EFFICIENT CONSTRAINED SIGNAL RECONSTRUCTION
BY RANDOMIZED EPIGRAPHICAL PROJECTION

Shunsuke Ono

Tokyo Institute of Technology

ABSTRACT

This paper proposes a randomized optimization framework for constrained signal reconstruction, where the word "constrained" implies that data-fidelity is imposed as a hard constraint instead of adding a data-fidelity term to an objective function to be minimized. Such formulation facilitates the selection of regularization terms and hyperparameters, but due to the non-separability of the data-fidelity constraint, it does not suit block-coordinate-wise randomization as is. To resolve this, we give another expression of the data-fidelity constraint via epigraphs, which enables to design a randomized solver based on a stochastic proximal algorithm with randomized epigraphical projection. Our method is very efficient especially when the problem involves non-structured large matrices. We apply our method to CT image reconstruction, where the advantage of our method over the deterministic counterpart is demonstrated.

Index Terms— signal reconstruction, constrained optimization, stochastic optimization, epigraphical projection

1. INTRODUCTION

Signal reconstruction from incomplete and/or degraded observation is a fundamental problem arising from various applications, such as medical imaging, microscopy, tomography, spectral imaging and computational photography. Such a problem is often reduced to a convex optimization problem that involves a regularization term, modeling some desirable properties on the signal of interest, and a data fidelity term, enforcing consistency with observed data.

Proximal splitting algorithms [1, 2] have played a central role in solving such problems. These methods, especially primal-dual splitting type algorithms [3–8], are efficient in the sense that they require only simple operations like matrix-vector multiplications and evaluation of proximity operators. However, even such a simple operation becomes computationally expensive in many applications. A typical case is computed tomography (CT), where a matrix representing the observation process is large and is not structured[1], resulting in large computational costs and memory requirements [9, 10].

Recently, *stochastic primal-dual splitting algorithms* have been intensively studied for stochastic optimization [11–14]. Roughly speaking, at each iteration, the algorithms activate only the operations associated with randomly chosen variables, so that the said costs are significantly reduced. Actually, several studies show the utility of such block-coodinate-wise randomization in image restoration [14–16].

Incidentally, the above studies aim at *unconstrained formulation*, i.e., minimizing the sum of a regularization and a data-fidelity term. On the other hand, *constrained formulation*, i.e., minimizing a regularization term subject to a hard constraint on data-fidelity has an important advantage over the unconstrained one in terms of facilitating the selection of regularization terms and hyperparameters, as addressed in [17–24]. However, such a data-fidelity constraint is *not separable* as is, i.e., it cannot be decomposed into block-coordinate-wise constraints, so that it does not suit randomized activation.

In this paper, we bridge the gap between the randomized nature of stochastic proximal methods and the non-separability of data-fidelity constraints by leveraging *epigraphical projection* [23, 25]. We focus on the ℓ_2 data-fidelity constraint, and introduce its equivalent expression via certain epigraphs, which enables us to deal with the constrained formulation by stochastic proximal splitting algorithms with *randomized* epigraphical projection. Specifically, we develop a randomized solver for the problem based on a stochastic primal-dual hybrid gradient algorithm [14]. The efficiency of our method is demonstrated on CT image reconstruction.

2. PRELIMINARIES

2.1. Proximal Tools

The *proximity operator* [26] of index $\gamma > 0$ of a proper lower semicontiuous convex function $f \in \Gamma_0(\mathbb{R}^N)$[2] is defined as

$$\text{prox}_{\gamma f} : \mathbb{R}^N \to \mathbb{R}^N : \mathbf{x} \mapsto \underset{\mathbf{y}}{\text{argmin}} \, f(\mathbf{y}) + \frac{1}{2\gamma}\|\mathbf{y} - \mathbf{x}\|^2.$$

---

[1]Here the word "structured" means that the matrix-vector multiplication can be computed efficiently via some operation. An instance of structured matrices is a uniform blur matrix, which can be diagonalized by FFT.

[2]The set of all proper lower semicontinuous convex functions on $\mathbb{R}^N$ is denoted by $\Gamma_0(\mathbb{R}^N)$.

The indicator function of a nonempty closed convex set C, denoted by ι_C, is defined as

$$\iota_C(\mathbf{x}) := \begin{cases} 0 & \text{if } \mathbf{x} \in C \\ \infty & \text{otherwise.} \end{cases}$$

Since the function returns ∞ when the input vector is outside of C, it acts exactly as the hard constraint represented by C in minimization. The proximity operator of ι_C is equivalent to the (metric) projection onto C, i.e.,

$$\operatorname{prox}_{\gamma \iota_C}(\mathbf{y}) = P_C(\mathbf{y}) := \underset{\mathbf{x} \in C}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{x}\|.$$

2.2. Stochastic Primal-Dual Hybrid Gradient Algorithm

A stochastic primal-dual hybrid gradient algorithm (SPDHG) [14] was proposed to optimize the following problem:

$$\min_{\mathbf{x} \in \mathbb{R}^N} \sum_{i=1}^{I} f_i(\mathbf{A}_i \mathbf{x}) + g(\mathbf{x}), \tag{1}$$

where f_i and g are proper lower semicontinuous convex functions, and $\mathbf{A}_i$ are bounded linear operators (matrices). Here we assume that the proximity operators of f_i and g are easy to compute.

