An Algorithmic Framework of Variable Metric Over-Relaxed Hybrid Proximal Extra-Gradient Method

Li Shen [1] **Peng Sun** [1] **Yitong Wang** [1] **Wei Liu** [1] **Tong Zhang** [1]

Abstract

We propose a novel algorithmic framework of **V**ariable **M**etric **O**ver-**R**elaxed **H**ybrid **P**roximal **E**xtra-gradient (VMOR-HPE) method with a global convergence guarantee for the maximal monotone operator inclusion problem. Its iteration complexities and local linear convergence rate are provided, which theoretically demonstrate that a large over-relaxed step-size contributes to accelerating the proposed VMOR-HPE as a byproduct. Specifically, we find that a large class of primal and primal-dual operator splitting algorithms are all special cases of VMOR-HPE. Hence, the proposed framework offers a new insight into these operator splitting algorithms. In addition, we apply VMOR-HPE to the Karush-Kuhn-Tucker (KKT) generalized equation of linear equality constrained multi-block composite convex optimization, yielding a new algorithm, namely nonsymmetric **P**roximal **A**lternating **D**irection **M**ethod of **M**ultipliers with a preconditioned **E**xtra-gradient step in which the preconditioned metric is generated by a blockwise **B**arzilai-**B**orwein line search technique (PADMM-EBB). We also establish iteration complexities of PADMM-EBB in terms of the KKT residual. Finally, we apply PADMM-EBB to handle the nonnegative dual graph regularized low-rank representation problem. Promising results on synthetic and real datasets corroborate the efficacy of PADMM-EBB.

1. Introduction

Maximal monotone operator inclusion, as an extension of the KKT generalized equations for nonsmooth convex optimization and convex-concave saddle-point optimization, encompasses a class of important problems and has extensive applications in statistics, machine learning, signal and image processing, and so on. More concrete applications can be found in the literature (Combettes & Pesquet, 2011; Boyd et al., 2011; Bauschke & Combettes, 2017) and the references therein. Let $\mathbb{X}$ be a finite-dimensional linear vector space. We focus on the operator inclusion problem:

$$0 \in T(x),\ x \in \mathbb{X}, \tag{1}$$

where $T : \mathbb{X} \rightrightarrows \mathbb{X}$ is a maximal monotone operator.

One of the most efficient algorithms for problem (1) is **P**roximal **P**oint **A**lgorithm (PPA) in the seminal work (Minty, 1962), which was further accelerated (Eckstein & Bertsekas, 1992) by attaching an over-relaxed parameter θ_k,

$$x^{k+1} := x^k + (1+\theta_k)\big(\mathcal{J}_{c_kT}(x^k) - x^k\big),\ \theta_k \in (-1, 1)$$

for a given positive penalty parameter c_k. Here, $\mathcal{J}_{c_kT}(\cdot) = (I + c_kT)^{-1}(\cdot)$ is called the resolvent operator (Bauschke & Combettes, 2017) of T. In addition, its inexact version

$$x^{k+1} := x^k + (1+\theta_k)\big(\overline{x}^k - x^k\big) \tag{2}$$

was proposed (Rockafellar, 1976) by requiring that either absolute error (3a) or relative error criterion (3b) holds,

$$\begin{cases} \big\|\overline{x}^k - \mathcal{J}_{c_kT}(x^k)\big\| \le \xi_k, & \text{(3a)} \\ \big\|\overline{x}^k - \mathcal{J}_{c_kT}(x^k)\big\| \le \xi_k\big\|\overline{x}^k - x^k\big\|, & \text{(3b)} \end{cases}$$

where $\sum_{k=1}^{\infty} \xi_k < \infty$. However, it is too flexible to preset the sequence $\{\xi_k\}$ which highly influences the level of the computational cost and quality of iteration (2). For more research on PPA and its inexact variants, we refer the readers to the literature (Güler, 1991; Burke & Qian, 1999; Corman & Yuan, 2014; Shen & Pan, 2015; Tao & Yuan, 2017).

Later on, a novel inexact PPA called **H**ybrid **P**roximal **E**xtra-gradient (HPE) algorithm (Solodov & Svaiter, 1999) was proposed. This algorithm first seeks a triple point $(y^k, v^k, \epsilon_k) \in \mathbb{X}\times\mathbb{X}\times\mathbb{R}_+$ satisfying error criterion (4a)-(4b):

$$\begin{cases} (y^k, v^k) \in \operatorname{gph} T^{[\epsilon_k]}, & \text{(4a)} \\ \big\|c_kv^k + (y^k - x^k)\big\|^2 + 2c_k\epsilon_k \le \sigma\big\|y^k - x^k\big\|^2, & \text{(4b)} \\ x^{k+1} := x^k - c_kv^k, & \text{(4c)} \end{cases}$$

---

[1]Tencent AI Lab, Shenzhen, China. Correspondence to: Li Shen <mathshenli@gmail.com>, Peng Sun <pengsun000@gmail.com>, Yitong Wang <yitongwang@tencent.com>, Wei Liu <wl2223@columbia.edu>, Tong Zhang <tongzhang@tongzhang-ml.org>.

Proceedings of the 35^{th} International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.

where $T^{[\epsilon]}$ is the enlargement operator (Burachik et al., 1997; 1998; Svaiter, 2000) of T and $\sigma \in [0,1)$ is a prespecified parameter, and then executes an extra-gradient step (4c) to ensure its global convergence. Whereafter, a new inexact criterion (5a)-(5b) is adopted, yielding an over-relaxed HPE algorithm (Svaiter, 2001; Parente et al., 2008) as below:

$$\begin{cases} (y^k, v^k) \in \operatorname{gph} T^{[\epsilon_k]}, & (5a)\\ \left\|c_k\mathcal{M}_k^{-1}v^k + (y^k - x^k)\right\|^2_{\mathcal{M}_k} + 2c_k\epsilon_k & (5b)\\ \qquad \le \sigma\left(\left\|y^k - x^k\right\|^2_{\mathcal{M}_k} + \left\|c_k\mathcal{M}_k^{-1}\right\|^2_{\mathcal{M}_k}\right), & \\ x^{k+1} := x^k - (1+\tau_k)a_k\mathcal{M}_k v^k, & (5c) \end{cases}$$

where $\tau_k \in (-1,1)$ is the over-relaxed step-size, $a_k = \left[\langle v^k, x^k - y^k\rangle - \epsilon_k\right] / \left\|\mathcal{M}_k^{-1}v^k\right\|^2_{\mathcal{M}_k}$, and $\mathcal{M}_k$ is a self-adjoint positive definite linear operator. An obvious defect of the above algorithm is that extra-gradient step-size a_k has to be adaptively determined to ensure its global convergence, which requires extra computation and may be time-consuming. In addition, Korpelevich's extra-gradient algorithm (Korpelevich, 1977), forward-backward algorithm (Passty, 1979), and forward-backward-forward algorithm (Tseng, 2000) are all shown to be special cases of the HPE algorithm in (Solodov & Svaiter, 1999; Svaiter, 2014).

In this paper, we propose a new algorithmic framework of **V**ariable **M**etric **O**ver-**R**elaxed **H**ybrid **P**roximal **E**xtra-gradient (VMOR-HPE) method with a global convergence guarantee for solving problem (1). This framework, in contrast to the existing HPE algorithms, generates the iteration sequences in terms of a novel relative error criterion and introduces an over-relaxed step-size in the extra-gradient step to improve its performance. In particular, the extra-gradient step-size and over-relaxed step-size here can both be set as a fixed constant in advance, instead of those obtained from a projection problem, which saves extra computation. Its global convergence, $\mathcal{O}(\frac{1}{\sqrt{k}})$ pointwise and $\mathcal{O}(\frac{1}{k})$ weighted iteration complexities, and the local linear convergence rate under some mild metric subregularity condition (Dontchev & Rockafellar, 2009) are also built. Interestingly, the coefficients of iteration complexities and linear convergence rate are inversely proportional to the over-relaxed step-size, which theoretically demonstrates that a large over-relaxed step-size contributes to accelerating the proposed VMOR-HPE as a byproduct. In addition, we rigorously show that a class of primal-dual algorithms, including **A**symmetric **F**orward **B**ackward **A**djoint **S**plitting **P**rimal-**D**ual (AFBAS-PD) algorithm (Latafat & Patrinos, 2017), Condat-Vu **P**rimal-**D**ual **S**plitting (Condat-Vu PDS) algorithm (Vũ, 2013; Condat, 2013), **P**rimal-**D**ual **F**ixed **P**oint (PDFP) algorithm (Chen et al., 2016), **P**rimal-**D**ual three **O**perator **S**plitting (PD3OS) algorithm (Yan, 2018), Combettes **P**rimal-**D**ual **S**plitting (Combettes PDS) algorithm (Combettes & Pesquet, 2012), **M**onotone+**S**kew **S**plitting (MSS) algorithm (Briceño Arias & Combettes, 2011), **P**roximal **A**lternating **P**redictor **C**orrector (PAPC) algorithm (Drori et al., 2015), and **P**rimal-**D**ual **H**ybrid **G**radient (PDHG) algorithm (Chambolle & Pock, 2011), all fall into the VMOR-HPE framework with specific variable metric operators $\mathcal{M}_k$ and T. Besides, **P**roximal-**P**roximal-**G**radient (PPG) algorithm (Ryu & Yin, 2017), **F**orward-**B**ackward-**H**alf **F**orward (FBHF) algorithm as well as its non self-adjoint metric extensions (Briceño-Arias & Davis, 2018), Davis-Yin three **O**perator **S**plitting (Davis-Yin 3OS) algorithm (Davis & Yin, 2015), **F**orward **D**ouglas-**R**achford **S**plitting (FDRS) algorithm (Briceño-Arias, 2015a),**G**eneralized **F**orward **B**ackward **S**plitting (GFBS) algorithm (Raguet et al., 2013), and **F**orward **D**ouglas-**R**achford **F**orward **S**plitting (FDRFS) algorithm (Briceño-Arias, 2015b) also fall into the VMOR-HPE framework. Thus, VMOR-HPE largely expands the HPE algorithmic framework to cover a large class of primal and primal-dual algorithms and their non self-adjoint metric extensions compared with (Solodov & Svaiter, 1999; Shen, 2017). As a consequence, the VMOR-HPE algorithmic framework offers a new insight into aforementioned primal and primal-dual algorithms and serves as a powerful analysis technique for establishing their convergences, iteration complexities, and local linear convergence rates.

