}} = \sum_{i,j} \|(\nabla(\mathbf{x}))_{i,j}\|_2$ [2]. Since the image is described as intensity values, a positivity constraint can also be imposed.

Researchers are often interested in combining multiple regularization terms, in order to enforce all available prior information about the signal to be estimated. However, it is not straightforward to adapt optimization algorithms that have been developed ad-hoc for a certain combination of regularizers to deal with new terms. Moreover, some of the proposed techniques are often riddled with heuristics that, if on one hand, allow efficient implementations, on the other hand do not guarantee convergence to optimal solutions.

Here, we illustrate how a recently proposed optimization framework [5, 6] for solving constrained convex optimization problems can be leveraged for mixtures of regularizers (1). This framework can deal with any linear equality, and inequality, constraints and multiple regularization terms in a principled way, with rigorous convergence guarantees, both on the decrease of the objective function value and on the feasibility of the constraints. It is based on smoothing the primal-dual gap function via Bregman distances and on a particular model-based gap reduction condition. Its iterations are efficient since they consist only of proximal computations and matrix multiplications with easy parallelization. We demonstrate its performance on a compressive magnetic resonance imaging problem with three regularizers and on an approach for improving the quality of analog-to-digital conversion of amplitude modulated signals.

1.1. Previous work

One of the first work on solving composite and constrained convex problems with non-smooth functions is [7], where the authors propose the Predictor Corrector Proximal Multiplier

This work was supported in part by the European Commission under grants MIRG-268398 and ERC Future Proof and by the Swiss Science Foundation under grants SNF 200021-146750 and SNF CRSII2-147633.

method which is based on Rockafellar's Proximal Point algorithm [8], whose theory is founded on inclusions of monotone operators. The method combines proximal steps in the dual as well as the primal problem, resembling the structure of the 1P2D algorithm proposed in [5], however its linear converge rate is guaranteed only by assuming a strong regularity condition which is impossible to check in practice.

Combettes and Pesquet [9] proposed a decomposition method for solving (1) by involving each function f_i independently via its proximity operator while the constraint can be dealt with by an indicator function. However, the proximity operator derived from the constraint may not be available in closed form and could require the solution of a smooth problem, for example by fast gradient method of Nesterov [10]. Furthermore, although a sequence produced by the algorithm is guaranteed to converge to a minimizer, which is not the case in [5, 6], neither convergence rate nor control of the feasibility gap is provided.

Combettes et al. [11] proposed an algorithm for computing the proximity operator of a sum of convex functions composed with linear operators that requires to compute proximity operators of each function independently. This algorithm could be used in combination with the proximal point algorithm, but again no convergence rates are obtained.

The primal-dual algorithm developed by Chambolle and Pock [12] for composite unconstrained convex problems with two terms could also be extended to tackle (1). However, since the algorithm was derived for solving a more general class of problems, it comes with weaker guarantees than the decomposition method of [5].

2. PRELIMINARIES

Scalars are denoted by lowercase letters, vectors by lowercase boldface letters and matrices by uppercase boldface letters. We say that the proper, closed and convex function $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ has tractable proximity operator if $\mathrm{prox}_f(\mathbf{w}) := \arg\min_{\mathbf{u}}\{f(\mathbf{u}) + (1/2)\|\mathbf{u} - \mathbf{w}\|^2\}$ can be computed efficiently for any $\mathbf{w}$ (e.g., by a closed form or a polynomial time algorithm) [13].

3. A PRIMAL-DUAL FRAMEWORK

We consider the general constrained convex optimization problem

$$f^* := \min_{\mathbf{x}} \{f(\mathbf{x}) : \mathbf{A}\mathbf{x} = \mathbf{b}\}, \quad (2)$$

where $f(\mathbf{x})$ is a convex, closed and proper function and $\mathbf{A}\mathbf{x} = \mathbf{b}$ is a linear constraint.