SPDHG does not solve (1) directly but solves the saddle point problem reformulated from (1), given by

$$\min_{\mathbf{x} \in \mathbb{R}^N} \sup_{\mathbf{y}_i \in \mathbb{R}^{M_i}} \sum_{i=1}^{I} \langle \mathbf{A}_i \mathbf{x}, \mathbf{y}_i \rangle - f_i^*(\mathbf{y}_i) + g(\mathbf{x}), \tag{2}$$

where f_i^* are the convex conjugate of f_i. We note that the proximity operator of f_i^* can be computed via that of f_i as $\operatorname{prox}_{\gamma f_i^*}(\mathbf{x}) = \mathbf{x} - \gamma \operatorname{prox}_{\frac{1}{\gamma} f_i}(\frac{1}{\gamma}\mathbf{x})$ [27, Theorem 14.3(ii)].

Let $\mathfrak{S} \subset \{1, \dots, I\}$ be a random subset of the indices of the dual variables $\mathbf{y}_i$ in (2), and define $\mathbf{A} := (\mathbf{A}_1, \dots, \mathbf{A}_I)$, $\mathbf{y} := (\mathbf{y}_1, \dots, \mathbf{y}_I)$ and $\boldsymbol{\Lambda} := \operatorname{diag}(p_1^{-1}\mathbf{I}, \dots, p_I^{-1}\mathbf{I})$ with the parameters p_i being probabilities that an index is selected in each iteration. Then SPDHG is formalized as follows: for given $\mathbf{x}^0, \mathbf{y}^0, \tau > 0, \rho_i > 0$, and $\overline{\mathbf{y}}^{(0)} = \mathbf{y}^0$, iterate

$$\left| \begin{aligned} &\mathbf{x}^{(k+1)} = \operatorname{prox}_{\tau g}(\mathbf{x}^{(k)} - \tau \mathbf{A}^* \overline{\mathbf{y}}^{(k)}), \\ &\text{Select } \mathfrak{S}^{(k+1)} \subset \{1, \dots, I\} \\ &\mathbf{y}_i^{(k+1)} = \begin{cases} \operatorname{prox}_{\rho_i f_i^*}(\mathbf{y}_i^{(k)} + \rho_i \mathbf{A}_i \mathbf{x}^{(k+1)}) & \text{if } i \in \mathfrak{S}^{(k+1)} \\ \mathbf{y}_i^{(k)} & \text{otherwise} \end{cases} \\ &\overline{\mathbf{y}}^{(k+1)} = \mathbf{y}^{(k+1)} + \boldsymbol{\Lambda}(\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}) \end{aligned} \right. \tag{3}$$

We should note that the matrix-vector multiplication $\mathbf{A}^* \overline{\mathbf{y}}^{(k)}$ in (3) can be computed by using only the selected $\mathbf{A}_i$ and the previous dual variable (see [14, Remark 1 and 2] for details). This means that each iteration requires both $\mathbf{A}_i$ and $\mathbf{A}_i^*$ to be evaluated only for each selected index $i \in \mathfrak{S}^{(k+1)}$. With a mild condition on the stepsizes τ and ρ_i, the algorithm converges to an optimal solution of (2) almost surely in the sense of the Bregman distance (see [14, Theorem 4.3] for details).

3. PROPOSED METHOD

3.1. Problem Formulation

Consider the following data observation model:

$$\mathbf{v} = \boldsymbol{\Phi} \bar{\mathbf{u}} + \mathbf{n}, \tag{4}$$

where $\bar{\mathbf{u}} \in \mathbb{R}^N$ is a latent signal we wish to estimate, $\boldsymbol{\Phi} \in \mathbb{R}^{M \times N}$ represents an observation process, $\mathbf{n} \in \mathbb{R}^M$ is an additive white Gaussian noise, and $\mathbf{v} \in \mathbb{R}^M$ is observed data.