In addition, we apply VMOR-HPE to the KKT generalized equation of linear equality constrained multi-block composite nonsmooth convex optimization as follows:

$$\begin{aligned} \min_{x_i \in \mathbb{X}_i} \quad & f(x_1, \ldots, x_p) + g_1(x_1) + \cdots + g_p(x_p) \qquad (6)\\ \text{s.t.} \quad & \mathcal{A}_1^* x_1 + \mathcal{A}_2^* x_2 + \cdots + \mathcal{A}_p^* x_p = b, \end{aligned}$$

where $\mathcal{A}_i^* : \mathbb{Y} \to \mathbb{X}_i$ is the adjoint linear operator of $\mathcal{A}_i$, $\mathbb{Y}$ and $\mathbb{X}_i$ are given finite-dimensional vector spaces, $g_i : \mathbb{X}_i \to (-\infty, +\infty]$ is a proper closed convex function, and $f : \mathbb{X}_1 \times \cdots \times \mathbb{X}_p \to \mathbb{R}$ is a gradient Lipschitz continuous convex function. Specifically, the proposed VMOR-HPE for solving problem (6) firstly generates points satisfying the relative inexact criterion in the VMOR-HPE framework by a newly developed nonsymmetric **P**roximal **A**lternating **D**irection **M**ethod of **M**ultipliers, and then performs an over-relaxed metric **E**xtra-gradient correction step to ensure its global convergence. Notably, metric $\mathcal{M}_k$ in the extra-gradient step is generated by using a blockwise **B**arzilai-**B**orwein line search technique (Barzilai & Borwein, 1988) to exploit the curvature information of the KKT generalized equation of (6). We thus name the resulting new algorithm as PADMM-EBB. Moreover, we establish the $\mathcal{O}(\frac{1}{\sqrt{k}})$ pointwise and $\mathcal{O}(\frac{1}{k})$ weighted iteration complexities and the local linear convergence rate for PADMM-EBB on the KKT residual of (6) by employing the VMOR-HPE framework. Besides, it is worth emphasizing that the derived iteration complexities do not need any assumption on the boundedness of the feasible set of (6). At last, we conduct experiments on the nonnegative dual graph regularized low-rank representation problem to verify the efficacy of PADMM-EBB, which shows great superiority over **P**roximal **L**inearized ADMM with **P**arallel **S**plitting and **A**daptive **P**enalty (PLADMM-PSAP) (Liu et al., 2013; Lin et al., 2015), **P**roximal **G**auss-**S**eidel ADMM (PGSADMM) with nondecreasing penalty, and **M**ixed **G**auss-**S**eidel and **J**acobi ADMM (M-GSJADMM) with nondecreasing penalty (Lu et al., 2017) on both synthetic and real datasets.

The major contributions of this paper are fourfold. **(i)** We propose a new algorithmic framework of VMOR-HPE for problem (1) and also establish its global convergence, iteration complexities, and local linear convergence rate. **(ii)** The proposed VMOR-HPE gives a new insight into a large class of primal and primal-dual algorithms and provides a unified analysis framework for their convergence properties. **(iii)** Applying VMOR-HPE to problem (6) yields a new convergent primal-dual algorithm whose iteration complexities on the KKT residual are also provided without requiring the boundedness of the feasible set of (6). **(iv)** Numerical experiments on synthetic and real datasets are conducted to demonstrate the superiority of the proposed algorithm.

2. Preliminaries

Given $\beta > 0$, a single-valued mapping $C : \mathbb{X} \to \mathbb{X}$ satisfing $\langle x - x', C(x) - C(x') \rangle \geq \beta \|C(x) - C(x')\|^2$ for all $x, x' \in \mathbb{X}$ is called a β-cocoercive operator. A set-valued mapping $T : \mathbb{X} \rightrightarrows \mathbb{X}$ satisfying $\langle x - x', v - v' \rangle \geq \alpha \|x - x'\|^2$ with $\alpha \geq 0$ for all $v \in T(x)$ and $v' \in T(x')$ is called α-strongly monotone operator if $\alpha > 0$, and is called a monotone operator if $\alpha = 0$. Moreover, T is called a maximal monotone operator if there does not exit any monotone operator T' satisfying $\operatorname{gph} T \subseteq \operatorname{gph} T'$, where $\operatorname{gph} T := \{(x, v) \in \mathbb{X} \times \mathbb{X} \mid v \in T(x), x \in \mathbb{X}\}$. In addition, given $\epsilon \geq 0$ and a maximal monotone operator T, the ϵ-enlargement $T^{[\epsilon]} : \mathbb{X} \rightrightarrows \mathbb{X}$ of T (Burachik et al., 1997; 1998; Svaiter, 2000) is defined as

$$T^{[\epsilon]}(x) := \{v \in \mathbb{X} \mid \langle w - v, z - x \rangle \geq -\epsilon, \forall w \in T(z)\}.$$

Below, we recall the definition of metric subregularity (Dontchev & Rockafellar, 2009) of set-valued mapping T.

Definition 1. *A set-valued mapping $T : \mathbb{X} \rightrightarrows \mathbb{X}$ is metric subregular at $(\overline{x}, \overline{y}) \in \operatorname{gph} T$ with modulus $\kappa > 0$, if there exists a neighborhood U of $\overline{x}$ such that for all $x \in U$,*

$$\operatorname{dist}(x, T^{-1}(\overline{y})) \leq \kappa \operatorname{dist}(\overline{y}, T(x)).$$

Given a self-adjoint positive definite linear operator $\mathcal{M}$, $\|\cdot\|_{\mathcal{M}}$ denotes the generalized norm induced by $\mathcal{M}$, which is defined as $\|\cdot\|_{\mathcal{M}} = \sqrt{\langle \cdot, \mathcal{M} \cdot \rangle}$. The generalized distance between a point z and a set Ω induced by $\mathcal{M}$ is defined as $\operatorname{dist}_{\mathcal{M}}(z, \Omega) := \inf_{x \in \Omega} \|x - z\|_{\mathcal{M}}$. Let $\mathcal{M} = \mathcal{I}$. $\operatorname{dist}_{\mathcal{M}}(z, \Omega)$ reduces to the standard distance function as $\operatorname{dist}(z, \Omega) := \inf_{x \in \Omega} \|x - z\|$. In addition, given a proper closed convex function $g : \mathbb{X} \to (\infty, +\infty]$ and a non self-adjoint linear operator $\mathcal{R}$, $\operatorname{Prox}_{\mathcal{R}^{-1} g}(\cdot)$ denoting the generalized proximal mapping of g induced by $\mathcal{R}$ is the unique root of inclusion:

$$0 \in \partial g(x) + \mathcal{R}(x - \cdot), \ x \in \mathbb{X}.$$

Particularly, if $g(x) = \sum_{i=1}^n g_i(x_i)$ is decomposable, $\operatorname{Prox}_{\mathcal{R}^{-1} g}(\cdot)$ can be calculated in a Gauss-Seidel manner by merely setting $\mathcal{R}$ as a block lower-triangular linear operator.

3. VMOR-HPE Framework

In this section, we propose the algorithmic framework of VMOR-HPE (described in Algorithm 1), and establish its global convergence rate, iteration complexities, and local linear convergence rate. Let $\mathcal{M}_k = \mathcal{I}$ in VMOR-HPE. We recover an enhanced version of an over-relaxed HPE algorithm (Shen, 2017) by allowing a larger over-relaxed step-size θ_k.

Algorithm 1 VMOR-HPE Framework

Parameters: Given $\underline{\omega}$, $\overline{\omega} > 0$, $\underline{\theta} > -1$, $\sigma \in [0, 1)$ and $\xi_k \geq 0$ satisfying $\sum_{k=1}^{\infty} \xi_k < \infty$. Choose a self-adjoint operator $\mathcal{M}_0$ satisfying $\underline{\omega}\mathcal{I} \preceq \mathcal{M}_0 \preceq \overline{\omega}\mathcal{I}$ and $x^0 \in \mathbb{X}$.

for $k = 1, 2, \cdots$, **do**

Choose $c_k \geq \underline{c} > 0$, $\theta_k \in [\underline{\theta}, \infty)$. Find $(\epsilon_k, y^k, v^k) \in \mathbb{R}_+ \times \mathbb{X} \times \mathbb{X}$ satisfying the relative error criterion that

$$\begin{cases} (y^k, v^k) \in \operatorname{gph} T^{[\epsilon_k]}, & \text{(7a)} \\ \theta_k \|c_k \mathcal{M}_k^{-1} v^k\|_{\mathcal{M}_k}^2 + \|c_k \mathcal{M}_k^{-1} v + (y^k - x^k)\|_{\mathcal{M}_k}^2 & \\ \qquad + 2c_k\epsilon_k \leq \sigma \|y^k - x^k\|_{\mathcal{M}_k}^2. & \text{(7b)} \end{cases}$$

Let $x^{k+1} := x^k - (1 + \theta_k) c_k \mathcal{M}_k^{-1} v^k$.