Recently, [5] proposed an efficient primal-dual optimization algorithm for solving (2). Via Lagrange dualization, its dual problem is

$$d^* := \max_{\mathbf{y}} d(\mathbf{y}), \quad (3)$$

where $d(\mathbf{y}) := \min_{\mathbf{x}} \big\{f(\mathbf{x}) + \langle \mathbf{A}\mathbf{x} - \mathbf{b}, \mathbf{y}\rangle\big\}$. Let $\mathbf{x}^*$ and $\mathbf{y}^*$ be an optimal solution of (2) and (3), respectively. Under strong duality, we have $d^* = f^*$. The dual function d is concave but in general non-smooth. We define the smoothed dual function:

$$d_\gamma(\mathbf{y}) := \min_{\mathbf{x}} \big\{f(\mathbf{x}) + \langle \mathbf{A}\mathbf{x} - \mathbf{b}, \mathbf{y}\rangle + \frac{\gamma}{2}\|\mathbf{S}(\mathbf{x} - \mathbf{x}_c)\|^2\big\}, \quad (4)$$

where $\gamma > 0$ is sufficiently small, $\mathbf{S}$ can be either $\mathbf{I}$ or $\mathbf{A}$, and $\mathbf{x}_c$ will be specified later. For given d_γ and $\beta > 0$, we define the smoothed gap function:

$$G_{\gamma\beta}(\mathbf{x}, \mathbf{y}) := f(\mathbf{x}) + (1/(2\beta))\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 - d_\gamma(\mathbf{y}). \quad (5)$$

The goal is to generate a sequence $\{(\bar{\mathbf{x}}^k, \bar{\mathbf{y}}^k, \gamma_k, \beta_k)\}$ such that $G_{\gamma_k\beta_k}(\bar{\mathbf{x}}^k, \bar{\mathbf{y}}^k)$ decreases to zero and γ_k, β_k also tend to 0^+. It has been proved in [6] that, if the primal-dual updates satisfy a *model-based excessive gap* reduction condition, then $\|\mathbf{A}(\bar{\mathbf{x}}^k) - \mathbf{b}\| \le M_1\beta_k$ and $|f(\bar{\mathbf{x}}^k) - f^*| \le M_2\gamma_k$ for given constants M_1 and M_2.

In [6], two different primal-dual updates that satisfy the model-based excessive gap reduction condition were proposed, with either one primal and two dual steps (1P2D), or two primal and one dual steps (2P1D). For conciseness, we present only the more efficient 1P2D updates.

For given $\mathbf{x}_c$, $\mathbf{y}$, γ and $\mathbf{S}$, let us denote by $\mathbf{x}^*_\gamma(\mathbf{y}, \mathbf{x}_c)$ the solution of the minimization problem in (4). If $\mathbf{S} = \mathbf{I}$, solving this problem amounts to computing the proximity operator of f. When $\mathbf{S} = \mathbf{A}$, $\mathbf{x}^*_\gamma(\mathbf{y}, \mathbf{x}_c)$ can be computed iteratively using FISTA, which still requires prox_f. The updates are given by:

$$\begin{array}{ll} \hat{\mathbf{y}}^k & := (1-\tau_k)\bar{\mathbf{y}}^k + \tau_k\beta_k^{-1}(\mathbf{A}\bar{\mathbf{x}}^k - \mathbf{b}) \\ \bar{\mathbf{x}}^{k+1} & := (1-\tau_k)\bar{\mathbf{x}}^k + \tau_k\mathbf{x}^*_{\gamma_{k+1}}(\hat{\mathbf{y}}^k) \\ \bar{\mathbf{y}}^{k+1} & := \hat{\mathbf{y}}^k + \frac{\gamma_{k+1}}{L}\left(\mathbf{A}\mathbf{x}^*_{\gamma_{k+1}}(\hat{\mathbf{y}}^k) - \mathbf{b}\right), \end{array} \quad \text{(1P2D)}$$

where $L = 1$ if $\mathbf{S} = \mathbf{I}$ and $L = \|\mathbf{A}\|_2^2$ if $\mathbf{S} = \mathbf{A}$. The parameters are updated as $\beta_{k+1} = (1-\tau_k)\beta_k$ and $\gamma_{k+1} = (1 - c_k\tau_k)\gamma_k$, where also τ_k and c_k are updated appropriately, see [6] for further details. Now, we are ready to present a complete primal-dual algorithm for solving (2).