Based on the model in (4), we aim at the following form of constrained signal reconstruction:

$$\min_{\mathbf{u} \in \mathbb{R}^N} \sum_{j=1}^{J} \mathcal{R}_j(\boldsymbol{\Psi}_j \mathbf{u}) \text{ s.t. } \begin{cases} \|\boldsymbol{\Phi}\mathbf{u} - \mathbf{v}\|^2 \leq \bar{\varepsilon}, \\ \mathbf{u} \in [\underline{\mu}, \overline{\mu}]^N, \end{cases} \tag{5}$$

where $\mathcal{R}_j(\boldsymbol{\Psi}_j \cdot)$ are regularization terms with functions $\mathcal{R}_j \in \Gamma_0(\mathbb{R}^{P_j})$ and matrices $\boldsymbol{\Psi}_j \in \mathbb{R}^{P_j \times N}$, the first hard-constraint is ℓ_2 delity with the radius $\bar{\varepsilon} > 0$, and the second one is a range constraint. We assume that the proximity operators of $\mathcal{R}_j$ are available.

3.2. Reformulation via Epigraphs

Since the ℓ_2 data-fidelity constraint in (5) is not separable, we cannot directly solve the problem by stochastic algorithms with block-coordinate-wise randomization. To circumvent this, we give another expression of the constraint as follows:

$$\|\boldsymbol{\Phi}\mathbf{u} - \mathbf{v}\|^2 \leq \bar{\varepsilon} \Leftrightarrow \begin{cases} \|\boldsymbol{\Phi}_1 \mathbf{u} - \mathbf{v}_1\|^2 \leq \varepsilon_1, \\ \quad \vdots \\ \|\boldsymbol{\Phi}_L \mathbf{u} - \mathbf{v}_L\|^2 \leq \varepsilon_L, \\ \sum_{l=1}^{L} \varepsilon_l \leq \bar{\varepsilon}, \end{cases} \tag{6}$$

where $\varepsilon_l \in \mathbb{R}$ are additional variables, $(\boldsymbol{\Phi}_1^\top \cdots \boldsymbol{\Phi}_L^\top)^\top = \boldsymbol{\Phi}$ and $(\mathbf{v}_1^\top \cdots \mathbf{v}_L^\top)^\top = \mathbf{v}$. Let us define the epigraphs of $\mathbf{v}_l$-centered squared distance, denoted by $\mathcal{S}_l$, and a half space $\mathcal{V}$, as

$$\mathcal{S}_l := \{(\mathbf{x}, \eta) \in \mathbb{R}^{Q_l} \times \mathbb{R} | \|\mathbf{x} - \mathbf{v}_l\|^2 \leq \eta\} \tag{7}$$

$$\mathcal{V} := \{(\eta_1, \dots, \eta_L) \in \mathbb{R}^L | \sum_{l=1}^{L} \eta_l \leq \bar{\varepsilon}\}, \tag{8}$$

Then, with (6), (7), and (8), Problem (5) can be rewritten as

$$\min_{\mathbf{u}, \varepsilon_1, \dots, \varepsilon_I} \sum_{j=1}^{J} \mathcal{R}_j(\boldsymbol{\Psi}_j \mathbf{u}) \text{ s.t. } \begin{cases} (\boldsymbol{\Phi}_1 \mathbf{u}, \varepsilon_1) \in \mathcal{S}_1, \\ \quad \vdots \\ (\boldsymbol{\Phi}_L \mathbf{u}, \varepsilon_L) \in \mathcal{S}_L, \\ \mathbf{u} \in [\underline{\mu}, \overline{\mu}]^N, \\ (\varepsilon_1, \dots, \varepsilon_L) \in V. \end{cases} \tag{9}$$

By introducing the indicator functions of the hard constraints in (9), which are denoted by $\iota_{\mathcal{S}_l}$, $\iota_{[\underline{\mu}, \overline{\mu}]}$, and $\iota_{\mathcal{V}}$, we can further reformulate Problem (9) as

$$\min_{\mathbf{u}, \varepsilon_1, \dots, \varepsilon_L} \sum_{j=1}^{J} \mathcal{R}_j(\boldsymbol{\Psi}_j \mathbf{u}) + \sum_{l=1}^{L} \iota_{\mathcal{S}_l}(\boldsymbol{\Phi}_l \mathbf{u}, \varepsilon_l)$$
$$+ \iota_{[\underline{\mu}, \overline{\mu}]^N}(\mathbf{u}) + \iota_{\mathcal{V}}(\varepsilon_1, \dots, \varepsilon_L). \tag{10}$$

