Operator-Norm-Based Variable-Wise Diagonal Preconditioning for Automatic Stepsize Selection of A Primal-Dual Splitting Algorithm

1st Kazuki Naganuma
School of Computing, Department of Computer Science
Tokyo Institute of Technology
Kanagawa, Japan
naganuma.k.aa@m.titech.ac.jp

2nd Shunsuke Ono
School of Computing, Department of Computer Science
Tokyo Institute of Technology
Kanagawa, Japan
ono@c.titech.ac.jp

***Abstract*—We propose a diagonal preconditioning method for automatically selecting the step sizes of a primal-dual splitting method (PDS). The conventional preconditioning method for PDS has several limitations, such as the need to directly access all the entries of the matrices representing the linear operators in the target optimization problem, and the possibility that the proximity operator cannot be solved analytically due to the element-wise preconditioning. In this paper, we establish operator norm-based variable-wise diagonal preconditioning (ON-VW) to resolve these issues. ON-VW has two features that are preferred in real applications. First, the preconditioners constructed by ON-VW are defined using only (an upper bound of) the operator norm of the linear operators, which eliminates the need for their explicit matrix representations. Furthermore, the stepsizes automatically selected by our preconditioners are variable-wise, which allows us to keep the proximity operator computable. We also prove that our preconditioners satisfy the convergence condition of PDS and demonstrate its effectiveness through its application to denoising of hyperspectral images.**

***Index Terms*—Primal-dual splitting method (PDS), diagonal preconditioning, automatic stepsize selection, signal estimation**

I. Introduction

Many signal estimation problems, such as denoising, interpolation, decomposition and reconstruction, have been resolved by casting them as convex optimization problems [1], [2] of the form:

$$\min_{\substack{\mathbf{x}_1,\dots,\mathbf{x}_N,\\ \mathbf{y}_1,\dots,\mathbf{y}_M}} \sum\nolimits_{i=1}^{N} f_i(\mathbf{x}_i) + \sum\nolimits_{j=1}^{M} g_j\left(\mathbf{y}_j\right)$$
$$\text{s.t. } \mathbf{y}_1 = \textstyle\sum_{i=1}^{N} \mathfrak{L}_{1,i}(\mathbf{x}_i), \dots, \mathbf{y}_M = \sum_{i=1}^{N} \mathfrak{L}_{M,i}(\mathbf{x}_i), \quad (1)$$

where $f_i : \mathbb{R}^{n_i} \to (-\infty, +\infty]$ and $g_j : \mathbb{R}^{m_j} \to (-\infty, +\infty]$ are proximable[1]proper lower semicontinuous convex functions, and $\mathfrak{L}_{j,i} : \mathbb{R}^{n_i} \to \mathbb{R}^{m_j}$ are linear operators ($\forall i = 1, \dots, N$ and $\forall j = 1, \dots, M$,). The variables $\mathbf{x}_1, \dots, \mathbf{x}_N$ represent estimated signals or components, and $\mathbf{y}_1, \dots, \mathbf{y}_M$ are auxiliary variables for splitting.

This work was supported in part by JST CREST under Grant JPMJCR1662 and JPMJCR1666, in part by JST PRESTO under Grant JPMJPR21C4, and in part by JSPS KAKENHI under Grant 20H02145, 19H04135 and 18H05413.

[1]If an efficient computation of the proximity operator (see. Eq. (3)) of f is available, we call f proximable.

As a method for solving Prob. (1), a primal-dual splitting method (PDS) [3] has attracted attention due to its simple implementation without operator inversions[2]. The theoretical convergence of PDS is established in a primal-dual space equipped with a skewed metric, which is determined by the linear operators involved in the optimization problem and stepsizes (see [4], [6] for details). The convergence speed of PDS thus strongly depends both on the problem structure and the stepsize selection, leading to a problem-wise manual adjustment of the stepsizes for fast convergence.

To avoid such troublesome stepsize adjustment and improve the convergence, a preconditioned version of PDS (P-PDS) has been proposed for automatic stepsize selection based on certain diagonal preconditioners [7]. The entries of the diagonal preconditioners, corresponding to element-wise stepsizes, consist of the row/column absolute sum of the entries of the explicit matrices representing the $\mathfrak{L}_{j,i}$, and thus the resulting stepsizes can be different for each element in one variable. Such element-wise preconditioning has been proposed for various proximal splitting methods [8]–[10].

Although this preconditioning method is powerful and useful, there exist two limitations that are considerable in real applications. First, the method is difficult to apply in the case where (some of) the linear operators $\mathfrak{L}_{j,i}$ in Prob. (1) are not implemented as explicit matrices because it requires to access the whole entries of the matrices to construct the preconditioners. We often encounter such situations especially in imaging applications, where the linear operators are implemented not as explicit matrices but as procedures that compute forward and adjoint operations in an efficient manner, e.g., difference operators [11], [12] and frame transformations [13]–[15]. Second, some proximable functions are not completely separable for each element of the input variable, e.g., mixed norms and the indicator functions of norm balls [16]. For such functions, element-wise preconditioning might make the functions non-proximable.

