Journal of Convex Analysis
Volume 16 (2009), No. 3&4, 857–880

A Class of Variable Metric Decomposition Methods for Monotone Variational Inclusions

P. A. Lotito
*CONICET, Departamento de Matemática,
FCEIA, UNR, PLADEMA, UNICEN, Argentina*
plotito@exa.unicen.edu.ar

L. A. Parente
*CONICET, Departamento de Matemática,
FCEIA, UNR, Argentina*
lparente@fceia.unr.edu.ar

M. V. Solodov*
*IMPA – Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110,
Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil*
solodov@impa.br

Dedicated to Stephen Simons on the occasion of his 70th birthday.

Received: May 16, 2008

We extend the general decomposition scheme of [32], which is based on the hybrid inexact proximal point method of [38], to allow the use of variable metric in subproblems, along the lines of [23]. We show that the new general scheme includes as special cases the splitting method for composite mappings [25] and the proximal alternating directions method [13, 17] (in addition to the decomposition methods of [10, 42] that were already covered by [32]). Apart from giving a unified insight into the decomposition methods in question and openning the possibility of using variable metric, which is a computationally important issue, this development also provides linear rate of convergence results not previously available for splitting of composite mappings and for the proximal alternating directions methods.

Keywords: Proximal point methods, variable metric, maximal monotone operator, variational inclusion, splitting, decomposition

1. Introduction

We consider the classical problem of finding a zero of a maximal monotone operator $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, i.e., find $z \in \mathbb{R}^n$ such that

$$0 \in T(z). \tag{1}$$

As is well known, a wide variety of problems such as convex optimization, min-max problems, and monotone variational inequalities over convex sets, fall within this general

*This author is supported in part by CNPq Grants 301508/2005-4 and 471267/2007-4, by PRONEX–Optimization, and by FAPERJ.

ISSN 0944-6532 / $ 2.50

framework, see, e.g., [31]. Our central interest in this paper is the situation when the given operator T has some separable structure. In such cases, decomposition methods come into play. Many of those methods (e.g., [20, 40, 12, 41, 43, 32, 22]) are explicitly or implicitly derived from the proximal point algorithm (e.g., [21, 30, 18]) for solving (1).

Given $z^k \in \mathbb{R}^n$, the current approximation to a solution of (1), the proximal point method obtains the new iterate as the solution of the subproblem

$$0 \in c_k T(z) + z - z^k,$$

which can be stated as

$$\begin{cases} v \in T(z), \\ 0 = c_k v + z - z^k, \end{cases}$$

where $c_k > 0$ is the regularization parameter. To handle approximate solutions, which is a typical practical requirement, it is useful to relax both the inclusion and the equation in the above system, and to employ constructive relative error criteria to control the quality of approximation. One development in this direction is the Hybrid Inexact Proximal Point Method (HIPPM) [38] (see also related methods in [35, 34] and applications of HIPPM to Newton, bundle, and decomposition methods in [33, 37, 36, 39, 32, 22, 5]). In order to get more efficient algorithms, it is also attractive to allow for the use of a variable metric (or preconditioning), see [2, 26, 19, 9] for the special case where T is the subdifferential of a convex function, and [6, 7, 8, 23] for the general case. The variable metric HIPPM (VMHIPPM) of [23] combines both the use of variable metric and of relative error tolerance, and it is the following procedure.

Consider the *generalized* proximal subproblem

$$0 \in c_k M_k T(z) + z - z^k, \tag{2}$$

where M_k is a symmetric positive definite matrix (it is sometimes convenient to keep separated the c_k parameter). Given the error tolerance (relaxation) parameter $\sigma_k \in [0, 1)$, an iteration of VMHIPPM consists in finding $\hat{v}^k \in \mathbb{R}^n$, $\hat{z}^k \in \mathbb{R}^n$ and $\varepsilon_k \geq 0$ such that

$$\begin{cases} \hat{v}^k \in T^{\varepsilon_k}(\hat{z}^k), \\ \delta^k = c_k M_k \hat{v}^k + \hat{z}^k - z^k, \end{cases} \tag{3}$$

and

$$\|\delta^k\|^2_{M_k^{-1}} + 2c_k\varepsilon_k \leq \sigma_k^2 \left(\|c_k M_k \hat{v}^k\|^2_{M_k^{-1}} + \|\hat{z}^k - z^k\|^2_{M_k^{-1}} \right), \tag{4}$$

where by $\|\cdot\|_M$ we denote the norm induced by a symmetric positive definite matrix M, i.e.,

$$\|z\|_M = \sqrt{\langle z, Mz\rangle},$$

and the inclusion in (3) is relaxed by using the ε-enlargement of a maximal monotone operator T (see, e.g., [3, 4]), defined as

$$T^\varepsilon(z) := \{v \in \mathbb{R}^n \mid \langle w - v, y - z\rangle \geq -\varepsilon,\ \forall y \in \mathbb{R}^n,\ \forall w \in T(y)\},\ \varepsilon \geq 0.$$

The enlargement above can be seen as an outer approximation of T, as it holds that $T^0 \equiv T$ and $T(z) \subseteq T^\varepsilon(z)$, for any $z \in \mathbb{R}^n$ and any $\varepsilon \geq 0$. If f is a proper closed convex function, then $\partial_\varepsilon f(z) \subseteq (\partial f)^\varepsilon(z)$, where $\partial_\varepsilon f$ is the usual ε-subdifferential of f.

Note that if $\sigma_k = 0$ is chosen in (4) then the exact solution of (2) is obtained. Also, it should be noted that one can check the approximation criterion (4) without having to invert the matrix M_k, see [23].

Having computed the objects satisfying (3) and (4), the next iterate is then obtained by

$$z^{k+1} = z^k - \tau_k a_k M_k \hat{v}^k, \qquad \tau_k \in (0,2), \qquad a_k = \frac{\langle \hat{v}^k, z^k - \hat{z}^k \rangle - \varepsilon_k}{\|M_k \hat{v}^k\|^2_{M_k^{-1}}}. \tag{5}$$

If the approximation rule (4) is replaced by the more stringent one:

$$\|\delta^k\|^2_{M_k^{-1}} + 2c_k\varepsilon_k \le \sigma_k^2 \|\hat{z}^k - z^k\|^2_{M_k^{-1}}, \tag{6}$$

then there exists $\tau_k \in (0,2)$ such that $\tau_k a_k = c_k$ ([23, Proposition 3.1]), and we can take the next iterate as

$$z^{k+1} = z^k - c_k M_k \hat{v}^k.$$

Convergence of the method outlined above to an element of $T^{-1}(0) \neq \emptyset$ is guaranteed under some mild conditions imposed on the choice of the matrices M_k [23, Theorem 4.2]. Moreover, if T^{-1} satisfies a certain Lipschitzian property at zero (a condition which does not imply uniqueness of the solution) then the linear rate of convergence is obtained [23, Theorem 4.4]. The advantage of employing variable metric was illustrated in [23] in the context of a proximal Newton method. In this paper, introducing variable metric will also allow us to treat splitting of composite mappings [25] and proximal alternating directions algorithms [13, 17].

Let us now go back to the discussion of variational inclusions with structure. Operators of the form

$$T(x,y) = F(x,y) \times [G(x,y) + H(y)],$$

where $F : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ and $H : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ are maximal monotone, and $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ is Lipschitz-continuous, have been considered in [42, 32]. The decomposition scheme presented in [32] is the following iterative procedure, which is derived from HIPPM (i.e., from the inexact proximal method outlined above, with the choice of $M_k = I$). Given $(x^k, y^k) \in \mathbb{R}^n \times \mathbb{R}^m$, first a *forward-backward splitting* step (e.g., [20, 41, 11]) is performed with the x-part fixed:

$$\hat{y}^k = \left(I + c_k[H(\cdot) + G_1^k(\cdot)]\right)^{-1} \left(I - c_k[G(x^k,\cdot) - G_1^k(\cdot)]\right)(y^k), \tag{7}$$

where G_1^k is some adequate Lipschitz-continuous splitting function. To clarify the nature of this step and some options concerning the choice of G_1^k, suppose that H is the normal cone mapping associated to a closed convex set $C \subset \mathbb{R}^m$. In that case, (7) gives

$$\hat{y}^k = P_C\left(y^k - c_k(G_1^k(\hat{y}^k) - G_1^k(y^k) + G(x^k, y^k))\right).$$

If we take $G_1^k \equiv 0$, then

$$\hat{y}^k = P_C\left(y^k - c_k G(x^k, y^k)\right),$$

which is the standard projection step. If we take $G_1^k \equiv G(x^k, \cdot)$, then

$$\hat{y}^k = P_C\left(y^k - c_k G(x^k, \hat{y}^k)\right),$$

which is an implicit (proximal) step. Inbetween there are various intermediate (in terms of computational cost) choices of G_1^k. Which particular G_1^k should be used depends on the structure of $G(x^k,\cdot)$ and of $H(\cdot)$, see [42] for a more detailed discussion and examples.