3.3. Algorithm

Now we define $\boldsymbol{\varepsilon} := (\varepsilon_1, \ldots, \varepsilon_L)$ and $\mathbf{x} := (\mathbf{u}, \boldsymbol{\varepsilon})$, let $\mathbf{e}_1, \ldots, \mathbf{e}_L$ be the canonical basis of $\mathbb{R}^L$, and set $I := J + L$,

$$f_i(\mathbf{A}_i\mathbf{x}) := \mathcal{R}_i(\boldsymbol{\Psi}_i\mathbf{u}) \ \text{ for } i = 1, \ldots, J,$$
$$f_i(\mathbf{A}_i\mathbf{x}) := \iota_{\mathcal{S}_{i-J}}(\boldsymbol{\Phi}_{i-J}\mathbf{u}, \mathbf{e}_{i-J}^\top\boldsymbol{\varepsilon}) \ \text{ for } i = J+1, \ldots, I,$$
$$g(\mathbf{x}) := \iota_{[\underline{\mu}, \overline{\mu}]^N}(\mathbf{u}) + \iota_{\mathcal{V}}(\boldsymbol{\varepsilon}).$$

Then (10) is reduced to (1), so that we can solve (10) by SPDHG. We describe the whole algorithm in Algorithm 1.

Remark 1 (Note on Algorithm 1).
(a) Our algorithm is designed to randomly pick up two indices respectively from the two index sets: one is $j \in \{1, \ldots, J\}$ with probability $1/J$, and the other is $l \in \{1, \ldots, L\}$ with probability $1/L$, so that both the j-th regularization term and the l-th data-fidelity epigraph are evaluated in each iteration.
(b) The matrix-vector multiplications required in each iteration are only for the selected indices j and l. This implies that the computational cost and memory requirement of Algorithm 1 is much less than deterministic algorithms designed for solving (5).
(c) We give a stepsize setting rule as follows:

$$\rho_\psi := \frac{\gamma}{\max_j \|\boldsymbol{\Psi}_j\|} \ \forall j \in \{1, \ldots, J\},$$
$$\rho_\phi := \frac{\gamma}{\max_l \|\boldsymbol{\Phi}_l\|} \ \forall l \in \{1, \ldots, L\},$$
$$\tau := \frac{\gamma}{\max\{J, L\} \max\{\max_j \|\boldsymbol{\Psi}_j\|, \max_l \|\boldsymbol{\Phi}_l\|\}},$$

where $0 < \gamma < 1$. Similar stepsize setting is adopted in [14] but our rule is simpler.

Remark 2 (Projection computations in Algorithm 1).
Since the proximity operator of the indicator function of a nonempty closed convex set C equals to the projection onto C, we need to compute $P_{[\underline{\mu}, \overline{\mu}]^N}$, $P_{\mathcal{V}}$, and $P_{\mathcal{S}_i}$ in our algorithm.
(a) The projection onto $[\underline{\mu}, \overline{\mu}]^N$ can be calculated by just pushing each entry of the input vector into $[\underline{\mu}, \overline{\mu}]$.
(b) The projection onto $\mathcal{V}$ is given by [28, (3.3-10)]:

$$P_{\mathcal{V}}(\boldsymbol{\varepsilon}) := \begin{cases} \boldsymbol{\varepsilon} & \text{if } \mathbf{1}^\top\boldsymbol{\varepsilon} \leq \bar{\varepsilon} \\ \boldsymbol{\varepsilon} + \frac{\bar{\varepsilon} - \mathbf{1}^\top\boldsymbol{\varepsilon}}{L}\mathbf{1} & \text{otherwise}, \end{cases}$$

where $\mathbf{1}$ is the all-one vector of size L.
(c) The projection onto $\mathcal{S}_l$ is given as follows.

Proposition 1 (Epigraphical projection of squared distance).
Let $\mathbf{z} \in \mathbb{R}^N$ and let $\mathcal{S} := \{(\mathbf{x}, \eta) \in \mathbb{R}^N \times \mathbb{R} | \|\mathbf{x} - \mathbf{z}\|^2 \leq \eta\}$. Then, for every $(\mathbf{y}, \zeta) \in \mathbb{R}^N \times \mathbb{R}$, by letting $d := \|\mathbf{y} - \mathbf{z}\|$, the projection onto $\mathcal{S}$ is given by

$$P_{\mathcal{S}}(\mathbf{y}, \zeta) = (\alpha\mathbf{y} + (1 - \alpha)\mathbf{z}, \max\{\alpha^2 d^2, \zeta\}), \quad (11)$$

where

$$\alpha = \begin{cases} 1 & d^2 \leq \zeta, \\ \frac{\beta}{d} & \text{otherwise}, \end{cases} \quad (12)$$

$$\beta = \left(\frac{d}{4} + \left(\frac{d^2}{16} - \left(\frac{\zeta}{3} - \frac{1}{6}\right)^3\right)^{\frac{1}{2}}\right)^{\frac{1}{3}} + \frac{\left(\frac{\zeta}{3} - \frac{1}{6}\right)}{\left(\frac{d}{4} + \left(\frac{d^2}{16} - \left(\frac{\zeta}{3} - \frac{1}{6}\right)^3\right)^{\frac{1}{2}}\right)^{\frac{1}{3}}}. \quad (13)$$