[2]This algorithm has been generalized by Condat [4] and Vu [5], where smooth convex functions are optimized by using their Lipschitzian gradients.

For addressing the above issues, this paper proposes an operator norm-based variable-wise diagonal preconditioning (ON-VW) method. Our method has two features preferred in a number of real applications. First, our preconditioners can be computed from (upper bounds of) the operator norms of the linear operators $\mathfrak{L}_{j,i}$, meaning that our method does not need their explicit matrix representations. This is because (upper bounds of) the operator norms are often known or can be estimated without matrix implementation for typical linear operators used in signal estimation, including the ones mentioned above. Second, since the ith set of linear operators $\mathfrak{L}_{1,i}, \ldots, \mathfrak{L}_{M,i}$ exactly corresponds to the ith variable $\mathbf{x}_i$, the entries of the diagonal preconditioner for the ith variable obtained by our method take the same value, resulting in variable-wise stepsizes. This maintains the proximablity of the functions in the target optimization problem. We also prove that the sequence generated by P-PDS with ON-VW converges the solution of Prob. (1), and demonstrate the effectiveness of ON-VW through a hyperspectral image denoising problem.

II. Preliminaries

A. Proximity Operator

Let $f : \mathbb{R}^N \to (-\infty, \infty]$ be a proximable proper lower semi-continuous convex function and $\mathbf{G} \in \mathbb{R}^{N\times N}$ be a symmetric and positive definite matrix. Then, the skewed proximity operator is defined as

$$\mathrm{prox}_{\mathbf{G},f}(\mathbf{x}) := \underset{\mathbf{y}}{\mathrm{argmin}} \frac{1}{2}\langle \mathbf{x}-\mathbf{y}, \mathbf{G}(\mathbf{x}-\mathbf{y})\rangle + f(\mathbf{y}), \quad (2)$$

where $\langle\cdot,\cdot\rangle$ is the Euclidean inner product. If $\mathbf{G}$ is a positive scalar matrix, i.e., $\mathbf{G} = \alpha\mathbf{I}$ ($\alpha > 0$), the skewed proximity operator is identical to the standard proximity operator:

$$\mathrm{prox}_{\alpha\mathbf{I},f}(\mathbf{x}) = \mathrm{prox}_{\frac{1}{\alpha}f}(\mathbf{x}) = \underset{\mathbf{y}}{\mathrm{argmin}} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \frac{1}{\alpha}f(\mathbf{y}). \quad (3)$$

The standard proximity operators of some popular convex functions, such as the mixed $\ell_{1,2}$-norm, have analytic solutions, but their computation is not completely separable element by element. In such cases, even if the preconditioner is diagonal (with different entries), the computation of the skewed proximity operator becomes difficult.

B. Preconditioned Version of PDS

P-PDS [7] computes the solution of Prob. (1) by Algorithm 1. The function f^* is the *Fenchel–Rockafellar conjugate function*[3] of f, $\mathfrak{L}^*$ is the adjoint operator of $\mathfrak{L}$, and $\mathbf{\Gamma}_{1,i}$ and $\mathbf{\Gamma}_{2,j}$ are symmetric and positive definite matricies called *preconditioners*.

Here, we introduce the convergence property of P-PDS.

[3]The Fenchel–Rockafellar conjugate function of f is defined as

$$f^*(\mathbf{x}) := \max_{\mathbf{y}} \langle \mathbf{x}, \mathbf{y}\rangle + f(\mathbf{y}).$$

Algorithm 1 P-PDS [7] for solving Prob. (1)

Input: $x_1^{(0)}, \ldots, x_N^{(0)}, y_1^{(0)}, \ldots, y_M^{(0)}$
$\Gamma_{1,1}, \ldots, \Gamma_{1,N}, \Gamma_{2,1}, \ldots, \Gamma_{2,M}$
Output: $x_1^{(k)}, \ldots, x_N^{(k)}, y_1^{(k)}, \ldots, y_M^{(k)}$

```
Initialize n = 0;
while A stopping criterion is not satisfied do
    for i = 1, ..., N do
        x'_i ← Σ_{j=1}^{M} 𝔏*_{j,i}(y_j^{(k)})
        x_i^{(k+1)} ← prox_{Γ_{1,i}^{-1}, f_i}(x^{(k)} − Γ_{1,i} x'_i);
    end for
    for j = 1, ..., M do
        y'_i ← Σ_{j=1}^{N} 𝔏_{j,i}(2x_i^{(k+1)} − x_i^{(k)})
        y_j^{(k+1)} ← prox_{Γ_{2,j}^{-1}, g_j^*}(y_j^{(k)} + Γ_{2,j} y'_i)
    end for
    k ← k + 1;
end while
```