Update $\mathcal{M}_{k+1}$ with $\underline{\omega}\mathcal{I} \preceq \mathcal{M}_{k+1} \preceq (1 + \xi_k)\mathcal{M}_k$.

end for

Remark 1. **(i)** $\theta_k \in [\underline{\theta}, \infty)$ *breaks the ceiling of over-relaxed step-sizes in the literature (Eckstein & Bertsekas, 1992; Chambolle & Pock, 2016; Bauschke & Combettes, 2017; Shen, 2017; Tao & Yuan, 2017) ,in which* $\theta_k \in (-1, 1)$. *Besides,* $\mathcal{M}_k$ *can exploit the curvature information of* T.
(ii) *Let* $\theta_k = -\sigma$ *in the VMOR-HPE framework. Criterion* (7a)-(7b) *coincides with* (5a)-(5b) *in (Parente et al., 2008), which makes the step-size* $(1 + \theta_k)$ *be* $(1 - \sigma)$ *that is too small to update* x^{k+1} *if* σ *is close to* 1. *That is the reason why* a_k *in* (5c) *has to be adaptively computed with extra computation instead of being a constant.*

3.1. Convergence Analysis

In this subsection, we build the global convergence for the algorithmic framework of VMOR-HPE, as well as its local linear convergence rate under a metric subregularity condition of T. In addition, its $\mathcal{O}(\frac{1}{\sqrt{k}})$ pointwise and $\mathcal{O}(\frac{1}{k})$ weighted iteration complexities depending solely on $(T^{-1}(0), x^0)$ are provided. Denote $\Xi := \prod_{i=0}^{\infty}(1+\xi_i) < \exp\left(\sum_{i=0}^{\infty}\xi_i\right) < \infty$.

Theorem 1. *Let $\{(x^k, y^k)\}$ be the sequence generated by the VMOR-HPE framework. Then, $\{x^k\}$ and $\{y^k\}$ both converge to a point x^∞ belonging to $T^{-1}(0)$.*

Theorem 2. *Let $\{(x^k, y^k)\}$ be the sequence generated by the VMOR-HPE framework. Assume that the metric subregularity of T at $(x^\infty, 0) \in \operatorname{gph} T$ holds with $\kappa > 0$. Then, there exits $\overline{k} > 0$ such that for all $k \geq \overline{k}$,*

$$\operatorname{dist}^2_{\mathcal{M}_{k+1}}\left(x^{k+1}, T^{-1}(0)\right) \leq \left(1 - \frac{\varrho_k}{2}\right)\operatorname{dist}^2_{\mathcal{M}_k}\left(x^k, T^{-1}(0)\right),$$

where $\varrho_k = \frac{(1-\sigma)(1+\theta_k)}{\left(1+\frac{\kappa}{\underline{c}}\sqrt{\frac{\Xi\overline{\omega}}{\underline{\omega}}}\right)^2\left(1+\sqrt{\sigma+\frac{4\max\{-\theta_k,0\}}{(1+\theta_k)^2}}\right)^2} \in (0,1).$

Polyhedra operators (Robinson, 1981) and strongly monotone operators all satisfy metric subregularity. For other sufficient conditions that guarantee metric subregulaity of T, we refer the readers to the monographs (Dontchev & Rockafellar, 2009; Rockafellar & Wets, 2009; Cui, 2016).

Point $x \in \mathbb{X}$ is called ε-solution (Monteiro & Svaiter, 2010) of problem (1) if there exists $(v, \epsilon) \in \mathbb{X} \times \mathbb{R}_+$ satisfying $v \in T^{[\epsilon]}(x)$ and $\max(\|v\|, \epsilon) \leq \varepsilon$. Below, we globally characterize the rate of $\max(\|v\|, \epsilon)$ decreasing to zero.

Theorem 3. *Let $\{(x^k, y^k, v^k)\}$ and $\{\epsilon_k\}$ be the sequences generated by the VMOR-HPE framework.*

(i) *There exists an integer $k_0 \in \{1, 2, \ldots, k\}$ such that $v^{k_0} \in T^{[\epsilon_{k_0}]}(y^{k_0})$ with v^{k_0} and $\epsilon_{k_0} \geq 0$ respectively satisfying*

$$\|v^{k_0}\| \leq \sqrt{\frac{4(1+\sum_{i=1}^k \xi_i)\Xi^2\overline{\omega}}{k(1-\sigma)(1+\underline{\theta})^3\underline{c}^2}}\|x^0 - x^*\|_{\mathcal{M}_0},$$

$$\text{and} \quad \epsilon_{k_0} \leq \frac{(1+\sum_{i=1}^k \xi_i)\Xi}{k(1-\sigma)(1+\underline{\theta})^2\underline{c}}\|x^0 - x^*\|^2_{\mathcal{M}_0}.$$

(ii) *Let $\{\alpha_k\}$ be the nonnegative weight sequence satisfying $\sum_{i=1}^k \alpha_i > 0$. Denote $\tau_i = (1+\theta_i)c_i$ and $\overline{y}^k = \frac{\sum_{i=1}^k \tau_i\alpha_i y^i}{\sum_{i=1}^k \tau_i\alpha_i}$,*

$$\overline{v}^k = \frac{\sum_{i=1}^k \tau_i\alpha_i v^i}{\sum_{i=1}^k \tau_i\alpha_i}, \overline{\epsilon}_k = \frac{\sum_{i=1}^k \tau_i\alpha_i\left(\epsilon_i + \langle y^i - \overline{y}^k, v^i - \overline{v}^k\rangle\right)}{\sum_{i=1}^k \tau_i\alpha_i}.$$

Then, it holds that $\overline{v}^k \in T^{[\overline{\epsilon}_k]}(\overline{y}^k)$ with $\overline{\epsilon}_k \geq 0$. Moreover, if $\mathcal{M}_k \leq (1+\xi_k)\mathcal{M}_{k+1}$, it holds that

$$\|\overline{v}^k\| \leq \frac{\max\limits_{1\leq i\leq k}\{\alpha_{i+1}\}\sum\limits_{i=1}^k \xi_i + \sum\limits_{i=1}^k\left|\alpha_i - \alpha_{i+1}\right| + \alpha_{k+1} + \alpha_1}{\underline{c}(1+\underline{\theta})\sum_{i=1}^k \alpha_i}M,$$

$$\overline{\epsilon}_k = \frac{(10+\underline{\theta})\max\limits_{1\leq i\leq k}\{\alpha_i\}\left(1+\sum\limits_{i=1}^k \xi_i\right) + (2+\underline{\theta})\sum\limits_{i=1}^k\left|\alpha_{i+1} - \alpha_i\right|}{\underline{c}(1+\underline{\theta})^2\sum_{i=1}^k \alpha_i}B,$$

where M and B are two constants which are respectively defined as $M = \Xi\overline{\omega}\left[\|x^*\| + \sqrt{\Xi/\underline{\omega}}\|x^0 - x^*\|_{\mathcal{M}_0}\right]$ and

$$B = \max\left\{\begin{array}{c} M,\ \Xi\left\|x^*\right\|^2 + \frac{\Xi^2}{\underline{\omega}}\left\|x^0 - x^*\right\|^2_{\mathcal{M}_0}, \\ \frac{\Xi^2}{(1-\sigma)\underline{\omega}}\left\|x^0 - x^*\right\|^2_{\mathcal{M}_0},\ \frac{\Xi}{(1-\sigma)}\left\|x^0 - x^*\right\|^2_{\mathcal{M}_0}\end{array}\right\}.$$

Remark 2. **(i)** *The iteration complexities in Theorem 3 merely depend on the solution set $T^{-1}(0)$ and initial point x^0. The upper bounds of $(v^{k_0}, \epsilon_{k_0})$ and $(\overline{v}_k, \overline{\epsilon}_0)$ are both inversely proportional to θ_k, which, in combination with Theorem 2, theoretically demonstrates that a large over-relaxed step-size contributes to accelerating VMOR-HPE.*

(ii) *Set $\alpha_k = 1$ or k. It holds that $\|\overline{v}^k\| \leq \mathcal{O}(\frac{1}{k})$ and $\overline{\epsilon}_k \leq \mathcal{O}(\frac{1}{k})$. However, setting $\alpha_k = k$ may lead to better performance than setting $\alpha_k = 1$, since $\alpha_k = k$ gives more weights on the latest generated points y^k and v^k.*

3.2. Connection to Existing Algorithms

First, we consider $\mathcal{M}_k = \mathcal{I}$. Under this situation, the proposed VMOR-HPE reduces to the over-relaxed HPE algorithm (Shen, 2017) which covers a number of primal first-order algorithms as special cases, such as FDRS algorithm, GFBS algorithm, FDRFS algorithm, *etc*. Hence, they are also covered by the algorithmic framework of VMOR-HPE. Below, we show a large collection of other primal and primal-dual algorithms which fall into VMOR-HPE.

3.2.1. PRIMAL ALGORITHMS

FBHF Algorithm tackles problem (1) as

$$0 \in T(x) = (A + B_1 + B_2)(x),\ x \in \Omega,$$

where A is a maximal monotone operator, $B_1 : \mathbb{X} \to \mathbb{X}$ is a β-cocoercive operator, $B_2 : \mathbb{X} \to \mathbb{X}$ is a monotone and L-Lipschitz continuous operator, and Ω is a subset of $\mathbb{X}$. The FBHF algorithm has the iterations:

$$\begin{cases} y^k := \mathcal{J}_{\gamma_k A}\left(x^k - \gamma_k(B_1 + B_2)x^k\right), \\ x^{k+1} := P_\Omega\left(y^k + \gamma_k B_2(x^k) - \gamma_k B_2(y^k)\right). \end{cases}$$

In the following, we focus on $\Omega = \mathbb{X}$ and replace x^{k+1} by

$$x^{k+1} := x^k + (1+\theta_k)\left(y^k - x^k + \gamma_k B_2(x^k) - \gamma_k B_2(y^k)\right)$$

to obtain an over-relaxed FBHF algorithm. The proposition below rigorously reformulates the over-relaxed FBHF algorithm as a specific case of the VMOR-HPE framework.