Algorithm 1 *Primal-dual decomposition algorithm* (DecOpt)

Inputs: Choose $\gamma_0 > 0$ and $c_0 \in (-1, 1]$.
1: $a_0 := (1 + c_0 + [4(1-c_0) + (1+c_0)^2]^{1/2})/2$ and $\tau_0 := a_0^{-1}$.
2: Set $\beta_0 := \gamma_0^{-1}L$.
3: Compute the initial point $(\bar{\mathbf{x}}^0, \bar{\mathbf{y}}^0)$ as described in [6].
For $k = 0$ **to** $k_{\max}$:
4: If the stopping criterion meets, then terminate.
5: Update γ_{k+1}.
6: Update $(\bar{\mathbf{x}}^{k+1}, \bar{\mathbf{y}}^{k+1})$ by (1P2D).
7: Update β_{k+1}. Also update c_{k+1} if necessary.
8: Update $a_{k+1} := (1 + c_{k+1} + [4a_k^2 + (1-c_k)^2]^{1/2})/2$ and set $\tau_{k+1} := a_{k+1}^{-1}$.
End For

The convergence rate of DecOpt is established by the following theorem.

Theorem 3.1 ([6]). *Let* $\left\{(\bar{\mathbf{x}}^k, \bar{\mathbf{y}}^k)\right\}_{k\geq 0}$ *be the sequence generated by Algorithm 1 after* $k \geq 1$ *iterations. Then, if* $\mathbf{S} = \mathbf{A}$*, we have:*

a) *If* $c_k := 0$ *for all* $k \geq 0$*,* $\gamma_0 := L = 1$*, then for all* $k \geq 0$*:*

$$\begin{cases} \|\mathbf{A}\bar{\mathbf{x}}^k - \mathbf{b}\|_2 \leq \frac{C_1}{(k+1)^2}, \\ -\frac{1}{2}\|\mathbf{A}\bar{\mathbf{x}}^k - \mathbf{b}\|_2^2 - C_2\|\mathbf{A}\bar{\mathbf{x}}^k - \mathbf{b}\|_2 \leq f(\bar{\mathbf{x}}^k) - f^\star \leq 0, \end{cases} \quad (6)$$

where $C_1, C_2 > 0$ *are defined in [6]. As a consequence, the worst-case analytical complexity of Algorithm 1 to achieve an* ε*-primal solution* $\bar{\mathbf{x}}^k$ *for* (2) *is* $\mathcal{O}\left(\varepsilon^{-1/2}\right)$.

b) *Alternatively, if* $\mathbf{S} = \mathbf{I}$*, we have: If* $\gamma_0 := \frac{2\sqrt{2L}}{K+1}$ *and* $c_k := 0$ *for all* $k = 0, \ldots, K$*, then:*

$$\begin{cases} \|\mathbf{A}\bar{\mathbf{x}}^K - \mathbf{b}\|_2 \leq \frac{C_3}{K+1}, \\ -C_4\|\mathbf{A}\bar{\mathbf{x}}^K - \mathbf{b}\|_2 \leq f(\bar{\mathbf{x}}^K) - f^\star \leq \frac{C_5}{K+1}. \end{cases} \quad (7)$$

where $C_3, C_4, C_5 > 0$ *are defined in [6]. As a consequence, the worst-case analytical complexity of Algorithm 1 to achieve an* ε*-primal solution* $\bar{\mathbf{x}}^k$ *for* (2) *is* $\mathcal{O}\left(\varepsilon^{-1}\right)$.