Theorem II.1. *[7, Theorem 1] Let* $\mathbf{\Gamma}_1 = \mathrm{diag}(\mathbf{\Gamma}_{1,1}, \ldots, \mathbf{\Gamma}_{1,N})$ *and* $\mathbf{\Gamma}_2 = \mathrm{diag}(\mathbf{\Gamma}_{2,1}, \ldots, \mathbf{\Gamma}_{2,M})$ *are symmetric and positive definite matrices satisfying*

$$\left\| \mathbf{\Gamma}_2^{\frac{1}{2}} \circ \mathfrak{L} \circ \mathbf{\Gamma}_1^{\frac{1}{2}} \right\|_{op}^2 \leq 1, \quad (4)$$

where $\mathfrak{A} \circ \mathfrak{B}$ *is the composition of* $\mathfrak{A}$ *and* $\mathfrak{B}$, $\|\cdot\|_{op}$ *is the operator norm and* $\mathfrak{L}$ *is the linear operator that includes* $\mathfrak{L}_{j,i}$*:*

$$\mathfrak{L} := \begin{bmatrix} \mathfrak{L}_{1,1} & \mathfrak{L}_{1,2} & \cdots & \mathfrak{L}_{1,N} \\ \mathfrak{L}_{2,1} & \mathfrak{L}_{2,2} & \cdots & \mathfrak{L}_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathfrak{L}_{M,1} & \mathfrak{L}_{M,2} & \cdots & \mathfrak{L}_{M,N} \end{bmatrix}. \quad (5)$$

Then, the sequence $(\mathbf{x}_1^{(n)}, \ldots, \mathbf{x}_N^{(n)}, \mathbf{y}_1^{(n)}, \ldots, \mathbf{y}_M^{(n)})$ *generated by Algorithm 1 converges to an optimal solution* $(\mathbf{x}_1^*, \ldots, \mathbf{x}_N^*, \mathbf{y}_1^*, \ldots, \mathbf{y}_M^*)$ *of Prob.* (1)*.*

The authors of [7] present a concrete method for constructing the preconditioners as follows. Each $\Gamma_{1,i}$ (and $\Gamma_{2,j}$) is a diagonal matrix consisting of the row/column absolute sum of the entries of the matrix representing $\mathfrak{L}_{j,i}$ (see Lemma 2 in [7]). This means that the diagonal entries of one $\Gamma_{1,i}$ (and $\Gamma_{2,j}$) may take different values, and the stepsizes of P-PDS will be different for each element for one variable in (1).

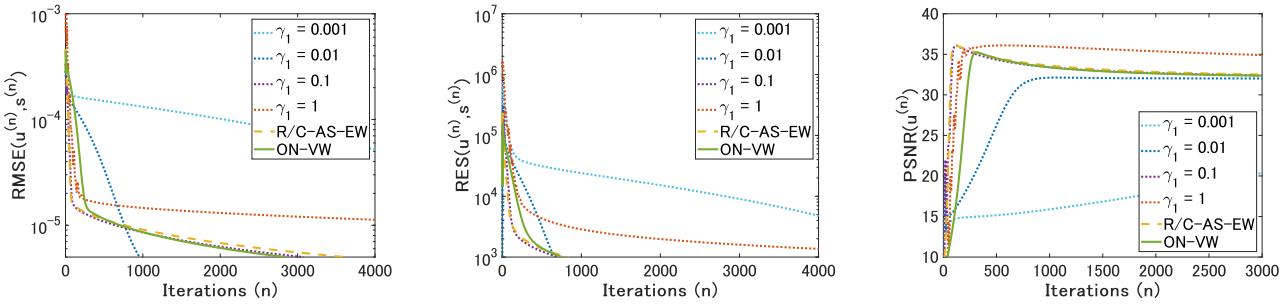
The standard PDS can be recovered by setting the preconditioners as $\mathbf{\Gamma}_{1,1} = \cdots = \mathbf{\Gamma}_{1,N} = \gamma_1\mathbf{I}$ and $\mathbf{\Gamma}_{2,1} = \cdots = \mathbf{\Gamma}_{2,M} = \gamma_2\mathbf{I}$ for positive scalars γ_1 and γ_2 that satisfy (4), that is,

$$\gamma_1\gamma_2\|\mathfrak{L}\|_{\mathrm{op}} \leq 1. \quad (6)$$

III. Proposed Oprator-Norm-Based Variable-Wise Diagonal Preconditioning (ON-VW)

Here we propose a novel diagonal preconditioning method, ON-VW, for P-PDS. Specifically, ON-VW constructs preconditioners given by

$$\mathbf{\Gamma}_{1,i} = \frac{1}{\sum_{j=1}^M \gamma_{j,i}^2}\mathbf{I}, \quad \mathbf{\Gamma}_{2,j} = \frac{1}{N}\mathbf{I}, \quad (7)$$