The forward-backward splitting step is followed by a hybrid inexact proximal step with the y-part fixed, that consists in finding a triplet $(\hat{u}^k,\hat{x}^k,\varepsilon_k)\in\mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}_+$ such that

$$\begin{cases}\hat{u}^k\in F^{\varepsilon_k}(\hat{x}^k,\hat{y}^k),\\ r^k=c_k\hat{u}^k+\hat{x}^k-x^k,\end{cases}$$

where the enlargement F^{ε_k} is in x for $\hat{y}^k$ fixed, and

$$\|r^k\|^2+\|s^k\|^2+2c_k\varepsilon_k\le\sigma_k^2\left(\|c_k\hat{u}^k\|^2+\|c_k\hat{w}^k\|^2+\|\hat{x}^k-x^k\|^2+\|\hat{y}^k-y^k\|^2\right),$$

with $\hat{w}^k=G(\hat{x}^k,\hat{y}^k)+\hat{h}^k$ and $s^k=c_k\hat{w}^k+\hat{y}^k-y^k$, where $\hat{h}^k$ is the element of $H(\hat{y}^k)$ computed in the forward-backward splitting step. The next iterates are obtained by setting

$$x^{k+1}=x^k-\tau_k a_k\hat{u}^k,$$
$$y^{k+1}=y^k-\tau_k a_k\hat{w}^k,$$

where

$$a_k=\frac{\langle(\hat{u}^k,\hat{w}^k),(x^k-\hat{x}^k,y^k-\hat{y}^k)\rangle-\varepsilon_k}{\|\hat{u}^k\|^2+\|\hat{w}^k\|^2},\qquad \tau_k\in(0,2).$$

The decomposition framework outlined above contains some instances of the scheme described in [42], as well as the proximal-based decomposition for convex minimization of [10], which we state next as an example that can be helpful for clarifying the nature of the general scheme. We refer the reader to [32] for justification of the relation in question. Consider the problem

$$\begin{array}{ll}\text{minimize} & f_1(x_1)+f_2(x_2)\\ \text{subject to} & Ax_1-Bx_2=0,\end{array}\tag{8}$$

where f_1 and f_2 are closed proper convex functions on $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, and $A:\mathbb{R}^n\to\mathbb{R}^l$, $B:\mathbb{R}^m\to\mathbb{R}^l$ are linear operators (matrices of appropriate dimensions). The method of [10] applies proximal point iterations to the subdifferential of the Lagrangian function $L(x_1,x_2,y)=f_1(x_1)+f_2(x_2)+\langle y,Ax_1-Bx_2\rangle$, alternately fixing the variables or the multipliers. Specifically, given some $(x_1^k,x_2^k,y^k)\in\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^l$, the method performs the following updates:

$$\begin{aligned}\hat{y}^k&=y^k+c_k\left(Ax_1^k-Bx_2^k\right),\\ x_1^{k+1}&=\arg\min_{x_1\in\mathbb{R}^n}\left\{f_1(x_1)+\langle A^\top\hat{y}^k,x_1\rangle+\frac{1}{2c_k}\|x_1-x_1^k\|^2\right\},\\ x_2^{k+1}&=\arg\min_{x_2\in\mathbb{R}^m}\left\{f_2(x_2)-\langle B^\top\hat{y}^k,x_2\rangle+\frac{1}{2c_k}\|x_2-x_2^k\|^2\right\},\\ y^{k+1}&=y^k+c_k\left(Ax_1^{k+1}-Bx_2^{k+1}\right).\end{aligned}\tag{9}$$

This method has some nice features not shared by previous decomposition algorithms, when the latter are applied to (8). In particular, the minimization is carried out separately in the spaces $\mathbb{R}^n$ and $\mathbb{R}^m$, and the two minimization problems decompose further according to the separable structure of the functions f_1 and f_2. Other methods do not achieve such a fine degree of decomposition for the given problem, see [42] for a more detailed discussion.

In this paper, we combine the ideas of decomposition from [32] with the use of variable metric from [23]. We emphasize that this development is worthwhile for a number of reasons. Apart from variable metrics and preconditioning being important in practice, we note that in this context it appears useful also for theoretical considerations. Specifically, splitting of composite mappings [25] and proximal alternating directions methods [13, 17] could not be analyzed within the previous decomposition framework of [32] (i.e., without introducing variable metric). Among other things, this analysis allows us to obtain rate of convergence results for the methods in consideration, which were not available previously, and to present a unified view of those seemingly different techniques. As an additional enhancement with respect to [32], we shall allow inexact computation in the forward-backward step.

We next introduce our notation. By $\mathcal{M}^n_{++}$ we denote the space of symmetric positive definite matrices, with the partial order $\preceq$ given by

$$A \preceq B \Leftrightarrow B - A \text{ is a positive semidefinite matrix.}$$

For $M \in \mathcal{M}^n_{++}$, $\lambda_{min}(M)$ and $\lambda_{max}(M)$ stand for the minimal and the maximal eigenvalues of M, respectively. For any $A \preceq B$, it holds that $\|z\|_A \leq \|z\|_B$. In particular, if

$$0 < \lambda_l \leq \lambda_{min}(M) \leq \lambda_{max}(M) \leq \lambda_u,$$

then for any $x \in \mathbb{R}^n$ it holds that

$$\lambda_l \|x\|^2 \leq \|x\|^2_M \leq \lambda_u \|x\|^2, \qquad \frac{1}{\lambda_u}\|x\|^2 \leq \|x\|^2_{M^{-1}} \leq \frac{1}{\lambda_l}\|x\|^2. \tag{10}$$

By $\langle x, y\rangle$ we denote the usual inner product between $x, y \in \mathbb{R}^n$. For a matrix $M \in \mathcal{M}^n_{++}$, we denote $\langle x, y\rangle_M = \langle Mx, y\rangle$. For a closed convex set $\Omega \subseteq \mathbb{R}^n$ and a matrix $M \in \mathcal{M}^n_{++}$, the "skewed" projection operator onto Ω under the matrix M is given by

$$P_{\Omega,M}(z) = \arg\min_{x\in\Omega} \frac{1}{2}\langle x - z, M(x - z)\rangle = \arg\min_{x\in\Omega} \frac{1}{2}\|x - z\|^2_M,$$

i.e., it is the projection operator with respect to the norm $\|\cdot\|_M$. The associated distance from $z \in \mathbb{R}^n$ to Ω is defined as

$$\text{dist}(z, \Omega)_M = \|z - P_{\Omega,M}(z)\|.$$

2. Variable Metric Hybrid Proximal Decomposition Method

Consider problem (1), where T has the following structure:

$$T : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m, \qquad T(z) = F(x, y) \times [G(x, y) + H(y)], \tag{11}$$

$F : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^n$, $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$, $H : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$, and we set $z = (x, y) \in \mathbb{R}^n \times \mathbb{R}^m$.

We make the following standing assumptions:

A1 G is a (single-valued) continuous function.
A2 H is maximal monotone.
A3 The mapping $(x, y) \mapsto F(x, y) \times G(x, y)$ is maximal monotone.
A4 $\mathrm{dom} H \subset \mathrm{rint}\{y \in \mathbb{R}^m \mid \exists x \in \mathbb{R}^n \text{ s.t. } F(x, y) \times G(x, y) \neq \emptyset\}$.

Under the stated assumptions, it follows from [28] that T is maximal monotone, and further that the mapping $x \to F(x, y)$ is also maximal monotone for any fixed $y \in \mathrm{dom} H$ [42, Lemma 2.1].