Proof sketch: Equation (11) can be easily obtained from [23, Proposition 4]. Using the same proposition, we can see that

$$\alpha = d^{-1} \text{prox}_{\frac{1}{2}(\max\{|\cdot|^2 - \zeta, 0\})^2}(d). \quad (14)$$

Clearly, $\alpha = 1$ when $d^2 < \zeta$. When $d^2 > \zeta$, by simple calculation, the solution of the proximity operator in (14) is reduced to the real solution of the following cubic equation:

$$2x^3 + (1 - 2\zeta)x - d = 0. \quad (15)$$

Finally, applying Cardano formula to (15) yields (13).[3]

Algorithm 1: Proposed algorithm for solving (5)

input : $\mathbf{u}^{(0)}, \boldsymbol{\varepsilon}^{(0)}, \mathbf{z}^{(0)}, \mathbf{w}^{(0)}, \mathbf{t}^{(0)}, \boldsymbol{\zeta}^{(0)}, \boldsymbol{\xi}^{(0)}$
initialize: $\overline{\mathbf{t}}^{(0)} = \mathbf{t}^{(0)}, \overline{\boldsymbol{\xi}}^{(0)} = \boldsymbol{\xi}^{(0)}$

1 **for** $k = 0, \ldots, K - 1$ **do**

2    $\mathbf{u}^{(k+1)} = P_{[\underline{\mu}, \overline{\mu}]^N}(\mathbf{u}^{(k)} - \tau\overline{\mathbf{t}}^{(k)})$;

3    $\boldsymbol{\varepsilon}^{(k+1)} = P_{\mathcal{V}}(\boldsymbol{\varepsilon}^{(k)} - \tau\overline{\boldsymbol{\xi}}^{(k)})$;

4   Select $j \in \{1, \ldots, J\}$ and $l \in \{1, \ldots, L\}$.;

5    $\tilde{\mathbf{z}}_j^{(k)} = \mathbf{z}_j^{(k)} + \rho_\psi\boldsymbol{\Psi}_j\mathbf{u}^{(k+1)}$;

6    $\mathbf{z}_j^{(k+1)} = \tilde{\mathbf{z}}_j^{(k)} - \rho_\psi\text{prox}_{\mathcal{R}_j/\rho_\psi}(\tilde{\mathbf{z}}_j^{(k)}/\rho_\psi)$;

7    $\hat{\mathbf{z}}_j^{(k)} = \boldsymbol{\Psi}_j^\top(\mathbf{z}_j^{(k+1)} - \mathbf{z}_j^{(k)})$;

8    $\tilde{\mathbf{w}}_l^{(k)} = \mathbf{w}_l^{(k)} + \rho_\phi\boldsymbol{\Phi}_l\mathbf{u}^{(k+1)}$;

9    $\tilde{\zeta}_l^{(k)} = \zeta_l^{(k)} + \rho_\phi\varepsilon_l^{(k+1)}$;

10    $(\mathbf{w}_l^{(k+1)}, \zeta_l^{(k+1)}) =$
     $(\tilde{\mathbf{w}}_l^{(k)}, \tilde{\zeta}_l^{(k)}) - \rho_\phi P_{\mathcal{S}_l}((\tilde{\mathbf{w}}_l^{(k)}, \tilde{\zeta}_l^{(k)})/\rho_\phi)$;

11    $\hat{\mathbf{w}}_l^{(k)} = \boldsymbol{\Phi}_l^\top(\mathbf{w}_l^{(k+1)} - \mathbf{w}_l^{(k)})$;

12    $\hat{\boldsymbol{\zeta}}_l^{(k)} = \mathbf{e}_l(\zeta_l^{(k+1)} - \zeta_l^{(k)})$;

13    $\mathbf{t}^{(k+1)} = \mathbf{t}^{(k)} + \hat{\mathbf{z}}_j^{(k)} + \hat{\mathbf{w}}_l^{(k)}$;

14    $\overline{\mathbf{t}}^{(k+1)} = \mathbf{t}^{(k+1)} + (1 + J)\hat{\mathbf{z}}_j^{(k)} + (1 + L)\hat{\mathbf{w}}_l^{(k)}$;

15    $\boldsymbol{\xi}^{(k+1)} = \boldsymbol{\xi}^{(k)} + \hat{\boldsymbol{\zeta}}_l^{(k)}$;

16    $\overline{\boldsymbol{\xi}}^{(k+1)} = \boldsymbol{\xi}^{(k+1)} + (1 + L)\hat{\boldsymbol{\zeta}}_l^{(k)}$

---

4. NUMERICAL EXPERIMENTS

We examined the performance of the proposed method by comparing it with the deterministic counterpart, i.e., the deterministic primal-dual hybrid gradient algorithm [3] on CT image reconstruction. All experiments were performed using MATLAB (R2017b), on a Windows 10 Pro laptop computer with an Intel Core i7 2.1 GHz processor and 16 GB of RAM.