(a) (b) (c)

Fig. 1. The comparison of PDS stepsizes (19), R/C-AS-EW [7], and ON-VW (Ours). (a): NRMSE versus iteration. (b): RES versus iteration. (c): PSNR versus iteration. Note that applying P-PDS with R/C-AS-EW (yellow dotted line) to Prob. (15) is not practical in terms of implementation. (the linear operators $\mathfrak{D}_v$, $\mathfrak{D}_h$, and $\mathfrak{D}_b$ are not usually implemented as explicit matrices)

where $\gamma_{j,i}$ ($\forall i = 1,\dots,N$ and $\forall j = 1,\dots,M$) are some upper bound of the operator norm of $\mathfrak{L}_{j,i}$, i.e.,

$$\gamma_{j,i} \in [\|\mathfrak{L}_{j,i}\|_{\mathrm{op}}, \infty). \tag{8}$$

Clearly, our preconditioners can be constructed by only using (upper bounds of) the operator norms of the linear operators $\mathfrak{L}_{j,i}$, implying that ON-VW does not require to directly access the entires of the explicit matrices representing $\mathfrak{L}_{j,i}$ as long as some $\gamma_{i,j}$ are available. In addition, the diagonal entries of one $\Gamma_{1,i}$ take the same value ($\Gamma_{2,j}$ as well), i.e., our preconditioners are in variable-wise, which does not violate proximability of the functions in (1). For the preconditioners defined in (7), the following theorem holds.

Theorem III.1. *If the preconditioners are set as Eq.* (7)*, then the inequality in* (4) *is satisfied.*

Proof. Since $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ are diagonal, their powers of one half are

$$\begin{aligned}\mathbf{\Gamma}_1^{\frac{1}{2}} &= \mathrm{diag}(\mathbf{\Gamma}_{1,1}^{\frac{1}{2}},\dots,\mathbf{\Gamma}_{1,N}^{\frac{1}{2}}),\\ \mathbf{\Gamma}_2^{\frac{1}{2}} &= \mathrm{diag}(\mathbf{\Gamma}_{2,1}^{\frac{1}{2}}\dots,\mathbf{\Gamma}_{2,M}^{\frac{1}{2}}).\end{aligned} \tag{9}$$

By matrix multiplication and Eq. (9), we have

$$\mathbf{\Gamma}_2^{\frac{1}{2}} \circ \mathfrak{L} \circ \mathbf{\Gamma}_1^{\frac{1}{2}} = [\mathbf{\Gamma}_{2,j}^{\frac{1}{2}} \circ \mathfrak{L}_{j,i} \circ \mathbf{\Gamma}_{1,i}^{\frac{1}{2}}], \begin{cases} i = 1,\dots,N, \\ j = 1,\dots,M. \end{cases} \tag{10}$$

Using the inequality of the operator norm of the block matrix [17] and Eq. (7), we obtain

$$\begin{aligned}\left\|\mathbf{\Gamma}_2^{\frac{1}{2}} \circ \mathfrak{L} \circ \mathbf{\Gamma}_1^{\frac{1}{2}}\right\|_{\mathrm{op}}^2 &\le \sum_{i=1}^{N}\sum_{j=1}^{M}\left\|\mathbf{\Gamma}_{2,j}^{\frac{1}{2}} \circ \mathfrak{L}_{j,i} \circ \mathbf{\Gamma}_{1,i}^{\frac{1}{2}}\right\|_{\mathrm{op}}^2\\ &= \sum_{i=1}^{N}\frac{1}{N}\frac{\sum_{j=1}^{M}\|\mathfrak{L}_{j,i}\|_{\mathrm{op}}^2}{\sum_{j=1}^{M}\gamma_{j,i}^2}.\end{aligned} \tag{11}$$

By using (8), it is satisfied that $\sum_{j=1}^{M}\|\mathfrak{L}_{j,i}\|_{\mathrm{op}}^2 \le \sum_{j=1}^{M}\gamma_{j,i}^2$ for any $i = 1,\dots,N$ and $j = 1,\dots,M$. Applying this inequality to Eq. (11), we obtain

$$\left\|\mathbf{\Gamma}_2^{\frac{1}{2}} \circ \mathfrak{L} \circ \mathbf{\Gamma}_1^{\frac{1}{2}}\right\|_{\mathrm{op}}^2 \le \sum_{i=1}^{N}\frac{1}{N} = 1. \tag{12}$$

□

Theorem III.1 asserts that the preconditioners defined in (7) satisfy the convergence condition in (4) of Algorithm 1. Therefore, Algorithm 1 with our preconditioners in (7) generates sequences that converge an optimal solution of Prob. (1). Here, it is better to set the smallest possible values to $\gamma_{j,i}$ in (7) since the convergence is faster as the stepsize is larger. Therefore, $\gamma_{j,i}$ is determined in the following manner.