Proposition 1. *Let $\{(x^k, y^k)\}$ be the sequence generated by the over-relaxed FBHF algorithm. Denote $\epsilon_k = \|x^k - y^k\|^2/(4\beta)$ and $v^k = \gamma_k^{-1}(x^k - y^k) - B_2(x^k) + B_2(y^k)$. Then,*

$$\begin{cases} (y^k, v^k) \in \operatorname{gph} T^{[\epsilon_k]} = \operatorname{gph}(A + B_1 + B_2)^{[\epsilon_k]}, \\ \theta_k\left\|\gamma_k v^k\right\|^2 + \left\|\gamma_k v^k + (y^k - x^k)\right\|^2 + 2\gamma_k\epsilon \leq \sigma\left\|y^k - x^k\right\|^2, \\ x^{k+1} = x^k - (1+\theta_k)\gamma_k v^k, \end{cases}$$

where (γ_k, θ_k) satisfies $\theta_k \leq \frac{\sigma - (\gamma_k L)^2 + \gamma_k/(2\beta))}{1+(\gamma_k L)^2}$.

Remark 3. **(i)** *If $\theta_k = 0$, γ_k reduces to $\gamma_k^2 L^2 + \gamma_k/(2\beta) \leq \sigma < 1 \Leftrightarrow 0 < \gamma_k < 4\beta/(1+\sqrt{1+16\beta^2 L^2})$ which coincides with the properties of γ_k in (Briceño-Arias & Davis, 2018).*

(ii) *By (Solodov & Svaiter, 1999), a slightly modified VMOR-HPE by attaching an extra projection step P_Ω on x^{k+1} can cover the original FBHF algorithm.*
(iii) *Let $B_1 = 0$ or $B_2 = 0$. The over-relaxed FBHF algorithm reduces to over-relaxed Tseng's forward-backward-forward splitting algorithm (Tseng, 2000) or over-relaxed forward-backward splitting algorithm (Passty, 1979). Thus, they are special cases of VMOR-HPE by Proposition 1.*

nMFBHF Algorithm The **n**on self-adjoint **M**etric variant of FBHF (nMFBHF) algorithm takes the iterations:

$$\begin{cases} y^k := \mathcal{J}_{P^{-1}A}\big(x^k - P^{-1}(B_1+B_2)(x^k)\big), \\ x^{k+1} := P_\Omega^U\big(y^k + U^{-1}[B_2(x^k) - B_2(y^k) - S(x^k - y^k)]\big), \end{cases}$$

where P is a bounded linear operator, $U = (P+P^*)/2$, $S = (P-P^*)/2$, and P_Ω^U is the projection operator of Ω under the weighted inner product $\langle \cdot, U\cdot\rangle$. Similarly, let $\Omega = \mathbb{X}$. We obtain the over-relaxed nMFBHF algorithm by replacing the updating step x^{k+1} as the following form

$$x^{k+1} := x^k + (1+\theta_k)\big(y^k - x^k + U^{-1}[B_2(x^k) - B_2(y^k)] - U^{-1}[S(x^k - y^k)]\big).$$

Below, we show that the over-relaxed nMFBHF algorithm also falls into the VMOR-HPE framework. Notice that $B_2 - S$ preserves the monotonicity by the skew symmetry of S, and K is denoted as its Lipschitz constant.

Proposition 2. *Let $\{(x^k, y^k)\}$ be the sequence generated by the over-relaxed nMFBHF algorithm. Denote $\epsilon_k = \|x^k - y^k\|^2/(4\beta)$ and $v^k = P(x^k - y^k) + B_2(y^k) - B_2(x^k)$. The step-size θ_k satisfies $\theta_k + \frac{K^2(1+\theta_k)}{\lambda^2_{\min}(U)} + \frac{1}{2\beta\lambda_{\min}(U)} \le \sigma$. Then,*

$$\begin{cases} (y^k, v^k) \in \operatorname{gph} T^{[\epsilon_k]} = \operatorname{gph}(A + B_1 + B_2)^{[\epsilon_k]}, \\ \theta_k\big\|U^{-1}v^k\big\|_U^2 + \big\|U^{-1}v + (y^k - x^k)\big\|_U^2 + 2\epsilon \le \sigma\big\|y^k - x^k\big\|_U^2, \\ x^{k+1} = x^k - (1+\theta_k)U^{-1}v^k. \end{cases}$$

Let $\theta_k = 0$, and then $\theta_k + \frac{K^2(1+\theta_k)}{\lambda^2_{\min}(U)} + \frac{1}{2\beta\lambda_{\min}(U)} \le \sigma < 1$ reduces to $\frac{K^2}{\lambda^2_{\min}(U)} + \frac{1}{2\beta\lambda_{\min}(U)} < 1$, which coincides with the required condition in (Briceño-Arias & Davis, 2018).

PPG Algorithm Consider the following minimization of a sum of many smooth and nonsmooth convex functions

$$\min_{x\in\mathbb{X}} r(x) + \frac{1}{n}\sum_{i=1}^n f_i(x) + \frac{1}{n}\sum_{i=1}^n g_i(x). \tag{12}$$

Let $\alpha \in (0, \frac{3}{2L})$. The PPG algorithm takes iterations as

$$\begin{cases} x^{k+\frac12} := \operatorname{Prox}_{\alpha r}\big(\frac{1}{n}\sum_{i=1}^n z_i^k\big), \\ x_i^{k+1} := \operatorname{Prox}_{\alpha g_i}\big(2x^{k+\frac12} - z_i^k - \alpha\nabla f_i(x^{k+\frac12})\big),\ i = 1,\ldots,n, \\ z_i^{k+1} := z_i^k + x_i^{k+1} - x^{k+\frac12},\ i = 1,2,\ldots,n, \end{cases}$$

where $g_i, r : \mathbb{X} \to (-\infty, +\infty]$ are proper closed convex functions, and $f_i : \mathbb{X} \to (-\infty, +\infty)$ is a differentiable convex function satisfying $\|\nabla f_i(x) - \nabla f_i(y)\| \le L\|x - y\|$ for all i.

Denote $\overline{f}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^n f_i(x_i)$, $\overline{g}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^n g_i(x_i)$ and $\overline{r}(\mathbf{x}) = \mathbf{1}_V(\mathbf{x}) + \frac{1}{n}\sum_{i=1}^n r(x_i)$, where $\mathbf{1}_V(\mathbf{x})$ is an indicator function over V. $V = \{\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{X}^n \mid \mathbb{X}^n = \mathbb{X}\times\mathbb{X}\times\ldots\times\mathbb{X},\ x_1 = x_2 = \cdots = x_n\}$. Then, problem (12) is equivalent to $\min_{\mathbf{x}} \overline{f}(\mathbf{x}) + \overline{g}(\mathbf{x}) + \overline{r}(\mathbf{x})$ and

$$0 \in \nabla\overline{f}(\mathbf{x}) + \partial\overline{r}(\mathbf{x}) + \partial\overline{g}(\mathbf{x}), \mathbf{x} \in \mathbb{X}^n. \tag{14}$$

Following the notation in (Shen, 2017), for $\alpha > 0$ we define the set-valued mapping $\mathcal{S}_{\alpha,\nabla\overline{f}+\partial\overline{g},\partial\overline{r}} : \mathbb{X}^n \rightrightarrows \mathbb{X}^n$ as:

$$\operatorname{gph}\big(\mathcal{S}_{\alpha,\nabla\overline{f}+\partial\overline{g},\partial\overline{r}}\big) = \big\{(\mathbf{x}_1 + \alpha\mathbf{y}_2, \mathbf{x}_2 - \mathbf{x}_1) \mid (\mathbf{x}_2, \mathbf{y}_2) \in \operatorname{gph}\partial\overline{r}, (\mathbf{x}_1, \mathbf{y}_1) \in \operatorname{gph}(\nabla\overline{f} + \overline{\partial} g), \mathbf{x}_1 + \alpha\mathbf{y}_1 = \mathbf{x}_2 - \alpha\mathbf{y}_2\big\}.$$

By the convexity of $\overline{f}, \overline{g}$ and $\overline{r}$, $\mathcal{S}_{\alpha,\nabla\overline{f}+\partial\overline{g},\partial\overline{r}}$ is a maximal monotone operator (Eckstein & Bertsekas, 1992). To obtain the over-relaxed PPG algorithm, we replace z_i^{k+1} by

$$z_i^{k+1} := z_i^k + (1+\theta_k)(x_i^{k+1} - x^{k+\frac12}),\ i = 1,\ldots,n.$$

Below, we show that the over-relaxed PPG algorithm is a specific case of the VMOR-HPE framework.