We are now in position to present our algorithm. In essence, it is the hybrid proximal decomposition method of [32], already explained above, with the following extensions. Variable metric is introduced in both the forward-backward splitting and proximal steps, and the former (in addition to the latter) is also allowed to be computed inexactly. We note that in general, in decomposition methods of this nature the regularization parameter c_k has to be sufficiently small, see [10, 42, 32]. The value of c_k is either determined by a suitable linesearch procedure or set according to some heuristic considerations. This is accounted for by the comment at the end of the proximal step of Algorithm 2.1 – if solution of the proximal subproblem does not satisfy the required criteria, the value of c_k has to be reduced. In the variable metric inexact setting of Algorithm 2.1, there also other parameters that affect the quality of solution of the proximal subproblem (specifically, the chosen metric Q_k and the error in the forward-backward splitting step e^k). Adjusting those may be enough without decreasing c_k. In any case, Theorem 2.2 below shows that appropriate values of c_k guarantee solution with the needed properties.

Algorithm 2.1 (VMHPDM).
Initialization: *Choose* $(x^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^m$ *and* $\theta \in (0, 1)$. *Set* $k := 0$.

Inexact Forward-Backward Splitting Step: *Choose a continuous monotone function* $G_1^k : \mathbb{R}^m \to \mathbb{R}^m$, *a scalar* $c_k > 0$ *and a symmetric* $m \times m$ *positive definite matrix* Q_k. *Compute* $\hat{y}^k \in \mathbb{R}^m$, $\hat{h}^k \in \mathbb{R}^m$ *and* $\varepsilon_k^y \geq 0$ *such that*

$$\begin{cases} \hat{h}^k \in (H^{\varepsilon_k^y} + G_1^k)(\hat{y}^k), \\ e^k = c_k Q_k \hat{h}^k + \hat{y}^k - \left(y^k - c_k Q_k \left[G(x^k, y^k) - G_1^k(y^k)\right]\right). \end{cases} \tag{12}$$

Inexact Proximal Step: *Choose the error tolerance parameter* $\sigma_k \in (0, 1)$ *and a symmetric* $n \times n$ *positive definite matrix* P_k. *Compute* $\hat{x}^k \in \mathbb{R}^n$, $\hat{u}^k \in \mathbb{R}^n$ *and* $\varepsilon_k^x \geq 0$ *such that*

$$\begin{cases} \hat{u}^k \in F^{\varepsilon_k^x}(\hat{x}^k, \hat{y}^k) \\ r^k = c_k P_k \hat{u}^k + \hat{x}^k - x^k, \end{cases} \tag{13}$$

where the enlargement $F^{\varepsilon_k^x}$ *is in* x *for* $\hat{y}^k$ *fixed, and*

$$\begin{aligned} &\|r^k\|^2_{P_k^{-1}} + \|s^k\|^2_{Q_k^{-1}} + 2c_k(\varepsilon_k^x + \varepsilon_k^y) \\ &\leq \sigma_k^2 \left(\|c_k P_k \hat{u}^k\|^2_{P_k^{-1}} + \|c_k Q_k \hat{w}^k\|^2_{Q_k^{-1}} + \|\hat{x}^k - x^k\|^2_{P_k^{-1}} + \|\hat{y}^k - y^k\|^2_{Q_k^{-1}} \right), \end{aligned} \tag{14}$$

with

$$\hat{w}^k = G(\hat{x}^k, \hat{y}^k) + \hat{h}^k - G_1^k(\hat{y}^k), \qquad s^k = c_k Q_k \hat{w}^k + \hat{y}^k - y^k.$$

(If the proximal subproblem (13) *is solved to "maximal possible precision" but* (14) *is not satisfied, decrease* c_k *or choose a new matrix* Q_k *to decrease* $\|Q_k\|$*, and/or compute* $\hat{y}^k$ *in the* `Inexact Forward-Backward Splitting Step` *with more accuracy and repeat the* `Inexact Proximal Step` *with this new* $\hat{y}^k$*.)*

`Iterates Update:` *Stop if* $\hat{x}^k = x^k$ *and* $\hat{y}^k = y^k$*. Otherwise, choose* $\tau_k \in (1-\theta, 1+\theta)$ *and define*

$$\begin{aligned} x^{k+1} &= x^k - \tau_k a_k P_k \hat{u}^k, \\ y^{k+1} &= y^k - \tau_k a_k Q_k \hat{w}^k, \end{aligned} \tag{15}$$

where

$$a_k = \frac{\langle (\hat{u}^k, \hat{w}^k), (x^k - \hat{x}^k, y^k - \hat{y}^k) \rangle - (\varepsilon_k^x + \varepsilon_k^y)}{\|P_k \hat{u}^k\|^2_{P_k^{-1}} + \|Q_k \hat{w}^k\|^2_{Q_k^{-1}}}.$$

Set $k := k+1$ *and go to* `Inexact Forward-Backward Splitting Step`*.*

We recall that if the stronger than (14) approximation is used, i.e.,

$$\|r^k\|^2_{P_k^{-1}} + \|s^k\|^2_{Q_k^{-1}} + 2c_k(\varepsilon_k^x + \varepsilon_k^y) \le \sigma_k^2 \left(\|\hat{x}^k - x^k\|^2_{P_k^{-1}} + \|\hat{y}^k - y^k\|^2_{Q_k^{-1}} \right), \tag{16}$$

then in (15) we can use the stepsize $\tau_k a_k = c_k$.

Apart from our ability to satisfy condition (14), which requires a proof, the other parts of the method are easily seen to be well-defined. Indeed, since G_1^k is a monotone continuous function, it follows that $H + G_1^k$ is maximal monotone. Thus $\hat{y}^k$ in the forward-backward splitting step is well-defined and $\hat{y}^k \in \text{dom}H$. As already noted above, for any $\hat{y}^k \in \text{dom}H$, the mapping $x \to F(x, \hat{y}^k)$ is maximal monotone under the stated assumptions. Thus the proximal point step is also well-defined. Furthermore, the stepsize choice of a_k is well-defined whenever $\hat{u}^k \neq 0$ or $\hat{w}^k \neq 0$. Now, if it were the case that $\hat{u}^k = 0$ and $\hat{w}^k = 0$, then it would follow that $r^k = \hat{x}^k - x^k$ and $s^k = \hat{y}^k - y^k$. But (14) then implies that $\hat{x}^k = x^k$ and $\hat{y}^k = y^k$ (because $\sigma_k \in (0,1)$), so that the stopping rule would have been activated (as will be shown in Theorem 2.2, in this case (x^k, y^k) is a solution of the problem).

Before proceeding to the convergence analysis we make one final assumption:

A5 It holds that

$$\left. \begin{array}{l} u \in F^{\varepsilon^x}(x, y) \\ w \in G(x, y) + H^{\varepsilon^y}(y) \end{array} \right\} \Rightarrow (u, w) = v \in T^{\varepsilon^x + \varepsilon^y}(z), \quad z = (x, y),$$

where $F^{\varepsilon^x}(x, y)$ is the ε-enlargement of $F(\cdot, y)$ at x with the y-part fixed.

This assumption is redundant if in VMHPDM we set $\varepsilon_k^x = \varepsilon_k^y = 0$ for all k. Furthermore, in [32] it is shown that A5 always holds for (set-valued) monotone variational inequalities with linear constraints. We shall refer to Assumption A5 with $\varepsilon^y = 0$ as Assumption **A5x**, and to Assumption A5 with $\varepsilon^x = 0$ as Assumption **A5y** .