- If the operator norm $\|\mathfrak{L}_{j,i}\|_{\mathrm{op}}$ is known, we set $\gamma_{j,i}$ to $\|\mathfrak{L}_{j,i}\|_{\mathrm{op}}$.
- If $\|\mathfrak{L}_{j,i}\|_{\mathrm{op}}$ is unknown, we set $\gamma_{j,i}$ to an upper bound of $\|\mathfrak{L}_{j,i}\|_{\mathrm{op}}$.
- If the linear operator is the composition function of two linear operators $\mathfrak{A}$ and $\mathfrak{B}$ whose operator norms (or their upper bounds) are known ($\|\mathfrak{A}\|_{op} \le \alpha_{\mathfrak{A}}, \|\mathfrak{B}\|_{op} \le \alpha_{\mathfrak{B}}$), we set $\gamma_{j,i}$ to $\alpha_{\mathfrak{A}}\alpha_{\mathfrak{B}}$.[4]

IV. Application to Hyperspectral Image Denoising

In this section, we apply the proposed ON-VW to a hyperspectral image denoising problem with the spatio-spectral total variation regularization (SSTV) [18], which has attracted attention as hyperspectral image regularization [19]–[22].

A. Problem Formulation

Consider that an observed hyperspectral image (of size $n_1 \times n_2 \times n_3$) $\mathbf{v} \in \mathbb{R}^{n_1 n_2 n_3}$ is modeled by

$$\mathbf{v} = \bar{\mathbf{u}} + \bar{\mathbf{s}} + \mathbf{n}, \tag{13}$$

where $\bar{\mathbf{u}}, \bar{\mathbf{s}}$, and $\mathbf{n}$ are the true hyperspectral image of interest, sparsely distributed noise such as outliers, and random noise, respectively. Based on this observation model, the SSTV regularized-denoising problem is formulated as a convex optimization problem with the following form:

$$\begin{aligned}\min_{\mathbf{u},\mathbf{s}} &\|\mathfrak{D}_v(\mathfrak{D}_b(\mathbf{u}))\|_1 + \|\mathfrak{D}_h(\mathfrak{D}_b(\mathbf{u}))\|_1 + \lambda\|\mathbf{s}\|_1\\ &\text{s.t.} \quad \mathbf{u} + \mathbf{s} \in B_{(\mathbf{v},\varepsilon)},\end{aligned} \tag{14}$$

[4]It is satisfied that $\|\mathfrak{A} \circ \mathfrak{B}\|_{\mathrm{op}} \le \|\mathfrak{A}\|_{\mathrm{op}}\|\mathfrak{B}\|_{\mathrm{op}}$ by the submultiplicity of the operator norm.

(a) (b) (c) PSNR=31.48 [dB] (d) PSNR=33.97 [dB] (e) PSNR=33.93 [dB]

Fig. 2. Denoising results. (a): Groundtruth. (b): Observed image. (c): Image estimated by PDS with adjusted stepsizes ($\gamma_1 = 0.01$). (d): Image estimated by P-PDS with R/C-AS-EW [7]. (e): Image estimated by P-PDS with ON-VW (Ours).

where $\mathfrak{D}_v$, $\mathfrak{D}_h$, and $\mathfrak{D}_b$ are the vertical, horizontal, and spectral difference operators, respectively. Here, $\|\cdot\|_1$ is the ℓ_1 norm, and $B_{(\mathbf{v},\varepsilon)}$ is the ℓ_2 norm ball that is defined as

$$B_{(\mathbf{v},\varepsilon)} := \{\mathbf{x} \in \mathbb{R}^{n_1 n_2 n_3} | \|\mathbf{v} - \mathbf{x}\|_2 \leq \varepsilon\}. \tag{15}$$

The terms $\|\mathfrak{D}_v(\mathfrak{D}_b(\mathbf{u}))\|_1 + \|\mathfrak{D}_h(\mathfrak{D}_b(\mathbf{u}))\|_1$ is the SSTV regularization. The positive value λ is a balancing parameter between the SSTV regularization and the sparse noise term. The hard constraint guarantees the ℓ_2 data fidelity to $\mathbf{v}$ with the radius ε.[5]