Proposition 3. *Let $(x^{k+\frac12}, x_i^k, z_i^k)$ be the sequence generated by the over-relaxed PPG algorithm. Denote $\mathbf{x}^k = (x_1^k, \cdots, x_n^k)$, $\mathbf{z}^k = (z_1^k, \cdots, z_n^k)$, $\mathbf{1} = (1, \cdots, 1) \in \mathbb{X}^n$, $\mathbf{y}^k = \mathbf{z}^k + \mathbf{x}^{\mathbf{k+1}} - x^{k+\frac12}\mathbf{1}$, $\mathbf{v}^k = x^{k+\frac12}\mathbf{1} - \mathbf{x}^{k+1}$ and $\epsilon_k = L\sum_{i=1}^n \|x_i^{k+1} - x^{k+\frac12}\|/4$. Parameters (θ_k, α) are constrained by $\theta_k + L\alpha/2 \le \sigma$. Then, it holds that*

$$\begin{cases} (\mathbf{y}^k, \mathbf{v}^k) \in \operatorname{gph}\mathcal{S}^{[\alpha\epsilon_k]}_{\alpha,\nabla\overline{f}+\partial\overline{g},\partial\overline{r}} = \operatorname{gph} T^{[\alpha\epsilon_k]}, \\ \theta_k\big\|\mathbf{v}^k\big\|^2 + \big\|\mathbf{v}^k + (\mathbf{y}^k - \mathbf{z}^k)\big\|^2 + 2\alpha\epsilon_k \le \sigma\big\|\mathbf{y}^k - \mathbf{z}^k\big\|^2, \\ \mathbf{z}^{k+1} = \mathbf{z}^k - (1+\theta_k)\mathbf{v}^k. \end{cases}$$

Remark 4. **(i)** *Let $\theta_k = 0$. $\alpha < 2/L$ can guarantee the global convergence of the original PPG algorithm, which largely expands the region $\alpha < 3/(2L)$ in (Ryu & Yin, 2017).*
(ii) *PPG algorithm has been shown to cover ADMM (Boyd et al., 2011) and Davis-Yin 3OS algorithm (Davis & Yin, 2015). Thus, they also fall into the VMOR-HPE framework.*

AFBAS Algorithm Let $A : \mathbb{X} \rightrightarrows \mathbb{X}$ be a maximally monotone operator, $M : \mathbb{X} \to \mathbb{X}$ be a linear operator, and $C : \mathbb{X} \to \mathbb{X}$ be a β-cocoercive operator with respect to $\|\cdot\|_P$ satisfying $\big\langle x - x', C(x) - C(x')\big\rangle \ge \beta\big\|C(x) - C(x')\big\|^2_{P^{-1}}$, respectively. The AFBAS algorithm solves problem (1) as below:

$$0 \in T(x) = (A + M + C)(x),\ x \in \mathbb{X}.$$

Let $S : \mathbb{X} \to \mathbb{X}$ be any self-adjoint positive definite linear operator and $K : \mathbb{X} \to \mathbb{X}$ be a skew adjoint operator, respectively. Denote $H = P + K$. Then, the AFBAS algorithm is defined as:

$$\begin{cases} \overline{x}^k := (H+A)^{-1}(H-M-C)x^k, \\ x^{k+1} := x^k + \alpha_k S^{-1}(H+M^*)(\overline{x}^k - x^k), \end{cases}$$

where $\alpha_k = \left[\lambda_k \|\overline{z}^k - z^k\|_P^2 \|\right] / \left[\|(H+M^*)(\overline{z}^k - z^k)\|_{S^{-1}}^2\right]$ and $\lambda_k \in [\underline{\lambda}, \overline{\lambda}] \leq [0, 2-1/(2\beta)]$. Throughout (Latafat & Patrinos, 2017), M is specified to a skew-adjoint linear operator, *i.e.*, $M^* = -M$.

Proposition 4. *Let $(x^k, \overline{x}^k)$ be the sequence generated by the AFBAS algorithm. Denote $\theta_k = \alpha_k - 1$, $v^k = (H + M^*)(x^k) - (H+M^*)(\overline{x}^k)$ and $\epsilon_k = \frac{\|\overline{z}^k - z^k\|_P^2}{4\beta}$. Then,*

$$\begin{cases} (\overline{x}^k, v^k) \in \operatorname{gph}(A+M+C)^{[\epsilon_k]}, \\ \theta_k \left\|S^{-1}v^k\right\|_S^2 + \left\|S^{-1}v + (\overline{x}^k - x^k)\right\|_S^2 + 2\epsilon \leq \sigma \left\|\overline{x}^k - x^k\right\|_S^2, \\ x^{k+1} := x^k - (1+\theta_k)S^{-1}v^k. \end{cases}$$

In (Latafat & Patrinos, 2017), a few new algorithms, such as forward-backward-forward splitting algorithm with only one evaluation of C, Douglas-Rachford splitting algorithm with an extra forward step, *etc*, are put forward based on the AFBAS algorithm. By Proposition 4, VMOR-HPE also covers these new splitting algorithms as special cases.

3.2.2. PRIMAL-DUAL ALGORITHMS

In this subsection, we focus on the existing primal-dual algorithms in the literature for solving the problem below:

$$\min f(x) + g(x) + h(Bx), \ x \in \mathbb{X}, \tag{18}$$

where $B: \mathbb{X} \to \mathbb{Y}$ is a linear operator, $g: \mathbb{X} \to (-\infty, +\infty]$ and $h: \mathbb{Y} \to (-\infty, +\infty]$ are closed proper convex functions, and $f: \mathbb{X} \to (-\infty, \infty)$ is a differentiable convex function satisfying $\|\nabla f(x) - \nabla f(x')\| \leq L\|x - x'\|$ for all $x, x' \in \mathbb{X}$. By introducing the dual variable $y \in \mathbb{Y}$ and denoting $\mathbb{Z} = \mathbb{X} \times \mathbb{Y}$, problem (18) can be formulated as:

$$0 \in T(z) = \begin{bmatrix} \partial g(x) \\ \partial h^*(y) \end{bmatrix} + \begin{bmatrix} \nabla f(x) + B^* y \\ -Bx \end{bmatrix}, z \in \mathbb{Z}. \tag{19}$$

Condat-Vu PDS Algorithm is proposed to solve problem (18) with the following iterations:

$$\begin{cases} \widetilde{x}^{k+1} := \operatorname{Prox}_{r^{-1}g}\left(x^k - r^{-1}\nabla f(x^k) - r^{-1}B^* y^k\right), \\ \widetilde{y}^{k+1} := \operatorname{Prox}_{s^{-1}h^*}\left(y^k + s^{-1}B(2\widetilde{x}^{k+1} - x^k)\right), \\ (x^{k+1}, y^{k+1}) := (x^k, y^k) + (1+\theta_k)\left((\widetilde{x}^{k+1}, \widetilde{y}^{k+1}) - (x^k, y^k)\right). \end{cases}$$

We denote $\mathcal{M}: \mathbb{Z} \to \mathbb{Z}$ as $\mathcal{M} = [r \ -B^*; -B \ s]$ and show that the Condat-Vu PDS algorithm is covered by VMOR-HPE.

Proposition 5. *Let $\{(x^k, y^k, \widetilde{x}^k, \widetilde{y}^k)\}$ be the sequence generated by the Condat-Vu PDS algorithm. Let $z^k = (x^k, y^k)$, $w^k = (\widetilde{x}^{k+1}, \widetilde{y}^{k+1})$. Parameters (r, s, θ_k) satisfy*

$$s - r^{-1}\|\mathcal{B}\|^2 > 0, \theta_k + L/[2(s - r^{-1}\|\mathcal{B}\|^2)] \leq \sigma. \tag{21}$$

Denote $v^k = \mathcal{M}(z^k - w^k)$, $\epsilon_k = L\|x^k - \widetilde{x}^{k+1}\|^2/4$. Then,

$$\begin{cases} v^k \in T^{[\epsilon_k]}(w^k), \\ \theta_k \left\|\mathcal{M}^{-1}v^k\right\|_{\mathcal{M}}^2 + \left\|\mathcal{M}^{-1}v^k + w^k - z^k\right\|_{\mathcal{M}}^2 + 2\epsilon_k \leq \sigma \left\|w^k - z^k\right\|_{\mathcal{M}}^2, \\ z^{k+1} = z^k - (1+\theta_k)\mathcal{M}^{-1}v^k. \end{cases}$$

Remark 5. **(i)** *The condition (21) is much more mild compared with $s - r^{-1}\|\mathcal{B}\|^2 > L/2, \theta_k + L/[2(s - r^{-1}\|\mathcal{B}\|^2)] < 1$ in (Condat, 2013; Vũ, 2013) and $s - r^{-1}\|\mathcal{B}\|^2 > L/2, \theta_k + L/[s - r^{-1}\|\mathcal{B}\|^2] < 1$ in (Chambolle & Pock, 2016).*
(ii) *The metric version of Condat-Vu PDS algorithm (Li & Zhang, 2016) with $(s = S, r = R)$ also falls into the VMOR-HPE framework by replacing condition (21) with $\|R^{-\frac{1}{2}}BS^{-\frac{1}{2}}\| < 1$ and $\theta_k + L/(2\lambda_{\min}(\mathcal{M})) \leq \sigma$.*
(iii) *If $f = 0$, the Condat-Vu PDS algorithm recovers PDHG algorithm (Chambolle & Pock, 2011) which is also covered by the VMOR-HPE framework.*

AFBAS-PD Algorithm Applying the AFBAS algorithm for (19) yields the **P**rimal-**D**ual (AFBAS-PD) algorithm:

$$\begin{cases} \overline{x}^k := \operatorname{Prox}_{\gamma_1 g}\left(x^k - \gamma_1 B^* y^k - \gamma_1 \nabla f(x^k)\right), \\ \overline{y}^k := \operatorname{Prox}_{\gamma_2 h^*}\left(y^k + \gamma_2 B((1-\theta)x^k + \theta\overline{x}^k)\right), \\ x^{k+1} := x^k + \alpha_k\left((\overline{x}^k - x^k) - \mu\gamma_1(2-\theta)B^*(\overline{y}^k - y^k)\right), \\ y^{k+1} := y^k + \alpha_k\left(\gamma_2(1-\mu)(2-\theta)B(\overline{x}^k - x^k) + (\overline{y}^k - y^k)\right), \end{cases}$$

where α_k is adaptively tuned and $(\gamma_1, \gamma_2, \theta, \mu)$ satisfy $\mu \in [0, 1]$, $\theta \in [0, \infty)$ and $\gamma_1^{-1} - \gamma_2\theta^2\|B\|^2/4 > L/4$.