For the original image $\bar{\mathbf{u}}$ in (4), we used a head CT scan image of size 128×128 ($N = 16384$) picked up from the CT dataset [31]. The matrix $\boldsymbol{\Phi}$ in (4) was set to a parallel beam projection (Radon transform) matrix with 60 projection angles. We would like to note that the nonzero entries of $\boldsymbol{\Phi}$

---

[3]One of the reviewers pointed out that the above result can also be proven by a combination of [29, Proposition 5.1] and [30, Example 3.8].
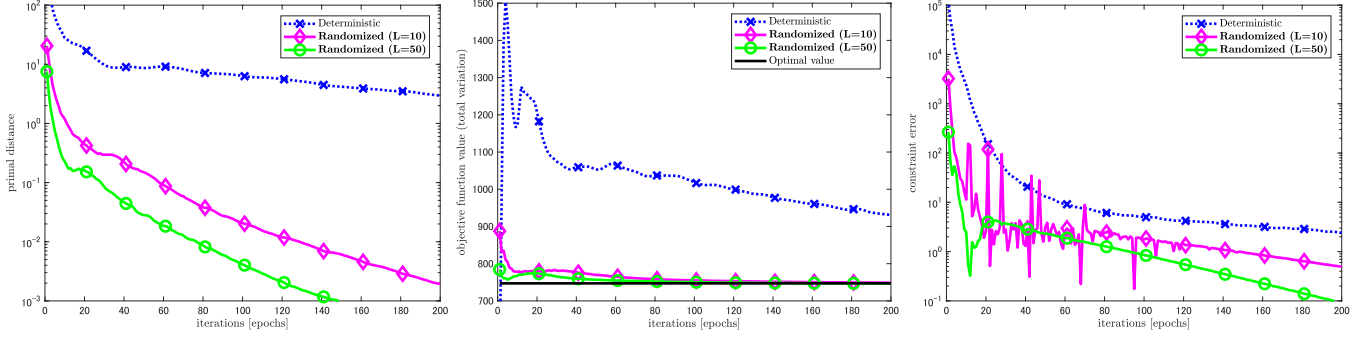
Fig. 1. Convergence profiles of Algorithm 1 ("Randomized") and its deterministic counterpart ("Deterministic") on CT image reconstruction in terms of the primal distance (left), objective function value (center), and constraint error (right). Note that the optimal value of the objective function (the black line in the center figure) was measured on $\mathbf{u}^{\star}$.
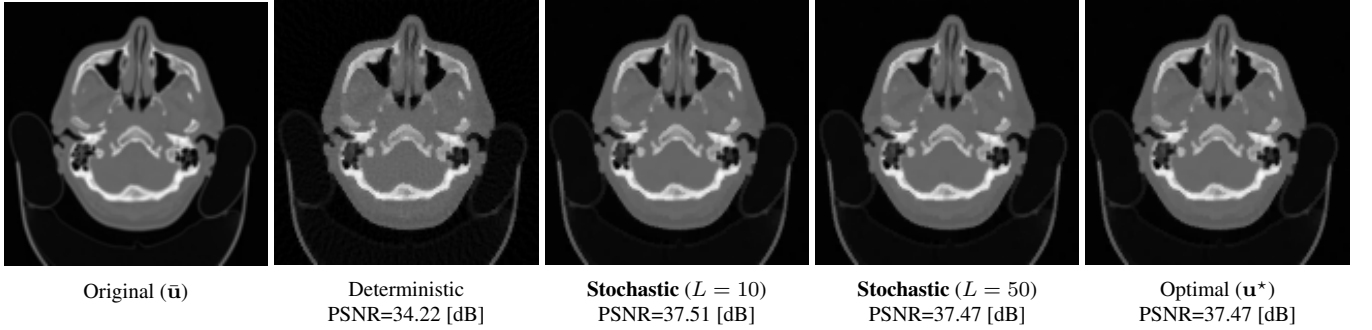

| Original ($\bar{\mathbf{u}}$) | Deterministic PSNR=34.22 [dB] | **Stochastic** ($L=10$) PSNR=37.51 [dB] | **Stochastic** ($L=50$) PSNR=37.47 [dB] | Optimal ($\mathbf{u}^{\star}$) PSNR=37.47 [dB] |

Fig. 2. Resulting images on CT image reconstruction (200 iterations [epochs]).

account for about 1.4% of all the entries, i.e., $\mathbf{\Phi}$ is sparse, so that this is advantageous for the deterministic algorithm, compared with the cases of dense $\mathbf{\Phi}$. We also note that we use a small image because the deterministic algorithm has to load full $\mathbf{\Phi}$ of size $M \times N$ ($M = 11100$, $N = 128^2$) in each iteration. The observed data was generated by adding white Gaussian noise with standard deviation $\sigma = 10/255$ to $\mathbf{\Phi}\bar{\mathbf{u}}$.