By using the indicator function[6] of $B_{(\mathbf{v},\varepsilon)}$, Prob. (14) is reduced to Prob. (1) through the following reformulation:

$$\begin{aligned} &\min_{\substack{\mathbf{u},\mathbf{s},\\ \mathbf{y}_1,\mathbf{y}_2,\mathbf{y}_3}} \lambda\|\mathbf{s}\|_1 + \|\mathbf{y}_1\|_1 + \|\mathbf{y}_2\|_1 + \iota_{\mathcal{B}_{\mathbf{v},\varepsilon}}(\mathbf{y}_3) \\ &\text{s.t. } \mathbf{y}_1 = \mathfrak{D}_v(\mathfrak{D}_b(\mathbf{u})), \mathbf{y}_2 = \mathfrak{D}_h(\mathfrak{D}_b(\mathbf{u})), \mathbf{y}_3 = \mathbf{u} + \mathbf{s}. \end{aligned} \tag{16}$$

By applying Algorithm 1 to Prob. (16), we can compute the solution of Prob. (14). Here, since it is satisfied that $\|\mathfrak{D}_v \circ \mathfrak{D}_b\|_{\text{op}} \leq 4$, $\|\mathfrak{D}_h \circ \mathfrak{D}_b\|_{\text{op}} \leq 4$ [7], and $\|\mathbf{I}\|_{\text{op}} = 1$, the preconditioners by ON-VW are derived as follows:

$$\begin{aligned} &\mathbf{\Gamma}_{1,1} = \frac{1}{4^2 + 4^2 + 1^2}\mathbf{I} = \frac{1}{33}\mathbf{I}, \mathbf{\Gamma}_{1,2} = \frac{1}{1^2}\mathbf{I} = \mathbf{I}, \\ &\mathbf{\Gamma}_{2,1} = \mathbf{\Gamma}_{2,2} = \mathbf{\Gamma}_{2,3} = \frac{1}{3}\mathbf{I}. \end{aligned} \tag{17}$$

B. Experimental Results and Discussion

We conducted hyperspectral image denoising experiments based on Prob. (14) to illustrate the effectiveness of our ON-VW. Specifically, we compare the standard PDS with various stepsizes, P-PDS with the preconditioners proposed in [7], and P-PDS with ours (ON-VW) in solving Prob. (14).

[5]The original SSTV-regularized denoising formulation proposed in [18] incorporates an ℓ_2 data-fidelity term as a part of the objective function, whereas the formulation in (14) imposes data fidelity as an ℓ_2-ball constraint. These two formulations are essentially the same with appropriate hyperparameters, but constrained formulation like (14) is preferred in experimental comparison and real applications because it facilitates hyperparameter settings, as has been addressed in [6], [23]–[25].

[6]For a given nonempty closed convex set C, the indicator function of C is defined by $\iota_C(\mathcal{X}) := 0$, if $\mathcal{X} \in C$; ∞, otherwise.

[7]These are derived by $\|\mathfrak{D}_v\| \leq 2$, $\|\mathfrak{D}_h\| \leq 2$, $\|\mathfrak{D}_b\| \leq 2$ [26], and the submultiplicity of the operator norm.

For the standard PDS, the stepsizes γ_1 and γ_2 were set as follows. Using the inequality of the operator norm of the block matrix [17], we have

$$\begin{aligned} \left\| \begin{bmatrix} \mathfrak{D}_v \circ \mathfrak{D}_b & \mathfrak{O} \\ \mathfrak{D}_h \circ \mathfrak{D}_b & \mathfrak{O} \\ \mathbf{I} & \mathbf{I} \end{bmatrix} \right\|_{\text{op}}^2 \leq & \|\mathfrak{D}_v \circ \mathfrak{D}_b\|_{\text{op}}^2 + \|\mathfrak{D}_h \circ \mathfrak{D}_b\|_{\text{op}}^2 \\ & + 2\|\mathbf{I}\|_{\text{op}}^2 \\ \leq & 4^2 + 4^2 + 2 \times 1^2 = 34, \end{aligned} \tag{18}$$

where $\mathfrak{O}$ is the zero operator. Therefore, the following positive values satisfy the inequality in (6):

$$\gamma_1 \in (0, \infty), \quad \gamma_2 = \frac{1}{34\gamma_1}. \tag{19}$$

We also derived the preconditioners proposed in [7], which we call Row/Columun Absolute Sum-based Element-Wise Preconditioning (R/C-AS-EW), for (16). Let us remark that since $\mathfrak{D}_v$, $\mathfrak{D}_h$, and $\mathfrak{D}_b$ in (16) are not usually implemented as explicit matrices, applying R/C-AS-EW to (16) is not practical in real applications. Let $\mathbf{x} \in \mathbb{R}^{n_1 n_2 n_3}$ be a vectorized cube data and $[x]_{i_1,i_2,i_3}$ the value of $\mathbf{x}$ of the location (i_1, i_2, i_3). Then the preconditioners were

$$\begin{aligned} &\mathbf{\Gamma}_{1,1} = \text{diag}(\mathbf{g}), \mathbf{I}, \mathbf{\Gamma}_{1,2} = \mathbf{I}, \\ &\mathbf{\Gamma}_{2,1} = \mathbf{\Gamma}_{2,2} = \mathbf{\Gamma}_{2,3} = \frac{1}{2}\mathbf{I}. \end{aligned} \tag{20}$$