Denote a linear operator $M: \mathbb{Z} \to \mathbb{Z}$ with $M = RS^{-1}$, where (R, S) are defined as $R = [\gamma_1^{-1} \ \ -B^*; \ \ (1-\theta)B \ \ \gamma_2^{-1}]$ and

$$S = \begin{bmatrix} 1 & -\mu\gamma_1(2-\theta)B^* \\ \gamma_2(1-\mu)(2-\theta)B & 1 \end{bmatrix}.$$

In addition, by (Horn & Johnson, 1990), it is easy to verify that M is a self-adjoint positive definite linear operator.

Proposition 6. *Let $\{(\overline{x}^k, \overline{y}^k, x^k, y^k)\}$ be the sequence generated by the AFBAS-PD algorithm. Denote $w^k = (\overline{x}^k, \overline{y}^k)$, $z^k = (x^k, y^k)$, $v^k = R(z^k - w^k)$, $\epsilon_k = L\|x^k - \overline{x}^k\|^2/4$, and $\theta_k = \alpha_k - 1$. Then, it holds that*

$$\begin{cases} v^k \in T^{[\epsilon_k]}(w^k), \\ \theta_k \left\|\mathcal{M}^{-1}v^k\right\|_{\mathcal{M}}^2 + \left\|\mathcal{M}^{-1}v^k + w^k - z^k\right\|_{\mathcal{M}}^2 + 2\epsilon_k \leq \sigma \left\|w^k - z^k\right\|_{\mathcal{M}}^2, \\ z^{k+1} = z^k - (1+\theta_k)\mathcal{M}^{-1}v^k. \end{cases}$$

The AFBAS-PD algorithm (Latafat & Patrinos, 2017) recovers: (i) the Condat-Vu PDS algorithm with an adaptive over-relaxed step-size if $\theta = 2$; (ii) the Combettes PDS algorithm if $\theta = 0$ and $\mu = \frac{1}{2}$; (iii) the MSS algorithm if $\theta = 0, \mu = 1/2$ and $h = 0$; (iv) the PAPC algorithm if $\theta = 1, \mu = 1$ and $f = 0$. Thus, they are also covered by VMOR-HPE.

To close this subsection 3.2.2, we make some comments on the PD3OS and PDFP algorithms which coincide with each other by (Tang & Wu, 2017). By Remark 4 and (OConnor & Vandenberghe, 2017), the PD3OS and PDFP algorithms are both covered by the algorithmic framework of VMOR-HPE.

4. PADMM-EBB Algorithm

The KKT generalized equation of problem (6) is defined as

$$T(z)=\begin{bmatrix}\partial g_1(x_1)\\ \vdots \\ \partial g_p(x_p)\\ b\end{bmatrix}+\begin{bmatrix}\nabla f_1(x)+\mathcal{A}_1 y\\ \vdots\\ \nabla f_p(x)+\mathcal{A}_p y\\ -\sum_{i=1}^p \mathcal{A}_i^* x_i\end{bmatrix},\ 0\in T(z), \quad (25)$$

where $\nabla f_i(x)$ is the i-th component of $\nabla f(x)$ and $y \in \mathbb{Y}$ is the Lagrange multiplier. Let $\mathbb{Z}=\mathbb{X}\times\mathbb{Y}$, $\mathbb{X}:=\mathbb{X}_1\times\cdots\times\mathbb{X}_p$, $x=(x_1,\ldots,x_p)\in\mathbb{X}$, $z=(x_1,\ldots,x_p,y)\in\mathbb{Z}$, and $\widehat{L}_{(\beta,\,x^k)}$ be the majorized augmented Lagrange function as

$$\widehat{L}_{(\beta_k,\,x^k)}(x,y) = f(x^k,x) + \left\langle \sum_{i=1}^p \mathcal{A}_i^* x_i - b, y\right\rangle \quad (26)$$
$$+\sum_{i=1}^p g_i(x_i)+\tfrac{\beta_k}{2}\left\| \sum_{i=1}^p \mathcal{A}_i x_i - b\right\|^2,$$

where $f(x^k,x) = f(x^k)+\langle\nabla f(x^k), x-x^k\rangle+\frac{1}{2}\|x-x^k\|^2_{\widehat{\Sigma}}$ and $\hat{\Sigma}$ is a self-adjoint positive semi-definite linear operator.

In the implementation of VMOR-HPE, generating (v^k, y^k, ϵ_k) satisfying (7a)-(7b) is equal to performing a non self-adjoint **P**roximal ADMM to problem (6) and $x^{k+1}:=x^k-(1+\theta_k)c_k\mathcal{M}_k^{-1}v^k$ in VMOR-HPE for problem (6) corresponds to performing an **E**xtra-gradient correction step to ensure the global convergence of PADMM. Additionally, $\mathcal{M}_k$ is determined by a **B**arzilai-**B**orwein line search technique to explore the curvature information of the KKT operator T. The PADMM-EBB is described in Algorithm 2.

Algorithm 2 PADMM-EBB Algorithm

Parameters: Given $\xi_k \geq 0$ satisfying $\sum_{i=1}^{\infty}\xi_i < \infty$, $\tau,\overline{\theta} > 0$, $-1 < \underline{\theta} < 0$, and $\overline{\sigma} \in [0,1)$. Choose a linear operator $\mathcal{M}_0 \succ 0$ and starting points $x^0\in\mathbb{X}$, $y^0\in\mathbb{Y}$.
for $k = 0,1,2,\ldots,$ **do**
  For $i = 1,2,\ldots,p$, $\widetilde{x}_i^{k+1}$ solves the inclusion as below
  $0\in\partial_{x_i}\widehat{L}_{(\beta_k,x^k)}(\ldots,\widetilde{x}_{i-1}^{k+1},x_i,x_{i+1}^k,\ldots,y^k)+P_i^k(x_i-x_i^k)$.
  $\widetilde{y}^{k+1} := y^k+\beta_k\big(\mathcal{A}_1^*\widetilde{x}_1^{k+1}+\mathcal{A}_2^*x_2^k+\ldots+\mathcal{A}_p^*x_p^k-b\big)$.
  Set $\theta_k\in[\theta_k^{\mathrm{fix}},\theta_k^{\mathrm{adap}}]$ with $\theta_k^{\mathrm{fix}}\in\big[\underline{\theta},\overline{\theta}_k\big]$ via (27b)-(27b).
  $z^{k+1}:=z^k+(1+\theta_k)\mathcal{M}_k^{-1}U_k(w^k-z^k)$, where (z^k,w^k) are defined as $z^k=(x^k,y^k)^\top$, $w^k=(\widetilde{x}^{k+1},\widetilde{y}^{k+1})^\top$.
  Update $\mathcal{M}_{k+1}^{-1}$=Diag$(M_1^{k+1},\cdots,M_p^{k+1},M_{p+1}^{k+1})$.
end for

In this algorithm, each M_i^{k+1} for $i=1,\ldots,p$ is defined as

$$M_i^{k+1}:=\min\left(\|\widetilde{x}_i^{k+1}-\widetilde{x}_i^k\|/\|s_{k+1}-s_k\|,(1+\xi_k)M_{i+1}^k\right),$$

where $s_{k+1} = (U^k(z^k-w^k))_i+\nabla f_i(\widetilde{x}_i^{k+1})-\nabla f_i(x_i^k)$. In addition, let $r_{k+1}=\beta_k^{-1}(y^k-\widetilde{y}^{k+1})+\sum_{i=2}^p\mathcal{A}_p^*(x_i^k-\widetilde{x}_i^{k+1})$. The metric M_{p+1}^{k+1} is defined as

$$M_{p+1}^{k+1}:=\min\left(\|\widetilde{y}^{k+1}-\widetilde{y}^k\|/\|r_{k+1}-r_k\|,(1+\xi_k)M_{p+1}^k\right).$$

Let $\mathcal{D}=\mathrm{Diag}(L_1\mathcal{I}\ \cdots\ L_p\mathcal{I}\ \ 0)$ and $\Gamma_k=U^k+(U^k)^*+(\overline{\sigma}-1)\mathcal{M}_k-\mathcal{D}/2$. Parameters $(\overline{\theta}_k,\theta_k^{\mathrm{adap}})$ are defined as

$$\begin{cases}\overline{\theta}_k=\max\left\{\theta \mid (\theta+1)(U^k)^*\mathcal{M}_k^{-1}U^k\preceq\Gamma_k\right\}, & (27a)\\ \theta_k^{\mathrm{adap}}=-1+\left\|z^k-w^k\right\|^2_{\Gamma_k}\Big/\left\|z^k-w^k\right\|^2_{(U^k)^*\mathcal{M}_k^{-1}U^k}. & (27b)\end{cases}$$