A full CT image was estimated by constrained total variation (TV) minimization, which is a special case of Problem (5). Specifically, we employed the anisotropic TV [32] for the regularization function $\mathcal{R}_j \circ \mathbf{\Psi}_j$ in Prob. (5). In this case, $J = 2$, and the matrix $\mathbf{\Psi}_1$ and $\mathbf{\Psi}_2$ are equal to the vertical and horizontal discrete gradient operators $\mathbf{D}_v$ and $\mathbf{D}_h$ with Neumann boundary, respectively. Both $\mathcal{R}_1$ and $\mathcal{R}_2$ are the ℓ_1 norm, and its proximity operator can be calculated by a simple $\mathcal{O}(N)$ soft-thresholding operation. We adopted an eight-bit dynamic range constraint $[0, 255]^N$ for $[\underline{\mu}, \overline{\mu}]^N$. For a fair comparison, the parameter $\bar{\varepsilon}$ was set to an oracle value, i.e., $\|\mathbf{\Phi}\bar{\mathbf{u}} - \mathbf{n}\|$. For the stepsizes of the deterministic algorithm, we employed the setting rule suggested in [3]. For our algorithm, see Remark 1(c) (Note: we choose $\gamma = 0.99$).

We adopted the following three convergence criteria:
(i) Primal distance: the squared distance between the current estimate $\mathbf{u}^{(k)}$ and an optimal solution $\mathbf{u}^{\star}$, i.e., $\|\mathbf{u}^{(n)} - \mathbf{u}^{\star}\|^2$. Since $\mathbf{u}^{\star}$ is analytically unavailable, it was pre-computed by the deterministic algorithm with 2×10^5 iterations.
(ii) Objective function value: the value of TV defined by $\|\mathbf{D}_v\mathbf{u}^{(k)}\|_1 + \|\mathbf{D}_h\mathbf{u}^{(k)}\|_1$.

(iii) Constraint error: the absolute difference between $\bar{\varepsilon}$ and $\|\mathbf{\Phi}\mathbf{u}^{(k)} - \mathbf{v}\|^2$. Since any optimal solution $\mathbf{u}^{\star}$ satisfies $\|\mathbf{\Phi}\mathbf{u}^{\star} - \mathbf{v}\|^2 = \bar{\varepsilon}$, this value should converge to zero.

The left of Fig. 1 shows the convergence of the primal distance, where "iterations [epochs]" means that the number of iterations is divided by L. We examined the cases of $L = 10, 50$ in this experiment. One sees that the proposed method ("Randomized") converges much faster than the deterministic counterpart ("Deterministic"). Similar convergence behavior can be observed in the center and right of Fig. 1, where the convergence profiles of the objective function value and the constraint error are plotted, respectively. The resulting images are depicted in Fig. 2, which illustrates that our algorithm properly works.

5. CONCLUSION

We have proposed an efficient constrained signal reconstruction framework based on a stochastic primal-dual splitting algorithm with randomized epigraphical projection. Since the proposed method does not require the multiplication of full $\mathbf{\Phi}$ and variables in each iteration, it would be a powerful choice when $\mathbf{\Phi}$ is large and not structured while keeping the benefits of the constrained formulation.

In this paper, we discuss only the ℓ_2 data-fidelity case but this framework can also be applied to other data fidelity constraints, for example, the ℓ_1 case, as long as their epigraphical projections are computable. Also, with a slight extension, our method would be able to handle signal decomposition models, such as image decomposition [33, 34].

6. REFERENCES

[1] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke *et al*, Ed., pp. 185–212. Springer-Verlag, 2011.

[2] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[3] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2010.

[4] P. L. Combettes and J.-C. Pesquet, "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators," *Set-Valued and Variational Analysis*, vol. 20, no. 2, pp. 307–330, 2012.

[5] L. Condat, "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *J. Opt. Theory Appl.*, vol. 158, no. 2, pp. 460–479, 2013.

[6] B. C. Vu, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Adv. Comput. Math.*, vol. 38, pp. 667–681, 2013.

[7] N. Komodakis and J.-C. Pesquet, "Playing with Duality: An overview of recent primal-dual approaches for solving large-scale optimization problems," *IEEE Signal Process. Magazine*, vol. 32, no. 6, pp. 31–54, 2015.

[8] S. Ono, "Primal-dual plug-and-play image restoration," *IEEE Signal Process. Lett.*, vol. 24, no. 8, pp. 1108–1112, 2017.