Here, $\mathbf{g} \in \mathbb{R}^{n_1 n_2 n_3}$ are defined as

$$[\mathbf{g}]_{i_1,i_2,i_3} = \begin{cases} \frac{1}{9}, & \text{if } i_1 \in I_1 \text{ and } i_2 \in I_2 \text{ and } i_3 \in I_3, \\ \frac{1}{3}, & \text{if } i_1 \in E_1 \text{ and } i_2 \in E_2 \text{ and } i_3 \in E_3, \\ \frac{1}{4}, & \text{if } i_3 \in E_3 \text{ and } \begin{cases} (i_1 \in E_1 \text{ and } i_2 \in I_2) \\ \text{or } (i_1 \in I_1 \text{ and } i_2 \in E_2), \end{cases} \\ \frac{1}{5}, & \text{if } i_1 \in E_1 \text{ and } i_2 \in E_2 \text{ and } i_3 \in I_3, \\ \frac{1}{7}, & \text{otherwise}, \end{cases} \tag{21}$$

where I_m and E_m $(m = 1, 2, 3)$ are $\{2, \ldots, n_m - 1\}$ and $\{1, n_m\}$, respectively.

As the groundtruth hyperspectral data, we used Moffett Field [27] of size $120 \times 120 \times 176$. The observed data was generated by adding white Gaussian noise with the standard deviation $\sigma = 0.05$ and salt and pepper noise with the percentage 10%. The parameters λ and ε were set to 0.1 and

$\|\mathbf{n}\|_2$, respectively. To check the convergence of PDS and P-PDS, we use the following RMSE criterion:

$$\mathrm{RMSE}(\mathbf{u}^{(n)}, \mathbf{s}^{(n)}) := \frac{\sqrt{\|\mathbf{u}^{(n)} - \mathbf{u}^*\|_2^2 + \|\mathbf{s}^{(n)} - \mathbf{s}^*\|_2^2}}{n_1 n_2 n_3}, \quad (22)$$

and the residual of the function values (RES):

$$\begin{aligned}&\mathrm{RES}(\mathbf{u}^{(n)}, \mathbf{s}^{(n)})\\ &:= |(\|\mathfrak{D}_v(\mathfrak{D}_b(\mathbf{u}^{(n)}))\|_1 + \|\mathfrak{D}_h(\mathfrak{D}_b(\mathbf{u}^{(n)}))\|_1 + \lambda\|\mathbf{s}^{(n)}\|_1)\\ &- (\|\mathfrak{D}_v(\mathfrak{D}_b(\mathbf{u}^*))\|_1 + \|\mathfrak{D}_h(\mathfrak{D}_b(\mathbf{u}^*))\|_1 + \lambda\|\mathbf{s}^*\|_1)|, \quad (23)\end{aligned}$$

where $\mathbf{u}^*$ and $\mathbf{s}^*$ are pseudo oracle solutions that were computed in advance by PDS with 1,000,000 iterations. For the quantitative evaluation of image qualities, we used peak signal to noise ratio (PSNR):

$$\mathrm{PSNR}(\mathbf{u}^{(n)}) := \frac{1}{n_3}\sum_{b=1}^{n_3} 10\log_{10}\frac{n_1 n_2}{\|\bar{\mathbf{u}}_b - \mathbf{u}_b^{(n)}\|_2^2}, \quad (24)$$

where $\mathbf{u}_b$ is the bth band of $\mathbf{u}$.

Fig. 1 plots the iteration versus NRMSE and RES, respectively. In terms of the convergence, PDSs with $\gamma_1 = 1$ and $\gamma_1 = 0.001$ are very slow, and PDS with $\gamma_1 = 0.01$ and P-PDSs with R/C-AS-EW and ON-VW are almost the same. PDS with $\gamma_1 = 0.1$ is the fastest in NRMSE, but is equal to ON-VW in RES. The PSNR values for PDSs $\gamma_1 = 0.1$, $\gamma_1 = 0.01$, R/C-AS-EW, and ON-VW are similar. The PSNR value for PDSs with $\gamma_1 = 1$ is high, but the sequence of variables do not converge, and the PSNR value decreases to around 33 [dB] as they converge. Although improving convergence is not the expected advantages of ON-VW, these observations imply that ON-VW performs as well as the standard PDS with hand-optimized stepsizes and R/C-AS-EW. Fig. 2 shows the denoising results and the PSNR values [dB] obtained by the standard PDS with $\gamma_1 = 0.01$, R/C-AS-EW, and ON-VW, where the stopping criterion is set to $\mathrm{NRMSE}(\mathbf{u}^{(n)}, \mathbf{s}^{(n)}) \leq 10^{-5}$. We can see that all results are equivalent in terms of the PSNR and the visual qualities.