Theorem 2.2. *Suppose that* $T^{-1}(0) \neq \emptyset$*, where* T *is given by* (11)*, and that Assumptions* A1–A4 *hold. Suppose further that either in Algorithm* 2.1 *we set* $\varepsilon_k^x = \varepsilon_k^y = 0$ *for all* k*, or we set* $\varepsilon_k^y = 0$ *and Assumption* A5x *holds, or we set* $\varepsilon_k^x = 0$ *and Assumption*

A5y *holds, or that Assumptions* A5 *holds. Let G be Lipschitz-continuous and G_1^k be Lipschitz-continuous uniformly in k.*

Suppose that

$$0 < \sigma_l \le \liminf_{k\to\infty} \sigma_k \le \limsup_{k\to\infty} \sigma_k \le \sigma_u < 1,$$

and the matrix sequences $\{P_k\}, \{Q_k\}$ satisfy the conditions

$$0 < \lambda_l \le \liminf_{k\to\infty} \lambda_{min}(P_k) \le \limsup_{k\to\infty} \lambda_{max}(P_k) \le \lambda_u,$$

$$0 < \lambda_l \le \liminf_{k\to\infty} \lambda_{min}(Q_k) \le \limsup_{k\to\infty} \lambda_{max}(Q_k) \le \lambda_u,$$

$$\frac{1}{1+\eta_k} Q_k \preceq Q_{k+1}, \qquad \frac{1}{1+\eta_k} P_k \preceq P_{k+1}, \qquad \eta_k > 0 \; \forall k, \; \sum_{k=0}^{\infty} \eta_k < +\infty. \tag{17}$$

Then there exists $c_u > 0$ such that if

$$0 < c_l \le \liminf_{k\to\infty} c_k \le \limsup_{k\to\infty} c_k \le c_u$$

and the forward-backward step is computed with sufficiently accuracy, then the sequence $\{(x^k, y^k)\}$ generated by Algorithm 2.1 *is well-defined and converges to an element of $T^{-1}(0)$.*

If, in addition, T^{-1} is Lipschitzian at zero, i.e., there exist $L_1 > 0$ and $L_2 > 0$ such that

$$T^{-1}(v) \subset T^{-1}(0) + L_1 \|v\| \mathcal{B} \quad \forall v \in L_2 \mathcal{B}, \tag{18}$$

where $\mathcal{B}$ is the unit ball in $\mathbb{R}^n$, then the algorithm parameters can be chosen to obtain the linear rate of convergence.

Finally, for any choice of the parameters $0 < \sigma_l \le \sigma_u < 1$, and $0 < c_l \le c_u$, we can choose the parameters $0 < \lambda_l \le \lambda_u$ such that if the matrices $\{P_k\}, \{Q_k\}$ satisfy, in addition to (17)*, the conditions*

$$Q_{k+1} \preceq (1+\eta_k) Q_k, \qquad P_{k+1} \preceq (1+\eta_k) P_k, \tag{19}$$

then there exists an index k_0 such that the convergence rate is linear in the norm induced by $M_{k_0}^{-1}$, where

$$M_k = \begin{pmatrix} P_k & 0 \\ 0 & Q_k \end{pmatrix}.$$

Proof. From (12) we obtain

$$\hat{y}^k - y^k = c_k Q_k \left[G_1^k(y^k) - G(x^k, y^k) - \hat{h}^k \right] + e^k,$$

so that

$$\begin{aligned} s^k &= c_k Q_k \hat{w}^k + \hat{y}^k - y^k \\ &= c_k Q_k \left[G(\hat{x}^k, \hat{y}^k) + \hat{h}^k - G_1^k(\hat{y}^k) \right] + \hat{y}^k - y^k \\ &= c_k Q_k \left[G_1^k(y^k) - G_1^k(\hat{y}^k) + G(\hat{x}^k, \hat{y}^k) - G(x^k, y^k) \right] + e^k. \end{aligned}$$

Suppose that the forward-backward splitting step is computed with sufficient accuracy, so that the following two conditions are satisfied:

$$\|e^k\| \le c_k L \|Q_k\| \|\hat{y}^k - y^k\|, \qquad \varepsilon_k^y \le \frac{c_k L^2 \|Q_k\|^2}{2\lambda_{min}(Q_k)} \|\hat{y}^k - y^k\|^2,$$

where $L > 0$ is the modulus of Lipschitz-continuity of G and G_1^k. Then

$$\begin{aligned} \|s^k\| &\le c_k L \|Q_k\| (\|\hat{y}^k - y^k\| + \|(\hat{x}^k, \hat{y}^k) - (x^k, y^k)\|) + \|e^k\| \\ &\le c_k L \|Q_k\| (2\|\hat{y}^k - y^k\| + \|(\hat{x}^k, \hat{y}^k) - (x^k, y^k)\|) \\ &\le 3 c_k L \|Q_k\| \|(\hat{x}^k, \hat{y}^k) - (x^k, y^k)\|. \end{aligned}$$

It follows that

$$\begin{aligned} & \|s^k\|^2_{Q_k^{-1}} + 2c_k \varepsilon_k^y \\ \le\ & \frac{1}{\lambda_{min}(Q_k)} \|s^k\|^2 + \frac{c_k^2 L^2 \|Q_k\|^2}{\lambda_{min}(Q_k)} \|\hat{y}^k - y^k\|^2 \\ \le\ & \frac{10 c_k^2 L^2 \lambda_{max}(Q_k)^2}{\lambda_{min}(Q_k)} (\|\hat{x}^k - x^k\|^2 + \|\hat{y}^k - y^k\|^2) \\ \le\ & \frac{10 c_k^2 L^2 \lambda_{max}(Q_k)^2}{\lambda_{min}(Q_k)} \left(\lambda_{max}(P_k) \|\hat{x}^k - x^k\|^2_{P_k^{-1}} + \lambda_{max}(Q_k) \|\hat{y}^k - y^k\|^2_{Q_k^{-1}} \right) \\ \le\ & \frac{10 c_u L^2 \lambda_u^3}{\lambda_l} \left(\|\hat{x}^k - x^k\|^2_{P_k^{-1}} + \|\hat{y}^k - y^k\|^2_{Q_k^{-1}} \right). \end{aligned}$$

Hence, if we choose $c_u > 0$ such that

$$\frac{10 c_u L^2 \lambda_u^3}{\lambda_l} \le \sigma_l^2,$$

then condition (14) can always be satisfied (it is enough to note that the exact solution of the proximal system (13), corresponding to $r^k = 0$ and $\varepsilon_k^x = 0$, satisfies (14)). This concludes the proof of the claim that Algorithm 2.1 is well-defined.

Since $\hat{u}^k \in F^{\varepsilon_k^x}(\hat{x}^k, \hat{y}^k)$ and $\hat{w}^k = G(\hat{x}^k, \hat{y}^k) + \hat{h}^k - G_1^k(\hat{y}^k)$, with $\hat{h}^k - G_1^k(\hat{y}^k) \in H^{\varepsilon_k^y}(\hat{y}^k)$, Assumption A5 guarantees that $(\hat{u}^k, \hat{w}^k) \in T^{\varepsilon_k}(\hat{x}^k, \hat{y}^k)$ for $\varepsilon_k = \varepsilon_k^x + \varepsilon_k^y$. The same inclusion is satisfied also if $\varepsilon_k^y = 0$ and Assumption A5x holds, or if $\varepsilon_k^x = 0$ and Assumption A5y holds, or if $\varepsilon_k^x = \varepsilon_k^y = 0$.

It now follows that with the identifications

$$z^k = (x^k, y^k), \quad \hat{z}^k = (\hat{x}^k, \hat{y}^k), \quad \hat{v}^k = (\hat{u}^k, \hat{w}^k), \quad \delta^k = (r^k, s^k), \quad M_k = \begin{pmatrix} P_k & 0 \\ 0 & Q_k \end{pmatrix},$$

Algorithm 2.1 falls within the VMHIPPM framework. The announced convergence results then essentially follow adapting [23, Theorems 4.2–4.4]. For the sake of completeness, and to take care of some necessary details, we include a streamlined proof.

From (3), by re-arranging terms, it is easy to see that (4) is equivalent to

$$\langle \hat{v}^k, z^k - \hat{z}^k \rangle - \varepsilon_k \ge \frac{1 - \sigma_k^2}{2c_k} \left(\|c_k M_k \hat{v}^k\|^2_{M_k^{-1}} + \|\hat{z}^k - z^k\|^2_{M_k^{-1}} \right). \tag{20}$$

By the definition of a_k, after some algebraic transformations, we obtain that

$$a_k \geq \frac{1-\sigma_k^2}{2c_k}\left(\frac{\|c_k M_k \hat{v}^k\|^2_{M_k^{-1}} + \|\hat{z}^k - z^k\|^2_{M_k^{-1}}}{\|M_k \hat{v}^k\|^2_{M_k^{-1}}}\right) \geq \frac{(1-\sigma_k^2)c_k}{1+\sqrt{1-(1-\sigma_k^2)^2}}. \tag{21}$$

And, since $\left(\|c_k M_k \hat{v}^k\|_{M_k^{-1}} - \|\hat{z}^k - z^k\|_{M_k^{-1}}\right)^2 \geq 0$, it holds that

$$a_k \|M_k \hat{v}^k\|_{M_k^{-1}} \geq (1-\sigma_k^2)\|\hat{z}^k - z^k\|_{M_k^{-1}}. \tag{22}$$