[9] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in Medicine & Biology*, vol. 53, no. 17, pp. 4777, 2008.

[10] X. Pan, E. Y. Sidky, and M. Vannier, "Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?," *Inverse problems*, vol. 25, no. 12, pp. 123009, 2009.

[11] J.-C. Pesquet and A. Repetti, "A class of randomized primal-dual algorithms for distributed optimization," *J. Nonlin. Convex Anal.*, vol. 16, no. 12, pp. 2453–2490, 2015.

[12] Y. Zhang and L. Xiao, "Stochastic primal-dual coordinate method for regularized empirical risk minimization," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 2939–2980, 2017.

[13] P. L. Combettes and J. Eckstein, "Asynchronous block-iterative primal-dual decomposition methods for monotone inclusions," *Math. Program.*, vol. 168, no. 1-2, pp. 645–672, 2018.

[14] A. Chambolle, M. J. Ehrhardt, P. Richtárik, and C.-B. Schonlieb, "Stochastic primal-dual hybrid gradient algorithm with arbitrary sampling and imaging applications," *SIAM J. Optim.*, vol. 28, no. 4, pp. 2783–2808, 2018.

[15] S. Ono, M. Yamagishi, T. Miyata, and I. Kumazawa, "Image restoration using a stochastic variant of the alternating direction method of multipliers," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2016, pp. 4523–4527.

[16] P. L. Combettes and J.-C. Pesquet, "Stochastic forward-backward and primal-dual approximation algorithms with application to online image restoration," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Aug 2016, pp. 1813–1817.

[17] D. C. Youla and H. Webb, "Image restoration by the method of convex projections: Part 1–theory," *IEEE Transactions on Medical Imaging*, vol. 1, no. 2, pp. 81–94, 1982.

[18] P. L. Combettes, "Inconsistent signal feasibility problems: Least-squares solutions in a product space," *IEEE Trans. Signal Process.*, vol. 42, no. 11, pp. 2955–2966, 1994.

[19] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, 2011.

[20] R. Stück, M. Burger, and T. Hohage, "The iteratively regularized gauss–newton method with convex constraints and applications in 4Pi microscopy," *Inverse Problems*, vol. 28, no. 1, pp. 015012, 2011.

[21] M. Carlavan and L. Blanc-Féraud, "Sparse Poisson noisy image deblurring," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1834–1846, 2012.

[22] T. Teuber, G. Steidl, and R. H. Chan, "Minimization and parameter estimation for seminorm regularization models with I-divergence constraints," *Inverse Problems*, vol. 29, no. 3, pp. 035007, 2013.

[23] G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu, "Epigraphical projection and proximal tools for solving constrained convex optimization problems," *Signal, Image and Video Process.*, vol. 9, no. 8, pp. 1737–1749, 2015.

[24] S. Ono and I. Yamada, "Signal recovery with certain involved convex data-fidelity constraints," *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 6149–6163, 2015.

[25] S. Ono and I. Yamada, "Second-order total generalized variation constraint," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2014, pp. 4938–4942.

[26] J. J. Moreau, "Fonctions convexes duales et points proximaux dans un espace hilbertien," *C. R. Acad. Sci. Paris Ser. A Math.*, vol. 255, pp. 2897–2899, 1962.

[27] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, Springer, New York, 2011.

[28] H. Stark and Y. Yang, *Vector Space Projections, a Numerical Approach to Signal and Image Processing, Neural Nets and Optics.*, Wiley Series in Telecommunications and Signal Processing. John Wiley & Sons, Inc., 1998.

[29] M. El Gheche, G. Chierchia, and J.-C. Pesquet, "Proximity operators of discrete information divergences," *IEEE Trans. Inform. Theory*, vol. 64, no. 2, pp. 1092–1104, 2018.

[30] P. L. Combettes and C. L. Mller, "Perspective functions: Proximal calculus and applications in high-dimensional statistics," *J. Math. Anal. Appl.*, vol. 457, no. 2, pp. 1283–1306, 2018.

[31] R. Chilamkurthy, S.and Ghosh, S. Tanamala, M. Biviji, N. G. Campeau, V.K. Venugopal, V. Mahajan, P. Rao, and P. Warier, "Development and validation of deep learning algorithms for detection of critical findings in head CT scans," *arXiv preprint arXiv:1803.05854*, 2018.

[32] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, no. 1-4, pp. 259–268, 1992.

[33] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher, "Structure-texture image decomposition - modeling, algorithms, and parameter selection," *Int. J. Comput. Vis.*, vol. 67, no. 1, pp. 111–136, 2006.

[34] S. Ono, T. Miyata, and I. Yamada, "Cartoon-texture image decomposition using blockwise low-rank texture characterization," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1128–1142, 2014.