V. Conclusion

We have proposed ON-VW, which automatically and easily adjusts the stepsizes in a variable-wise manner when the target optimization problem incorporates linear operators not represented as explicit matrices. We also proved the convergence of P-PDS with our preconditioners. An application of our method to hyperspectral image denoising was provided with experimental comparison, where we showed that ON-VW achieved the same convergence speed as the standard PDS with hand-optimized stepsizes and R/C-AS-EW.

References

[1] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Mach Learn.*, vol. 1, no. 3, pp. 127–239, 2014.

[2] P. L Combettes and J.-C. Pesquet, "Fixed point strategies in data science," *IEEE Trans. Signal Process.*, vol. 69, pp. 3878–3905, 2021.

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

[4] 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.

[5] B. C. Vu, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Adv. Comput. Math.*, vol. 38, no. 3, pp. 667–681, 2013.

[6] 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, Nov. 2015.

[7] T. Pock and A. Chambolle, "Diagonal preconditioning for first order primal-dual algorithms in convex optimization," in *IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 1762–1769.

[8] P. Giselsson and S. Boyd, "Acceleration of primal–dual methods by preconditioning and simple subproblem procedures," *Automatica*, vol. 62, pp. 1–10, 2015.

[9] P. Giselsson and S. Boyd, "Linear convergence and metric selection for Douglas-Rachford splitting and ADMM," *IEEE Trans. Autom. Control*, vol. 62, no. 2, pp. 532–544, Feg. 2017.

[10] Y. Liu, Y. Xu, and W. Yin, "Acceleration of primal–dual methods by preconditioning and simple subproblem procedures," *J. Sci. Comput.*, vol. 86, no. 21, pp. 1–34, Jan. 2021.

[11] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock, "An introduction to total variation for image analysis," in *Theoretical foundations and numerical methods for sparse recovery*, pp. 263–340. de Gruyter, 2010.

[12] K. Bredies and M. Holler, "Higher-order total variation approaches and generalisations," *Inverse Probl.*, vol. 36, no. 12, pp. 123001, Dec. 2020.

[13] J. Kovacevic and A. Chebira, "Life beyond bases: The advent of frames (part i)," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 86–104, Jul. 2007.

[14] J.-F. Cai, H. Ji, Z. Shen, and G.-B. Ye, "Data-driven tight frame construction and image denoising," *Appl. Comput. Harmon. Anal.*, vol. 37, no. 1, pp. 89–105, 2014.

[15] A. Parekh and I. W. Selesnick, "Convex denoising using non-convex tight frame regularization," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1786–1790, Oct. 2015.

[16] G. Chierchia, E. Chouzenoux, P. L. Combettes, and J.-C. Pesquet, "The proximity operator repository," *User's guide http://proximityoperator.net/download/guide.pdf (accessed October 3rd, 2021)*, 2020.

[17] R. Bhatia and F. Kittaneh, "Norm inequalities for partitioned operators and an application," *Math. Ann.*, vol. 287, no. 4, pp. 719–726, 1990.

[18] H. K. Aggarwal and A. Majumdar, "Hyperspectral image denoising using spatio-spectral total variation," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 442–446, Feb. 2016.

[19] H. Fan, C. Li, Y. Guo, G. Kuang, and J. Ma, "Spatial-spectral total variation regularized low-rank tensor decomposition for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6196–6213, Oct. 2018.

[20] W. He, H. Zhang, H. Shen, and L. Zhang, "Hyperspectral image denoising using local low-rank matrix recovery and global spatial-spectral total variation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 713–729, Mar. 2018.

[21] T. Ince, "Hyperspectral image denoising using group low-rank and spatial-spectral total variation," *IEEE Access*, vol. 7, pp. 52095–52109, Apr. 2019.

[22] M. Wang, Q. Wang, J. Chanussot, and D. Hong, "l_0-l_1 hybrid total variation regularization and its applications on hyperspectral image mixed noise removal and compressed sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7695–7710, Sep. 2021.

[23] 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, Mar. 2011.

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

[25] S. Ono, "Efficient constrained signal reconstruction by randomized epigraphical projection," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*. IEEE, 2019, pp. 4993–4997.

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

[27] "AVIRIS," https://aviris.jpl.nasa.gov/data/free_data.html.