In addition, $P_i^k:\mathbb{X}_i\to\mathbb{X}_i$ for $i=1,2,\ldots,p$ are non self-adjoint linear operators, $\mathcal{T}_i=\widehat{\Sigma}_i+P_i^k+\beta_k\mathcal{A}_i\mathcal{A}_i^*$, and $U_k:\mathbb{Z}\to\mathbb{Z}$ is a block linear operator defined as below

$$U^k=\begin{pmatrix}\widehat{\Sigma}_1+P_1^k & 0 & \ldots & 0 & 0\\ 0 & \mathcal{T}_2 & \ldots & 0 & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & \beta_k\mathcal{A}_p\mathcal{A}_2^* & \cdots & \mathcal{T}_p & 0\\ 0 & \mathcal{A}_2^* & \cdots & \mathcal{A}_n^* & \beta_k^{-1}\mathcal{I}\end{pmatrix}.$$

Remark 6. *To ensure $1+\theta_k>0$, P_i^k should be chosen to make $U^k+(U^k)^*\succ\mathcal{D}/2$. In addition, the non self-adjoint linear operator P_i^k in inclusion with respect to $\widetilde{x}_i^{k+1}$ is chosen to approximate $\beta_k\mathcal{A}_i\mathcal{A}_i^*+\widehat{\Sigma}$ more tightly and make the inclusion easier to solve than the common settings.*

Theorem 4. *Let $(\widetilde{x}^k,\widetilde{y}^k,x^k,y^k)$ be the sequence generated by the PADMM-EBB algorithm. Denote $v^k=U^k(z^k-w^k)$, $\epsilon_k=\|x^k-\widetilde{x}^{k+1}\|_{\mathcal{D}}/4$ and operator T as (25). Then, it holds*

$$\begin{cases}v^k\in T^{[\epsilon_k]}(w^k),\\ \theta_k\left\|\mathcal{M}_k^{-1}v^k\right\|^2_{\mathcal{M}_k}+\left\|\mathcal{M}_k^{-1}v^k+w^k-z^k\right\|^2_{\mathcal{M}_k}+2\epsilon_k\leq\sigma\left\|w^k-z^k\right\|^2_{\mathcal{M}_k},\\ z^{k+1}=z^k-(1+\theta_k)\mathcal{M}_k^{-1}v^k.\end{cases}$$

Besides, **(i)** *$(x^k,\widetilde{x}^k)$ and $(y^k,\widetilde{y}^k)$ converge to x^∞ and y^∞ belonging to the primal-dual solution set of problem (6).*
(ii) *There exits an integer $\overline{k}\in\{1,2,\ldots,k\}$ such that*

$$\sum_{i=1}^p\mathrm{dist}\big((\partial g_i+\nabla f_i)(\widetilde{x}^{\overline{k}})+\mathcal{A}_i\widetilde{y}^{\overline{k}},0\big)+\Big\|b-\sum_{i=1}^p\mathcal{A}_i^*\widetilde{x}_i^{\overline{k}}\Big\|\leq\mathcal{O}\Big(\frac{1}{\sqrt{k}}\Big).$$

(iii) *Let $\alpha_i=1$* or *i. There exists $0\leq\overline{\epsilon}_k^{x_i}\leq\mathcal{O}(\frac{1}{k})$ such that*

$$\sum_{i=1}^p\mathrm{dist}\big((\partial g_i+\nabla f_i)_{\overline{\epsilon}_k^{x_i}}(\overline{x}^k)+\mathcal{A}_i\overline{y}^k,0\big)+\Big\|b-\sum_{i=1}^p\mathcal{A}_i^*\overline{x}_i^k\Big\|\leq\mathcal{O}\Big(\frac{1}{k}\Big),$$

where $\overline{x}^k=\frac{\sum_{i=1}^k(1+\theta_i)\alpha_i\widetilde{x}^{i+1}}{\sum_{i=1}^k(1+\theta_i)\alpha_i}$ and $\overline{y}^k=\frac{\sum_{i=1}^k(1+\theta_i)\alpha_i\widetilde{y}^{i+1}}{\sum_{i=1}^k(1+\theta_i)\alpha_i}$.
(iv) *If T satisfies metric subregularity at $\big((x^\infty,y^\infty),0\big)\in$ gph$T$ with modulus $\kappa>0$. Then, there exits $\overline{k}>0$ such that*

$$\mathrm{dist}_{\mathcal{M}_{k+1}}\big((x^{k+1},y^{k+1}),T^{-1}(0)\big)$$
$$\leq\Big(1-\frac{\varrho_k}{2}\Big)\mathrm{dist}_{\mathcal{M}_k}\big((x^k,y^k),T^{-1}(0)\big),\ \forall k\geq\overline{k},$$

where $\varrho_k=\frac{(1-\sigma)(1+\theta_k)}{\left(1+\kappa\sqrt{\frac{\Xi\overline{\omega}}{\underline{\omega}}}\right)^2\left(1+\sqrt{\sigma+\frac{4\max\{-\theta_k,0\}}{(1+\theta_k)^2}}\right)^2}\in(0,1)$.

Remark 7. *By Proposition 3, the constants in $\mathcal{O}(\frac{1}{\sqrt{k}})$ pointwise iteration complexity and $\mathcal{O}(\frac{1}{k})$ weighted iteration complexity both depend merely on the primal-dual solution set of problem (6) without requiring the boundedness of $(\mathbb{X},\mathbb{Y})$.*

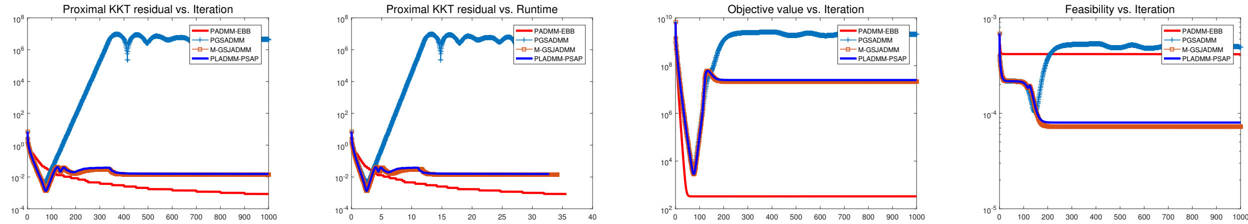

Figure 1. The above four figures illustrate the proximal KKT residual vs. iteration, proximal KKT residual vs. runtime, objective value vs. iteration, and feasibility vs. iteration on the synthetic dataset with parameters $(\lambda, \mu, \gamma) = (10^3, 10^4, 10^4)$, respectively.

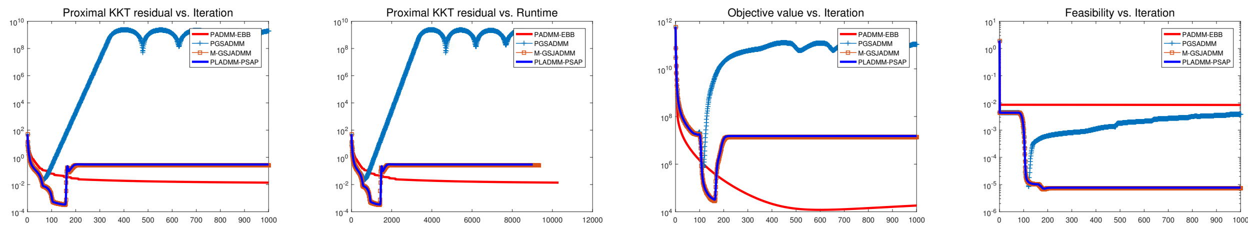

Figure 2. The above four figures illustrate the proximal KKT residual vs. iteration, proximal KKT residual vs. runtime, objective value vs. iteration, and feasibility vs. iteration on the real dataset PIE_pose27 with parameters $(\lambda, \mu, \gamma) = (10^3, 10^4, 10^4)$, respectively.

4.1. Experiments

We verify the efficacy of the proposed PADMM-EBB algorithm by solving the nonnegative dual graph regularized low-rank representation problem (Yin et al., 2015) as below:

$$\begin{aligned} &\min \|Z\|_* + \|G\|_* + \lambda\|E\|_1 + \frac{\mu}{2}\|Z\|_{L_Z}^2 + \frac{\gamma}{2}\|G\|_{L_G}^2 \\ &\text{s.t. } X = XZ + GX + E, Z \geq 0,\ G \geq 0, \end{aligned} \tag{29}$$

where (X, L_Z, L_G) are given parameters and (λ, μ, γ) are the parameters to control the level of the reconstruction error and graph regularization. It is obvious that problem (29) can be formulated as problem (6) with f being quadratic and $p = 3$. Define the proximal KKT residual of problem (6) as

$$R(z) = \begin{bmatrix} x_1 - \text{Prox}_{g_1}(x_1 - \nabla f_1(x) - \mathcal{A}_1 y) \\ \vdots \\ x_p - \text{Prox}_{g_p}(x_p - \nabla f_p(x) - \mathcal{A}_p y) \\ b - \sum_{i=1}^p \mathcal{A}_i^* x_i \end{bmatrix}. \tag{30}$$

The proximal KKT residual, as a complete characterization of optimality for constrained optimization, simultaneously evaluates the performance in terms of the feasibilities of primal-dual equalities, violation of nonnegativity, and complementarity condition of nonnegativity for problem (29).

We compare PADMM-EBB with three existing state-of-the-art primal-dual algorithms which are suitable for problem (6), namely PLADMM-PSAP (Liu et al., 2013; Lin et al., 2015), PGSADMM and M-GSJADMM (Lu et al., 2017) in terms of the objective value, feasibility, and proximal KKT residual $R(z)$ over iteration and runtime. Notably, PGSADMM and PADMM-EBB are performed with a full Gauss-Seidel updating for the majorized augmented Lagrange function (26). We conduct experiments on a synthetic dataset $X = \text{randn}(200, 200)$ and a real dataset PIE_pose27[1]. Graph matrices (L_Z, L_G) and parameters $(\lambda, \mu, \gamma) = (10^3, 10^4, 10^4)$ are directly borrowed from (Yin et al., 2015). In the implementation, we strictly follow the advice in (Lin et al., 2015; Lu et al., 2017) to adaptively tune the penalty parameter β_k for PLADMM-PSAP, PGSADMM and M-GSJADMM.