In order to accelerate Algorithm 1, we adaptively compute $\mathbf{x}^*_{\gamma_{k+1}}(\hat{\mathbf{y}}^k, \mathbf{x}_c^k)$: The center point $\mathbf{x}_c^k$ is updated as $\mathbf{x}_c^{k+1} := \mathbf{x}^*_{\gamma_{k+1}}(\hat{\mathbf{y}}^k, \mathbf{x}_c^k) + \mathbf{d}^k$, where $\mathbf{d}^k$ is the gradient descent direction obtained by linearizing the augmented Lagrangian term $\langle \mathbf{A}\mathbf{x} - \mathbf{b}, \mathbf{y}\rangle + (\gamma/2)\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$ as in preconditioned ADMM [12]. We also set $c_k = 1.05$ [6].

4. VARIABLE SPLITTING

In order to solve problem (1) with DecOpt, we formulate it as

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimize}} \quad f(\mathbf{x}) := \sum_{i=0}^{p} f_i(\mathbf{A}_i\mathbf{x} + \mathbf{b}_i) \\ &\text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned} \quad (8)$$

For example, $\|\mathbf{W}\mathbf{x}\|_1$ can be seen as $f_i = \|\cdot\|_1$ with $\mathbf{A}_i = \mathbf{W}$ and $\mathbf{b}_i = \mathbf{0}$. We now consider the splitting $\mathbf{x}_i = \mathbf{A}_i\mathbf{x} + \mathbf{b}_i$, which leads to

$$\begin{aligned} &\underset{\mathbf{x}, \tilde{\mathbf{x}} = [\mathbf{x}_0^T, \ldots, \mathbf{x}_p^T]^T}{\text{minimize}} \quad f(\tilde{\mathbf{x}}) := \sum_{i=0}^{p} f_i(\mathbf{x}_i) \\ &\text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \\ &\qquad\qquad \mathbf{A}_i\mathbf{x} + \mathbf{b}_i = \mathbf{x}_i, \forall i \end{aligned} \quad (9)$$

where each function f_i now depends only on the variable $\mathbf{x}_i$. As seen in Algorithm 1, we need to compute the proximal operators of $f(\tilde{\mathbf{x}})$, which due to the splitting in (9), amounts to calculating the proximal operator of each function $f_i(\mathbf{x}_i)$ and concatenating them:

$$\begin{aligned} \text{prox}_{\gamma f}(\tilde{\mathbf{x}}) &= \underset{\tilde{\mathbf{y}} = [\mathbf{y}_0 \ldots \mathbf{y}_p]^T}{\text{argmin}} \sum_{i=0}^{p} \gamma f_i(\mathbf{y}_i) + \sum_{i=0}^{p} \frac{1}{2}\|\mathbf{y}_i - \mathbf{x}_i\|_2^2 \\ &= [\text{prox}_{\gamma f_0}(\mathbf{x}_0)^T, \ldots, \text{prox}_{\gamma f_p}(\mathbf{x}_p)^T]^T \end{aligned} \quad (10)$$

With this approach we can add as many penalties as we wish as long as their proximal operators are tractable. Furthermore, each $\text{prox}_{\gamma f_i}(\mathbf{x}_i)$, can be computed in parallel, leading to significant computational gains.

5. COMPRESSIVE MAGNETIC RESONANCE IMAGING

Here, we describe a mixture of regularizers used in compressive Magnetic Resonance Imaging (MRI) and how the proposed approach can deal with these formulations. MRI is based on reconstructing an image from its Fourier samples. There are three types of structures that are currently used to improve the quality of MR images: smoothness, captured by the total variation semi-norm [14], sparsity in wavelet domain, measured by ℓ_1 norm of the wavelet coefficients and hierarchical structures over the wavelet quad-tree encoded by group structures over the wavelet coefficients [15, 16]. Combining these terms together with a standard data fit terms and a positivity constraint leads to the following problem:

$$\min_{\mathbf{x}\geq 0} \frac{1}{2}\|\mathbf{M}\mathbf{x} - \mathbf{y}\|_2^2 + \alpha\|\mathbf{x}\|_{\text{TV}} + \mu\|\mathbf{W}\mathbf{x}\|_1 + \beta\|\mathbf{W}\mathbf{x}\|_{\text{tree}}, \quad (11)$$

where $\mathbf{M} \in \mathbb{C}^{m\times N}$ is the measurement matrix, $\mathbf{y} \in \mathbb{C}^m$ is the measurement vector, $\mathbf{W}$ is the wavelet transform matrix and $\mathbf{x}$ is the vectorized image that we wish to recover. The tree norm is defined as:

$$\|\mathbf{x}\|_{\text{tree}} := \sum_{i=1}^{s} \|\mathbf{x}_{g_i}\|_2, \quad (12)$$

where g_i, $i = 1, \ldots, s$ are all the parent-child pairs according to the wavelet tree and $\|\mathbf{x}_{g_i}\|_2$ is the 2-norm of the vector consisting of the elements $\mathbf{x}$ indexed by g_i.

To fit (11) to our model in (9), we coin the components of the objective function as $f_0(\mathbf{x}_0) = \|\mathbf{x}_0\|_2^2$, $f_1(\mathbf{x}_1) = \alpha\|\mathbf{x}_1\|_{\text{TV}}$ and $f_2(\mathbf{x}_2) = \mu\|\mathbf{x}_2\|_1$, whose proximal operators are well known and easy to compute. However, the computation of $\text{prox}_{\|\cdot\|_{\text{tree}}}$ is not straightforward because the parent-child groups overlap with each other. The replication approach [17] is to define an auxiliary variable $\mathbf{z}$ of size $\sum_{i=1}^{s}|g_i|$ such that $\mathbf{z}$ is the concatenation of $\mathbf{x}_{g_i}$ for all $i = 1, \ldots, s$ and then define non-overlapping groups $\tilde{g}_i$ over $\mathbf{z}$. The matrix $\mathbf{G}$ performs the mapping from $\mathbf{x}$ to $\mathbf{z}$, that is $\mathbf{z} = \mathbf{G}\mathbf{x}$. We then define:

$$\|\mathbf{W}\mathbf{x}\|_{\text{tree}} = \sum_{i=1}^{s} \|\left(\mathbf{G}\mathbf{W}\mathbf{x}\right)_{\tilde{g}_i}\|_2, \quad (13)$$

$$f_3(\mathbf{x}_3) = \beta\sum_{i=1}^{s} \|\left(\mathbf{x}_3\right)_{\tilde{g}_i}\|_2, \quad (14)$$

with the additional constraint $\mathbf{x}_3 = \mathbf{G}\mathbf{W}\mathbf{x}_1$. The remaining constraints are $\mathbf{W}\mathbf{x}_1 = \mathbf{x}_2$ and $\mathbf{x}_4 = \mathbf{M}\mathbf{x}_1 - \mathbf{y}$. We can then impose $\mathbf{x}_3 = \mathbf{G}\mathbf{x}_2$. Therefore, the linear constraint $\mathbf{A}\mathbf{x} = \mathbf{b}$ in (9) encodes the variable splittings:

$$\mathbf{A} = \begin{bmatrix} \mathbf{W} & -\mathbf{I} & 0 & 0 \\ 0 & \mathbf{G} & -\mathbf{I} & 0 \\ \mathbf{M} & 0 & 0 & -\mathbf{I} \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ \mathbf{y} \end{bmatrix}. \tag{15}$$

Based on the Fast Composite Splitting Algorithm [18], Chen and Huang [19] proposed the *Wavelet Tree Sparsity MRI* (WaTMRI) algorithm for solving (11). However, since their method cannot directly deal with constraints, instead of $\mathbf{z} = \mathbf{G}\mathbf{W}\mathbf{x}$, they added the augmented Lagrangian term $\frac{\lambda}{2}\|\mathbf{z} - \mathbf{G}\mathbf{W}\mathbf{x}\|_2$ with a fixed value of λ, which does not guarantee that the constraint will be met at convergence.