It is only necessary to consider the case when the right-hand side in (20) is positive (if it is zero, (3) and (4) imply that $\hat{z}^k \in T^{-1}(0)$ and the method stops). If the right-hand side in (20) is positive then z^k is not contained in the closed halfspace

$$\mathcal{H}_k = \{z \in \mathbb{R}^n \mid \langle \hat{v}^k, z - \hat{z}^k\rangle - \varepsilon_k \leq 0\}.$$

Since $\hat{v}^k \in T^{\varepsilon_k}(\hat{z}^k)$, it holds that for any $z^* \in T^{-1}(0)$ we have $\langle \hat{v}^k - 0, \hat{z}^k - z^*\rangle \geq -\varepsilon_k$. In particular, $z^* \in \mathcal{H}_k$. By the properties of the skewed projection onto $\mathcal{H}_k$, we have

$$\bar{z} := P_{\mathcal{H}_k, M_k^{-1}}(z^k) = z^k - \frac{\langle \hat{v}^k, z^k - \hat{z}^k\rangle - \varepsilon_k}{\|M_k \hat{v}^k\|^2_{M_k^{-1}}} M_k \hat{v}^k = z^k - a_k M_k \hat{v}^k,$$

$$\bar{z} - z^{k+1} = (\tau_k - 1) a_k M_k \hat{v}^k, \quad \text{and} \quad \langle z^* - \bar{z}, \hat{v}^k\rangle \leq 0.$$

Using these relations (after adding and subtracting adequate terms), we obtain

$$\begin{aligned} \|z^* - z^{k+1}\|^2_{M_k^{-1}} &\leq \|z^* - z^k\|^2_{M_k^{-1}} - (1-(1-\tau_k)^2) a_k^2 \|M_k \hat{v}^k\|^2_{M_k^{-1}} \\ &\leq \|z^* - z^k\|^2_{M_k^{-1}} - (1-\theta^2)\|a_k M_k \hat{v}^k\|^2_{M_k^{-1}}. \end{aligned} \tag{23}$$

Condition (17) implies that $M_{k+1}^{-1} \preceq (1+\eta_k) M_k^{-1}$ forall k, and that $\prod_{k=0}^{\infty}(1+\eta_k) = p < \infty$. Hence, from (10) and (23),

$$\begin{aligned} \lambda_u^{-1}\|z^* - z^{k+1}\|^2 &\leq \|z^* - z^{k+1}\|^2_{M_{k+1}^{-1}} \\ &\leq (1+\eta_k)\|z^* - z^k\|^2_{M_k^{-1}} - (1-\theta^2)\|a_k M_k \hat{v}^k\|^2_{M_k^{-1}}. \end{aligned}$$

Applying this inequality consecutively, we obtain that

$$\lambda_u^{-1}\|z^* - z^{k+1}\|^2 \leq \prod_{i=0}^{k}(1+\eta_i)\|z^* - z^0\|^2_{M_0^{-1}} - (1-\theta^2)\sum_{i=0}^{k}\|a_i M_i \hat{v}^i\|^2_{M_i^{-1}}, \tag{24}$$

and, for any k,

$$\|z^* - z^k\|^2 \leq \lambda_u \prod_{i=0}^{k-1}(1+\eta_i)\|z^* - z^0\|^2_{M_0^{-1}} \leq \frac{p\lambda_u}{\lambda_l}\|z^* - z^0\|^2, \tag{25}$$

which shows that the sequence $\{z^k\}$ is bounded. Therefore, it has some accumulation point, say $\tilde{z} \in \mathbb{R}^n$. Passing onto the limit when $k \to \infty$ in these inequalities, we obtain that

$$\sum_{k=0}^{\infty}\|a_k M_k \hat{v}^k\|^2 \leq \lambda_u \sum_{k=0}^{\infty}\|a_k M_k \hat{v}^k\|^2_{M_k^{-1}} \leq \frac{p\lambda_u}{1-\theta^2}\|z^* - z^0\|^2_{M_0^{-1}} < \infty,$$

and, as a consequence, we have

$$\lim_{k\to\infty} \|a_k M_k \hat{v}^k\| = 0, \qquad \lim_{k\to\infty} \|M_k \hat{v}^k\| = 0 \quad \text{and} \quad \lim_{k\to\infty} \|\hat{z}^k - z^k\| = 0.$$

Since the matrices M_k are uniformly positive definite, we also have that $\lim_{k\to\infty} \hat{v}^k = 0$. And, since $\varepsilon_k \leq \langle \hat{v}^k, z^k - \hat{z}^k\rangle$, it follows that $\lim_{k\to\infty} \varepsilon_k = 0$.

Let $\{z^{k_j}\}$ be any subsequence converging to $\tilde{z}$. It then holds that $\hat{z}^{k_j} \to \tilde{z}$. For any $z \in \mathbb{R}^n$ and any $u \in T(z)$, $\langle u - v^{k_j}, z - \hat{z}^{k_j}\rangle \geq -\varepsilon_{k_j}$. Hence,

$$\langle u - 0, z - \hat{z}^{k_j}\rangle \geq \langle v^{k_j}, z - \hat{z}^{k_j}\rangle - \varepsilon_{k_j},$$

and passing onto the limit when $j \to \infty$ we obtain

$$\langle u - 0, z - \tilde{z}\rangle \geq 0.$$

As $z \in \mathbb{R}^n$ and $u \in T(z)$ were arbitrarily chosen, and T is maximal monotone, the above relation shows that $0 \in T(\tilde{z})$, i.e., $\tilde{z}$ is a solution. The proof of uniqueness of the accumulation point is standard.

Assume now the Lipschitzian property (18) of T^{-1}. Let $\xi^k, \psi^k \in T(\xi^k)$ be the exact solution of the proximal system $\psi \in T(\xi)$, $0 = b_k M_k \psi + \xi - z^k$, where $b_k = \tau_k a_k$. Since $\hat{v}^k \in T^{\varepsilon_k}(\hat{z}^k)$, by [23, Lemma 4.3], by the definitions of ψ^k, $\hat{v}^k$ and a_k, and (22), it follows that

$$\begin{aligned}
& \|\xi^k - \hat{z}^k\|^2_{M_k^{-1}} + \|\xi^k - z^{k+1}\|^2_{M_k^{-1}} \\
= \;& \|\xi^k - \hat{z}^k\|^2_{M_k^{-1}} + b_k^2 \|M_k \hat{v}^k - M_k \psi^k\|^2_{M_k^{-1}} \\
\leq \;& \|\hat{z}^k - z^k\|^2_{M_k^{-1}} + (\tau_k^2 - 2\tau_k)\|a_k M_k \hat{v}^k\|^2_{M_k^{-1}} \\
\leq \;& \left(\tau_k^2 - 2\tau_k + \frac{1}{(1-\sigma_k^2)^2}\right) \|a_k M_k \hat{v}^k\|^2_{M_k^{-1}}.
\end{aligned} \tag{26}$$

Since $\hat{v}^k \to 0$, the first inequality in (26) implies that $\psi^k \to 0$. Hence, there exists $k_1 \in \mathbb{N}$ such that $\|\psi^k\| < L_2$ for all $k > k_1$. By (18), we then have that

$$\operatorname{dist}(\xi^k, T^{-1}(0)) \leq L_1 \|\psi^k\| \quad \forall k > k_1.$$

Therefore, for $k > k_1$,

$$\operatorname{dist}(\xi^k, T^{-1}(0))^2_{M_k^{-1}} \leq \frac{L_1^2}{\lambda_l^2}\|\psi^k\|^2_{M_k} = \frac{L_1^2}{\lambda_l^2 b_k^2}\|z^k - \xi^k\|^2_{M_k^{-1}}. \tag{27}$$