According to Figures 1 and 2, we know that PADMM-EBB is slightly better than PLADMM-PSAP, PGSADMM and M-GSJADMM in terms of the proximal KKT residual and the objective value due to the efficient block Barzilai-Borwein technique, which exploits the curvature information of the KKT generalized equation (25) and the Gauss-Seidel updating for primal variables. PGSADMM, PLADMM-PSAP and M-GSJADMM have lower feasibilities since their penalty parameters β_k are increasing as iterations proceed to force the equality constraint to hold. More experimental results are placed into the supplementary material.

5. Conclusions

In this paper, we proposed a novel algorithmic framework of **V**ariable **M**etric **O**ver-**R**elaxed **H**ybrid **P**roximal **E**xtra-gradient (VMOR-HPE) method and established its global convergence, iteration complexities, and local linear convergence rate. This framework covers a large class of primal and primal-dual algorithms as special cases, and serves as a powerful analysis technique for characterizing their convergences. In addition, we applied the VMOR-HPE framework to linear equality constrained optimization, yielding a new convergent primal-dual algorithm. The numerical experiments on synthetic and real datasets demonstrate the efficacy of the proposed algorithm.

[1] http://dengcai.zjulearning.org:8081/Data/FaceDataPIE.html

References

Barzilai, J. and Borwein, J. M. Two-point step size gradient methods. *IMA journal of numerical analysis*, 8(1):141–148, 1988.

Bauschke, H. H. and Combettes, P. L. *Convex analysis and monotone operator theory in Hilbert spaces*, volume 408. Springer, 2017.

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

Briceño Arias, L. M. and Combettes, P. L. A monotone+ skew splitting model for composite monotone inclusions in duality. *SIAM Journal on Optimization*, 21(4):1230–1250, 2011.

Briceño-Arias, L. M. Forward-Douglas-Rachford splitting and forward-partial inverse method for solving monotone inclusions. *Optimization*, 64(5):1239–1261, 2015a.

Briceño-Arias, L. M. Forward-partial inverse-forward splitting for solving monotone inclusions. *Journal of Optimization Theory and Applications*, 166(2):391–413, 2015b.

Briceño-Arias, L. M. and Davis, D. Forward-backward-half forward algorithm with non self-adjoint linear operators for solving monotone inclusions. *SIAM Journal on Optimization*, 2018.

Burachik, R. S., Iusem, A. N., and Svaiter, B. F. Enlargement of monotone operators with applications to variational inequalities. *Set-Valued and Variational Analysis*, 5(2):159–180, 1997.

Burachik, R. S., Sagastizábal, C. A., and Svaiter, B. F. ε-enlargements of maximal monotone operators: Theory and applications. In *Reformulation: nonsmooth, piecewise smooth, semismooth and smoothing methods*, pp. 25–43. Springer, 1998.

Burke, J. V. and Qian, M. A variable metric proximal point algorithm for monotone operators. *SIAM Journal on Control and Optimization*, 37(2):353–375, 1999.

Chambolle, A. and Pock, T. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

Chambolle, A. and Pock, T. On the ergodic convergence rates of a first-order primal–dual algorithm. *Mathematical Programming*, 159(1-2):253–287, 2016.

Chen, P., Huang, J., and Zhang, X. A primal-dual fixed point algorithm for minimization of the sum of three convex separable functions. *Fixed Point Theory and Applications*, 1(1):1–18, 2016.

Combettes, P. L. and Pesquet, J. C. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*, pp. 185–212. Springer, 2011.

Combettes, P. L. and Pesquet, J. C. Primal-dual splitting algorithm for solving inclusions with mixtures of composite, lipschitzian, and parallel-sum type monotone operators. *Set-Valued and variational analysis*, 20(2):307–330, 2012.

Condat, L. A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. *Journal of Optimization Theory and Applications*, 158(2):460–479, 2013.

Corman, E. and Yuan, X. A generalized proximal point algorithm and its convergence rate. *SIAM Journal on Optimization*, 24(4):1614–1638, 2014.

Cui, Y. *Large scale composite optimization problems with coupled objective functions: theory, algorithms and applications*. PhD thesis, 2016.

Davis, D. and Yin, W. A three-operator splitting scheme and its optimization applications. *Set-Valued and Variational Analysis*, pp. 1–30, 2015.

Dontchev, A. L. and Rockafellar, R. T. *Implicit Functions and Solution Mappings : a view from variational analysis*. Springer US, 2009.

Drori, Y., Sabach, S., and Teboulle, M. A simple algorithm for a class of nonsmooth convex–concave saddle-point problems. *Operations Research Letters*, 43(2):209–214, 2015.

Eckstein, J. and Bertsekas, D. P. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1):293–318, 1992.

Güler, O. On the convergence of the proximal point algorithm for convex minimization. *SIAM Journal on Control and Optimization*, 29(2):403–419, 1991.

Horn, R. A. and Johnson, C. R. *Matrix analysis*. Cambridge university press, 1990.

Korpelevich, G. M. Extragradient method for finding saddle points and other problems. volume 13, pp. 35–49. ME SHARPE INC 80 BUSINESS PARK DR, ARMONK, NY 10504, 1977.

Latafat, P. and Patrinos, P. Asymmetric forward–backward–adjoint splitting for solving monotone inclusions involving three operators. *Computational Optimization and Applications*, 68(1):57–93, 2017.

Li, Q. and Zhang, N. Fast proximity-gradient algorithms for structured convex optimization problems. *Applied and Computational Harmonic Analysis*, 41(2):491–517, 2016.

Lin, Z., Liu, R., and Li, H. Linearized alternating direction method with parallel splitting and adaptive penalty for separable convex programs in machine learning. *Machine Learning*, 2(99):287–325, 2015.

Liu, R., Lin, Z., and Su, Z. Linearized alternating direction method with parallel splitting and adaptive penalty for separable convex programs in machine learning. In *Asian Conference on Machine Learning*, pp. 116–132, 2013.

Lu, C., Feng, J., Yan, S., and Lin, Z. A unified alternating direction method of multipliers by majorization minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

Minty, G. J. Monotone (nonlinear) operators in hilbert space. *Duke Mathematical Journal*, 29(3):341–346, 1962.

Monteiro, R. D. C. and Svaiter, B. F. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. *SIAM Journal on Optimization*, 20 (6):2755–2787, 2010.

OConnor, D. and Vandenberghe, L. On the equivalence of the primal-dual hybrid gradient method and Douglas-Rachford splitting. 2017.

Parente, L. A., Lotito, P. A., and Solodov, M. V. A class of inexact variable metric proximal point algorithms. *SIAM Journal on Optimization*, 19(1):240–260, 2008.

Passty, G. B. Ergodic convergence to a zero of the sum of monotone operators in hilbert space. *Journal of Mathematical Analysis and Applications*, 72(2):383–390, 1979.

Raguet, H., Fadili, J., and Peyré, G. A generalized forward-backward splitting. *SIAM Journal on Imaging Sciences*, 6(3):1199–1226, 2013.

Robinson, S. M. Some continuity properties of polyhedral multifunctions. *Mathematical Programming at Oberwolfach*, pp. 206–214, 1981.

Rockafellar, R. T. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14(5):877, 1976.

Rockafellar, R. T. and Wets, R. J. B. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

Ryu, E. K. and Yin, W. Proximal-proximal-gradient method. *arXiv preprint arXiv:1708.06908*, 2017.

Shen, L. Over relaxed hybrid proximal extragradient algorithm and its application to several operator splitting methods. *Journal of Mathematical Analysis and Applications*, 2(448):727–749, 2017.

Shen, L. and Pan, S. Linear convergence of the generalized PPA and several splitting methods for the composite inclusion problem. *arXiv preprint arXiv:1508.05156*, 2015.

Solodov, M. V. and Svaiter, B. F. A hybrid approximate extragradient–proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Analysis*, 7(4):323–345, 1999.

Svaiter, B. F. A family of enlargements of maximal monotone operators. *Set-Valued and Variational Analysis*, 8(4): 311–328, 2000.

Svaiter, B. F. A unified framework for some inexact proximal point algorithms. *Numerical Functional Analysis & Optimization*, 22(7):1013–1035, 2001.

Svaiter, B. F. A class of fejér convergent algorithms, approximate resolvents and the hybrid proximal-extragradient method. *Journal of Optimization Theory and Applications*, 162(1):133–153, 2014.

Tang, Y. C. and Wu, G. R. A general framework for solving convex optimization problems involving the sum of three convex functions. *arXiv preprint arXiv:1705.06164*, 2017.

Tao, M. and Yuan, X. On the optimal linear convergence rate of a generalized proximal point algorithm. *Journal of Scientific Computing*, pp. 1–25, 2017. doi: 10.1007/s10915-017-0477-9.

Tseng, P. A modified forward-backward splitting method for maximal monotone mappings. *SIAM Journal on Control and Optimization*, 38(2):431–446, 2000.

Vũ, B. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Advances in Computational Mathematics*, 38(3):667–681, 2013.

Yan, M. A new primal–dual algorithm for minimizing the sum of three functions with a linear operator. *Journal of Scientific Computing*, pp. 1–20, 2018.

Yin, M., Gao, J., Lin, Z., Shi, Q., and Guo, Y. Dual graph regularized latent low-rank representation for subspace clustering. *IEEE Transactions on Image Processing*, 24 (12):4918–4933, 2015.