We compare the performance of DecOpt for solving (11) to WaTMRI. Note that although we use the same coefficient values for α, β, μ as in [19], WaTMRI addresses the augmented problem without the constraint. We consider a $N = 128 \times 128$ MRI brain image sampled via a partial Fourier operator at a subsampling ratio of 0.2. In Fig. 1, we show that DecOpt converges both in the objective value and in the constraint feasibility gap. We also report the objective function obtained for WaTMRI (without including the augmented term), which is observed to achieve a slightly lower objective value at the expense of violating the constraint $\mathbf{z} = \mathbf{G}\mathbf{W}\mathbf{x}$. We observe that our solution gives a more stable Signal-to-Noise Ratio (SNR) performance.

6. ANALOG-TO-DIGITAL CONVERSION

We describe an application of a mixture of regularizers with inequality constraints for improving the quality of the digital signals obtained by analog-to-digital converters (ADC).

We consider a continuous time signal $\mathbf{x}(t)$ as input of the ADC. The output is the discrete signal $\mathbf{x}[t_i]$, where $t_i = t_1, ..., t_N$ are the sampling times. Since the ADC has a fixed output resolution, the output signal $\mathbf{x}[t_i]$ is between two consecutive quantization levels: the low level ℓ_i and the upper level u_i. We call this range bucket. For instance, an 8 bit ADC has 256 levels, then $\mathbf{x}[t_i]$ can be in one of 255 buckets. We consider 1024 samples from a random band-limited signal with cut off frequency $f_{\text{cut}} = 30$MHz, which is amplitude modulated (AM) with carrier frequency $f_{\text{car}} = 1$GHz. To simulate a real ADC with jitter noise, we add white noise to the original signal before sampling to obtain a SNR of 30dB. Due to quantization, the sampled signal has SNR of 28.9dB. We expect the spectrum of the original signal to be sparse and clustered around the carrier frequency f_{car}. These properties can be favoured by two regularization terms $f_1(\mathbf{x}) = \|\mathbf{D}\mathbf{x}\|_1$ and $\|\mathbf{D}\mathbf{x}\|_{\text{TV}}$, where $\mathbf{D}$ is the Discrete Cosine Transform operator and $\|\cdot\|_{\text{TV}}$ is the discrete 1D Total Variation norm. Moreover, we allow the values $\mathbf{x}[t_i]$ to be occasionally corrupted and hence lie between $\ell_i - s_i$ and $u_i + s_i$, where $s_i > 0$ is a variable that accounts for bucket correction and which we assume to be sparse. Combining all these terms, we obtain the following optimization problem:

$$\begin{array}{ll} \text{argmin}_{\mathbf{x},\mathbf{s}\in\mathbb{R}^{1024}} & \lambda_1\|\mathbf{s}\|_1 + \lambda_2\|\mathbf{D}\mathbf{x}\|_1 + \|\mathbf{D}\mathbf{x}\|_{\text{TV}} \\ \text{subject to} & \boldsymbol{\ell} - \mathbf{s} \le \mathbf{x} \le \mathbf{u} + \mathbf{s}, \quad \mathbf{s} \ge 0, \end{array} \tag{16}$$

where we set $\lambda_1 = 2$ and $\lambda_2 = 1$. The above problem can be cast in the mixture model (9) as follows:

$$\begin{array}{ll} \underset{\mathbf{x},\mathbf{x}_1,\mathbf{x}_2,\mathbf{s},\mathbf{r}_\ell,\mathbf{r}_u}{\text{argmin}} & 2\|\mathbf{s}\|_1 + \|\mathbf{x}_1\|_1 + \|\mathbf{x}_2\|_{\text{TV}} + \delta_{\ge\boldsymbol{\ell}}(\mathbf{r}_\ell) + \delta_{\le\mathbf{u}}(\mathbf{r}_u) \\ \text{subject to} & \mathbf{s} \ge 0, \quad \mathbf{x}_1 = \mathbf{D}\mathbf{x}, \quad \mathbf{x}_2 = \mathbf{D}\mathbf{x} \\ & \mathbf{r}_\ell = \mathbf{x} + \mathbf{s}, \quad \mathbf{r}_u = \mathbf{x} - \mathbf{s}, \end{array}$$

where $\delta_{\ge\boldsymbol{\ell}}(\mathbf{r}_\ell) = 0$ if $(\mathbf{r}_\ell)_i \ge \boldsymbol{\ell}_i$, $\forall i$ and $+\infty$ otherwise. Likewise for the last term.

In Fig. 2, we report the distributions of the errors for the ideal ADC, the jittery ADC and the estimate via (16). It can be seen that the convex estimate obtains a distribution which is more concentrated around zero. The signal via the convex problem (16) achieves a SNR of 34dB, an improvement greater than 5dB over the quantized signal.

7. CONCLUSION

We presented the application of a general primal-dual method for solving inverse problems via mixtures of regularizers. The described approach not only allows a fast implementation, but also has theoretical convergence guarantees on both the objective residual and the feasibility gap. The numerical experiments show that this approach can constitute a reliable and efficient framework for dealing with decomposable constrained convex signal processing problems in a principled and unified way. We are currently developing a general convergence theory that explains the effects of the adaptive enhancements.

REFERENCES

[1] P.L. Combettes and J.-C. Pesquet, "A douglas–rachford splitting approach to nonsmooth convex variational signal recovery," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 1, no. 4, 2007.

[2] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse mri: The application of compressed sensing for rapid mr imaging," *Magnetic resonance in medicine*, vol. 58, no. 6, 2007.

[3] H. Stark, *Image recovery: theory and application*, Elsevier, 1987.

[4] L. Baldassarre, J. Mourao-Miranda, and M. Pontil, "Structured sparsity models for brain decoding from fmri data," in *International Workshop on Pattern Recognition in NeuroImaging*, 2012.

[5] Q. Tran-Dinh and V. Cevher, "A primal-dual algorithmic framework for constrained convex minimization," *Tech. Report., LIONS*, 2014.

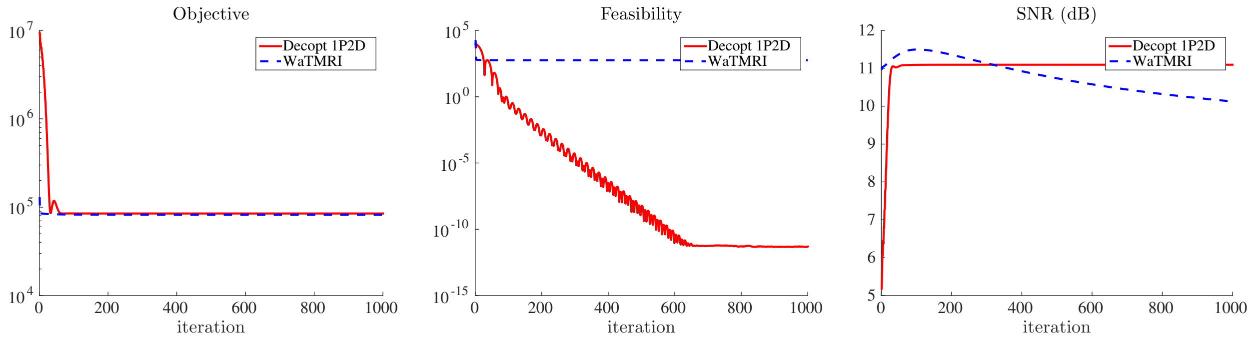

Fig. 1. MRI experiment. (a) Objective function. (b) Feasibility gap $\|\mathbf{z} - \mathbf{GWx}\|$. (c) Signal-to-noise ratio of the iterates. 1000 iterations took 244.12 seconds for Decopt and 325.74 seconds for WatMRI on a 2.6 GHz OS X system with 16 GB RAM.