Let $\bar{\xi}^k := P_{T^{-1}(0), M_k^{-1}}(\xi^k)$. Then, for $k > k_1$, by combining the Cauchy-Schwarz inequality with (21), (22) and (23), we have that

$$\begin{aligned}
& \operatorname{dist}(z^{k+1}, T^{-1}(0))_{M_k^{-1}} \\
\leq \;& \|z^{k+1} - \bar{\xi}^k\|_{M_k^{-1}} \\
\leq \;& \|z^{k+1} - \xi^k\|_{M_k^{-1}} + \operatorname{dist}(x^k, T^{-1}(0))_{M_k^{-1}} \\
\leq \;& \|z^{k+1} - \xi^k\|_{M_k^{-1}} + \frac{L_1}{\lambda_l b_k}\left(\|\xi^k - \hat{z}^k\|_{M_k^{-1}} + \|\hat{z}^k - z^k\|_{M_k^{-1}}\right), \\
\leq \;& \mu_k \|a_k M_k \hat{v}^k\|_{M_k^{-1}},
\end{aligned} \tag{28}$$

where

$$\mu_k := \sqrt{\alpha_k^2 + 1}\sqrt{\beta_k^2 - 1} + \alpha_k\beta_k, \tag{29}$$

with

$$\alpha_k := \frac{L_1\left(1+\sqrt{1-(1-\sigma_k^2)^2}\right)}{\lambda_l c_k(1-\sigma_k^2)(1-\theta)} \leq \frac{L_1\left(1+\sqrt{1-(1-\sigma_u^2)^2}\right)}{\lambda_l c_l(1-\sigma_u^2)(1-\theta)} =: \alpha. \tag{30}$$

$$\text{and} \quad \beta_k := \frac{1}{1-\sigma_k^2} \leq \frac{1}{1-\sigma_u^2} =: \beta. \tag{31}$$

Let $\bar{z}^k := P_{T^{-1}(0),M_k^{-1}}(z^k)$. From (23) and (28), we obtain

$$\begin{aligned} \operatorname{dist}(z^k, T^{-1}(0))^2_{M_k^{-1}} &\geq \operatorname{dist}(z^{k+1}, T^{-1}(0))^2_{M_k^{-1}} + (1-\theta^2)\|a_k M_k \hat{v}^k\|^2_{M_k^{-1}} \\ &\geq \left(1 + \frac{1-\theta^2}{\mu_k^2}\right) \operatorname{dist}(z^{k+1}, T^{-1}(0))^2_{M_k^{-1}}. \end{aligned} \tag{32}$$

Therefore,

$$\operatorname{dist}(z^{k+1}, T^{-1}(0)) \leq \frac{\mu_k\sqrt{\lambda_u}}{\sqrt{\lambda_l(\mu_k^2 + 1 - \theta^2)}} \operatorname{dist}(z^k, T^{-1}(0)). \tag{33}$$

Let $\gamma > 1$. By the definitions (30) and (31), taking c_k sufficiently large we can make α_k arbitrarily small, and by taking σ_k sufficiently small we can make β_k arbitrarily close to one, so we can make μ_k sufficiently small to satisfy

$$\frac{\mu_k}{\sqrt{\mu_k^2 + 1 - \theta^2}} < \frac{1}{\gamma}.$$

Now note that by choosing $0 < \sigma_l < \sigma_u$, $c_u > c_l > 0$, $\lambda_u \geq \lambda_l > 0$ such that

$$\sqrt{\lambda_u/\lambda_l} < \gamma, \qquad \lambda_u \leq \sigma_l/(L\gamma\sqrt{10c_u}),$$

we satisfy the condition in the part of the proof that shows that the method is well-defined. In addition, (33) now establishes that $\{\operatorname{dist}(z^k, T^{-1}(0))\}$ converges linearly to zero. For Fejér-monotone sequences, this is equivalent to the linear convergence rate of $\{z^k\}$ to its limit (see, e.g., [1]).

Assume now that the additional conditions (19) hold. Then,

$$\begin{aligned} \frac{1}{(1+\eta_k)} \operatorname{dist}(z, T^{-1}(0))^2_{M_k^{-1}} &= \inf_{y \in T^{-1}(0)} \frac{1}{(1+\eta_k)} \|z - y\|^2_{M_k^{-1}} & (34) \\ &\leq \inf_{y \in T^{-1}(0)} \|z - y\|^2_{M_{k+1}^{-1}} \\ &\leq \inf_{y \in T^{-1}(0)} (1+\eta_k)\|z - y\|^2_{M_k^{-1}} \\ &= (1+\eta_k) \operatorname{dist}(z, T^{-1}(0))^2_{M_k^{-1}}. & (35) \end{aligned}$$

Define $\mu = \sqrt{\alpha^2 + 1}\sqrt{\beta^2 - 1} + \alpha\beta$. Note that $\mu > \mu_k$ for all k.

Since $\prod_{i=0}^{\infty}(1+\eta_i) < \infty$, there exists $k_2 \in \mathbb{N}$ such that

$$\prod_{i=k_2}^{\infty}(1+\eta_i) < \frac{\sqrt{\mu^2+1-\theta^2}}{\mu}.$$

From (32), applying (34) consecutively, for any $k \geq k_0 := \max\{k_1, k_2\}$, we have that

$$\operatorname{dist}(z^{k+1}, T^{-1}(0))_{M_{k_0}^{-1}} \leq \nu \operatorname{dist}(z^k, T^{-1}(0))_{M_{k_0}^{-1}},$$

where

$$\nu := \frac{\mu}{\sqrt{\mu^2+1-\theta^2}} \prod_{i=k_0}^{\infty}(1+\eta_i) < 1,$$

as claimed. □

3. Applications

We next show that in addition to decomposition methods of [10] and [42], already covered by the scheme of [32], the proposed variable metric framework also includes splitting of composite mappings [25] and proximal alternating directions methods [13, 17]. This provides a unified view of all those techniques, some of which are seemingly unrelated, as well as adds some new convergence rate results.

3.1. Splitting Method for Composite Mappings

Consider the following variational inclusion in the composite form: Find $x \in \mathbb{R}^n$ such that

$$0 \in A^\top \Gamma A(x), \tag{36}$$

where $A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear operator (a matrix of appropriate dimensions), and $\Gamma : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ is a maximal monotone (set-valued) operator. Problems of this form arise frequently in applications [31, 27, 24]. For example, for a composite function $f \circ A$, where f is convex, under appropriate regularity conditions it holds that $\partial(f \circ A) = A^\top \partial f A$. Also, a sum of operators $T_1 + T_2$ is a special case corresponding to taking $Ax = (x, x)$ and $\Gamma(x_1, x_2) = T_1(x_1) \times T_2(x_2)$, so that $A^\top(y_1, y_2) = y_1 + y_2$ and, hence, $A^\top \Gamma A(x) = T_1(x) + T_2(x)$. (It should be noted that the converse is also true, i.e., a composite map can be reformulated as a sum, e.g., [16].) However, unlike for sums of operators for which a wealth of splitting methods have been proposed (e.g., [20, 41, 12, 11, 14, 43, 5]), decomposition of inclusions in the composite form appears to be much less developed. The first proposal in this direction seems to be the method of [25], which is derived as an application of the Method of Partial Inverses [40], with the decomposition of the space $\mathbb{R}^m = \operatorname{rge} A \oplus \ker A^\top$.

Let $A^\dagger$ be the pseudo-inverse of A, i.e., for each $y \in \mathbb{R}^m$ it gives the minimal-norm solution of the least-squares problem $\min_x \|Ax - y\|^2$. We next state the method of [25] (we note that in [25] the more general setting of Hilbert spaces is considered).

Algorithm 3.1.

1. Choose some $y_1^0 \in \operatorname{rge} A$ and $y_2^0 \in \ker A^\top$. Set $k := 1$.

2. Compute $\hat{y}_1^k$ such that

$$0 \in \Gamma(\hat{y}_1^k) + \hat{y}_1^k - (y_1^k + y_2^k).$$

3. Set

$$\begin{aligned} x^{k+1} &= A^{\dagger}(\hat{y}_1^k), \\ y_1^{k+1} &= A(x^{k+1}), \\ y_2^{k+1} &= y_2^k + y_1^{k+1} - \hat{y}_1^k. \end{aligned}$$

Set $k := k+1$, and go to Step 2.

The above is indeed an attractive splitting method for solving the composite inclusion (36), as it achieves a full decomposition between A and Γ: one computes proximal steps for Γ and solves least-squares problems for A. We refer the reader to [25] for a detailed discussion and some applications.

We next show that the splitting Algorithm 3.1 is a special case in the VMHPDM framework of Section 2. Among other things, this gives rate of convergence results for Algorithm 3.1 that were not available previously.

We first define the appropriate mappings H, G and F, that put the inclusion with composite structure (36) in the form of finding a zero of T given by (11).

Fix some $\nu > 2$ and define

$$H : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^m, \quad H\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \Gamma(y_1) \\ 0 \end{pmatrix},$$

$$G : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^m, \quad G\begin{pmatrix} x \\ y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} -y_2 + \nu^{-1}(y_1 - Ax) \\ y_1 - Ax \end{pmatrix},$$

$$F : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^n, \quad F\begin{pmatrix} x \\ y_1 \\ y_2 \end{pmatrix} = A^{\top} y_2 + \nu^{-1} A^{\top}(Ax - y_1).$$

We claim that, with the definitions above, the problem

$$0 \in T(z) = T(x, y_1, y_2) = F(x, y_1, y_2) \times [G(x, y_1, y_2) + H(y_1, y_2)]$$

is equivalent to the composite inclusion (36). Indeed, for $z = (x, y_1, y_2) \in T^{-1}(0)$ it holds that

$$\begin{aligned} 0 = y_1 - Ax \quad &\Rightarrow \quad y_1 = Ax \in \mathrm{rge} A, \quad A^{\top}(Ax - y_1) = 0, \\ 0 = A^{\top} y_2 + \nu^{-1} A^{\top}(Ax - y_1) = A^{\top} y_2 \quad &\Rightarrow \quad y_2 \in \ker A^{\top}, \\ 0 \in \Gamma(y_1) - y_2 + \nu^{-1}(y_1 - Ax) = \Gamma(y_1) - y_2 \quad &\Rightarrow \quad y_2 \in \Gamma(y_1). \end{aligned}$$

Combining these relations, we obtain that

$$A^{\top} \Gamma A(x) = A^{\top} \Gamma(y_1) \ni A^{\top} y_2 = 0.$$

Conversely, if $x \in \mathbb{R}^n$ verifies $0 \in A^\top \Gamma A(x)$, there must exist $y_1 \in \mathrm{rge}A$ and $y_2 \in \ker A^\top$ such that $y_1 = Ax$ and $0 \in \Gamma(y_1) - y_2$. It is easily seen that $0 \in T(x, y_1, y_2)$. We conclude that the two problems are equivalent.

Furthermore, the Assumptions A1, A2 and A4 of Section 2 hold trivially. We next show that Assumption A3 holds as well. As $F \times G$ is single-valued and continuous, it is enough to prove its monotonicity. To this end,

$$
\begin{aligned}
&\left\langle \begin{pmatrix} x - x' \\ y_1 - y_1' \\ y_2 - y_2' \end{pmatrix}, \begin{pmatrix} A^\top(y_2 - y_2') + \nu^{-1}A^\top(A(x - x') - (y_1 - y_1')) \\ y_2' - y_2 + \nu^{-1}(y_1 - y_1' - A(x - x')) \\ y_1 - y_1' - A(x - x') \end{pmatrix} \right\rangle \\
&= \nu^{-1}(\|A(x - x')\|^2 - 2\langle y_1 - y_1', A(x - x')\rangle + \|y_1 - y_1'\|^2) \\
&= \nu^{-1}\|A(x - x') - (y_1 - y_1')\|^2 \geq 0.
\end{aligned}
$$

Thus, the chosen operator T satisfies all the assumptions of Section 2.

Next note that $A^\top A$ is a symmetric positive semidefinite matrix, which is positive definite if $\ker A = \{0\}$. If $\ker A \neq \{0\}$, we can make the decomposition $\mathbb{R}^n = \ker A \times (\ker A)^\perp$ and write

$$
A^\top A = \begin{pmatrix} R & 0 \\ 0 & 0 \end{pmatrix}, \tag{37}
$$

where R is a symmetric positive definite matrix that acts on $(\ker A)^\perp$.

In the VMHPDM Algorithm 2.1, we shall now choose the following parameters:

$$
c_k = 1, \qquad G_1^k \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} -y_2 \\ y_1 + \frac{1}{\nu - 1} y_2 \end{pmatrix}, \tag{38}
$$

$$
Q_k = \begin{pmatrix} \nu & 0 \\ 0 & 1 \end{pmatrix}, \qquad P_k = \frac{\nu}{\nu - 1} P, \tag{39}
$$

with $P = (A^\top A)^{-1}$ if $\ker A = \{0\}$ and $P = \begin{pmatrix} R^{-1} & 0 \\ 0 & I \end{pmatrix}$ otherwise, where R is the symmetric positive definite matrix defined in (37). Note that G_1^k is Lipschitz-continuous and monotone (since $\nu > 2$).

Suppose that $\{y_1^k\}_{k=0}^\infty$, $\{y_2^k\}_{k=0}^\infty$ and $\{x^k\}_{k=1}^\infty$ are the iterates generated by Algorithm 3.1. Observe that $y_2^0 \in \ker A^\top$ and formally define $x^0 = A^\dagger y_1^0 \in (\ker A)^\perp$, so that $y_1^0 = Ax^0$ (this is done because VMHPDM needs also the x-part of the starting point, while the splitting Algorithm 3.1 employs only the y-part to start). We next show that VMHPDM Algorithm 2.1 applied to the operator T, with (x^0, y_1^0, y_2^0) as the starting point and with appropriate choice of the algorithm parameters, generates the same iterates as the splitting Algorithm 3.1.

Suppose that Algorithms 2.1 and 3.1 have the same iterates until the one indexed by some $k \geq 0$. We shall show that the $(k + 1)$ iterate is then also the same.

`Forward-Backward Splitting Step` (i.e., (12) with $\varepsilon_k^y = 0$, $e^k = 0$ and other parameters defined in (38)–(39)) computes $\hat{y}_1^k \in \mathbb{R}^m, \hat{y}_2^k \in \mathbb{R}^m$ such that

$$\begin{cases} \begin{pmatrix} \hat{h}_1^k \\ \hat{h}_2^k \end{pmatrix} \in (H + G_1^k)(\hat{y}^k) = \begin{pmatrix} \Gamma(\hat{y}_1^k) \\ 0 \end{pmatrix} + \begin{pmatrix} -\hat{y}_2^k \\ \hat{y}_1^k + \frac{1}{\nu-1}\hat{y}_2^k \end{pmatrix}, \\ 0 \;=\; Q_k\hat{h}^k + \hat{y}^k - \left(y^k - Q_k\left[G(x^k,y^k) - G_1^k(y^k)\right]\right) \\ \quad = \begin{pmatrix} \nu\hat{h}_1^k + \hat{y}_1^k - \left(y_1^k - \nu\left[-y_2^k + \nu^{-1}(y_1^k - Ax^k) + y_2^k\right]\right) \\ \hat{h}_2^k + \hat{y}_2^k - \left[y_2^k - (y_1^k - Ax^k - y_1^k - \frac{1}{\nu-1}y_2^k)\right] \end{pmatrix}. \end{cases}$$

Since $y_1^k - Ax^k = 0$ (by the construction of Algorithm 3.1, given that the k-th iterates coincide), from the latter relations we obtain that

$$\begin{cases} \begin{pmatrix} \hat{h}_1^k \\ \hat{h}_2^k \end{pmatrix} \in \begin{pmatrix} \Gamma(\hat{y}_1^k) - \hat{y}_2^k \\ \hat{y}_1^k + \frac{1}{\nu-1}\hat{y}_2^k \end{pmatrix}, \\ 0 = \nu\hat{h}_1^k + \hat{y}_1^k - y_1^k, \\ 0 = \hat{h}_2^k - y_1^k + \hat{y}_2^k - \frac{\nu}{\nu-1}y_2^k. \end{cases}$$

Hence,

$$\hat{h}^k = \begin{pmatrix} \nu^{-1}(y_1^k - \hat{y}_1^k) \\ \hat{y}_1^k + \frac{1}{\nu-1}\hat{y}_2^k \end{pmatrix} \in \begin{pmatrix} \Gamma(\hat{y}_1^k) - \hat{y}_2^k \\ y_1^k - \hat{y}_2^k + \frac{\nu}{\nu-1}y_2^k \end{pmatrix}.$$