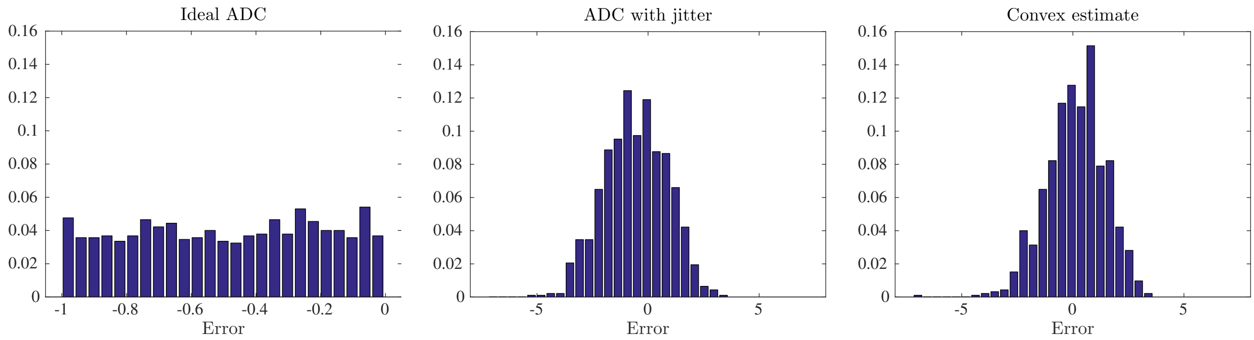

Fig. 2. ADC simulation. Errors distributions of the ideal ADC, of the ADC with jitter noise and of the convex estimate obtained via (16) from the samples obtained by the jittery ADC.

[6] Q. Tran-Dinh and V. Cevher, "Constrained convex minimization via model-based excessive gap," in *Neural Information Processing Systems conference*, 2014.

[7] G. Chen and M. Teboulle, "A proximal-based decomposition method for convex minimization problems," *Mathematical Programming*, vol. 64, no. 1-3, 1994.

[8] R Tyrrell Rockafellar, "Monotone operators and the proximal point algorithm," *SIAM journal on control and optimization*, vol. 14, no. 5, pp. 877–898, 1976.

[9] P. L Combettes and J.-C. Pesquet, "A proximal decomposition method for solving convex variational inverse problems," *Inverse problems*, vol. 24, no. 6, 2008.

[10] Y. Nesterov, "A method of solving a convex programming problem with convergence rate $\mathcal{O}(1/k^2)$," in *Soviet Mathematics Doklady*, 1983, vol. 27.

[11] P.L. Combettes, D. Dũng, and B.C. Vũ, "Proximity for sums of composite functions," *Journal of Mathematical Analysis and Applications*, vol. 380, no. 2, 2011.

[12] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, 2011.

[13] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, 2013.

[14] L.I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, 1992.

[15] N.S. Rao, R.D. Nowak, S.J. Wright, and N.G. Kingsbury, "Convex approaches to model wavelet sparsity patterns," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE, 2011.

[16] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach, "Proximal methods for hierarchical sparse coding," *The Journal of Machine Learning Research*, vol. 12, 2011.

[17] L. Jacob, G. Obozinski, and J.P. Vert, "Group lasso with overlap and graph lasso," in *International Conference on Machine Learning*, 2009.

[18] J. Huang, S. Zhang, and D. Metaxas, "Efficient mr image reconstruction for compressed mr imaging," *Medical Image Analysis*, vol. 15, no. 5, 2011.

[19] C. Chen and J. Huang, "Compressive sensing mri with wavelet tree sparsity," in *Advances in neural information processing systems*, 2012.