In particular, we have that

$$\hat{y}_2^k = y_2^k + \frac{\nu-1}{\nu}(y_1^k - \hat{y}_1^k). \tag{40}$$

`Inexact Proximal Step` (i.e., (13) with $\varepsilon_k^x = 0$ and other parameters defined in (38)–(39)) computes $\hat{x}^k \in \mathbb{R}^n$ and $\hat{u}^k \in \mathbb{R}^n$ such that

$$\begin{cases} \hat{u}^k \in F(\hat{x}^k, \hat{y}^k) = A^\top\hat{y}_2^k + \nu^{-1}A^\top(A\hat{x}^k - \hat{y}_1^k), \\ r^k = P_k\hat{u}^k + \hat{x}^k - x^k, \end{cases} \tag{41}$$

and

$$\|r^k\|^2_{P_k^{-1}} + \|\hat{s}^k\|^2_{Q_k^{-1}} \le \sigma_k^2\left(\|\hat{x}^k - x^k\|^2_{P_k^{-1}} + \|\hat{y}^k - y^k\|^2_{Q_k^{-1}}\right), \tag{42}$$

where $\hat{y}^k = (\hat{y}_1^k, \hat{y}_2^k)$, $y^k = (y_1^k, y_2^k)$ and

$$\hat{w}^k = G(\hat{x}^k, \hat{y}^k) + \hat{h}^k - G_1^k(\hat{y}^k) = \begin{pmatrix} \nu^{-1}(y_1^k - A\hat{x}^k) \\ \hat{y}_1^k - A\hat{x}^k \end{pmatrix},$$

$$\hat{s}^k = Q_k\hat{w}^k + \hat{y}^k - y^k = \begin{pmatrix} \hat{y}_1^k - A\hat{x}^k \\ \hat{y}_1^k - A\hat{x}^k + \hat{y}_2^k - y_2^k \end{pmatrix}.$$

In the above, instead of the approximation condition (14) in the VMHPDM framework we use the stronger version (16), corresponding to (42), so that we can define the new iterates using the stepsize $\tau_k a_k = c_k$ in (15), i.e., `Iterates Update:` Stop if $\hat{x}^k = k^k$ and $\hat{y}^k = y^k$, otherwise

$$x^{k+1} = x^k - P_k\hat{u}^k = \hat{x}^k - r^k,$$
$$y^{k+1} = y^k - Q_k\hat{w}^k = \begin{pmatrix} A\hat{x}^k \\ y_2^k - \hat{y}_1^k + A\hat{x}^k \end{pmatrix}.$$

We next show that for adequate choices of the approximation parameter $\sigma_k \in (0,1)$, the element $\hat{x}^k = A^\dagger \hat{y}_1^k$ satisfies (41) and (42).

By the very definition of the pseudo-inverse, it holds that

$$A^\top (A\hat{x}^k - \hat{y}_1^k) = 0. \tag{43}$$

Then, using also (40), we have that (41) takes the form

$$\begin{cases} \hat{u}^k = A^\top \hat{y}_2^k = A^\top (y_2^k + \frac{\nu-1}{\nu}(y_1^k - \hat{y}_1^k)), \\ r^k = P_k\hat{u}^k + \hat{x}^k - x^k = \frac{\nu}{\nu-1}P\hat{u}^k + \hat{x}^k - x^k. \end{cases} \tag{44}$$

By the construction of Algorithm 3.1, and since the k-th iterates coincide, we have that $y_2^k \in \ker A^\top$ and $y_1^k = Ax^k$. Hence, $A^\top y_2^k = 0$ and

$$A^\top (y_1^k - \hat{y}_1^k) = A^\top A(x^k - \hat{x}^k) + A^\top (A\hat{x}^k - \hat{y}_1^k) = P^{-1}(x^k - \hat{x}^k),$$

where we have taken into account (43) and the fact that $\hat{x}^k - x^k \in (\ker A)^\perp$. By using these relations in (44), we obtain

$$\begin{cases} \hat{u}^k = \frac{\nu-1}{\nu}P^{-1}(x^k - \hat{x}^k), \\ r^k = 0. \end{cases}$$

This implies that $x^{k+1} = \hat{x}^k - r^k = \hat{x}^k = A^\dagger \hat{y}_1^k$.

Furthermore, condition (42) then becomes

$$\begin{aligned} &\frac{1}{\nu}\|\hat{y}_1^k - A\hat{x}^k\|^2 + \|\hat{y}_1^k - A\hat{x}^k + \hat{y}_2^k - y_2^k\|^2 \\ \le\ & \sigma_k^2\left(\frac{\nu-1}{\nu}\|A(\hat{x}^k - x^k)\|^2 + \frac{1}{\nu}\|\hat{y}_1^k - y_1^k\|^2 + \|\hat{y}_2^k - y_2^k\|^2\right). \end{aligned} \tag{45}$$

Note that, since $y_1^k = Ax^k$ and $A^\top(\hat{y}_1^k - A\hat{x}^k) = 0$, we have

$$\begin{aligned} &\|\hat{y}_1^k - A\hat{x}^k\|^2 \\ =\ & \|\hat{y}_1^k - y_1^k + Ax^k - A\hat{x}^k\|^2 \\ =\ & \|\hat{y}_1^k - y_1^k\|^2 + \|Ax^k - A\hat{x}^k\|^2 + 2\langle \hat{y}_1^k - y_1^k, Ax^k - A\hat{x}^k\rangle \\ =\ & \|\hat{y}_1^k - y_1^k\|^2 + \|Ax^k - A\hat{x}^k\|^2 + 2\langle A^\top(\hat{y}_1^k - A\hat{x}^k + A\hat{x}^k - Ax^k), x^k - \hat{x}^k\rangle \\ =\ & \|\hat{y}_1^k - y_1^k\|^2 + \|Ax^k - A\hat{x}^k\|^2 + 2\langle A\hat{x}^k - Ax^k, Ax^k - A\hat{x}^k\rangle \\ =\ & \|\hat{y}_1^k - y_1^k\|^2 - \|Ax^k - A\hat{x}^k\|^2. \end{aligned}$$

Also, by using (40) and similar transformations as above, we obtain

$$\begin{aligned}
& \|\hat{y}_1^k - A\hat{x}^k + \hat{y}_2^k - y_2^k\|^2 \\
= & \left\| \hat{y}_1^k - y_1^k + Ax^k - A\hat{x}^k + \frac{\nu - 1}{\nu}(y_1^k - \hat{y}_1^k) \right\|^2 \\
= & \left\| \frac{1}{\nu}(\hat{y}_1^k - y_1^k) + A(x^k - \hat{x}^k) \right\|^2 \\
= & \frac{1}{\nu^2}\|\hat{y}_1^k - y_1^k\|^2 + \|A(x^k - \hat{x}^k)\|^2 + \frac{2}{\nu}\langle \hat{y}_1^k - y_1^k, A(x^k - \hat{x}^k)\rangle \\
= & \frac{1}{\nu^2}\|\hat{y}_1^k - y_1^k\|^2 + \left(1 - \frac{2}{\nu}\right)\|A(x^k - \hat{x}^k)\|^2.
\end{aligned}$$

Hence, from these relations and (40), the inequality (45) is equivalent to

$$\begin{aligned}
& \frac{\nu + 1}{\nu^2}\|\hat{y}_1^k - y_1^k\|^2 + \frac{\nu - 3}{\nu}\|A(\hat{x}^k - x^k)\|^2 \\
\le\ & \sigma_k^2 \left(\frac{\nu^2 - \nu + 1}{\nu^2}\|\hat{y}_1^k - y_1^k\|^2 + \frac{\nu - 1}{\nu}\|A(\hat{x}^k - x^k)\|^2 \right).
\end{aligned} \tag{46}$$

Since we have chosen $\nu > 2$, it holds that $\nu + 1 < \nu^2 - \nu + 1$ and there exists $\sigma_k < 1$ such that

$$\frac{\nu + 1}{\nu^2} < \sigma_k \frac{\nu^2 - \nu + 1}{\nu^2}, \qquad \frac{\nu - 3}{\nu} < \sigma_k \frac{\nu - 1}{\nu},$$

so that the inequality (46) (and, hence, (45)) is automatic.

Finally, the new iterates of VMHPDM Algorithm 2.1 are given by

$$\begin{aligned}
x^{k+1} &= \hat{x}^k = A^\dagger \hat{y}_1^k, \\
y_1^{k+1} &= A\hat{x}^k = Ax^{k+1}, \\
y_2^{k+1} &= y_2^k + A\hat{x}^k - \hat{y}_1^k = y_2^k + y_1^{k+1} - \hat{y}_1^k,
\end{aligned}$$

which is the same as in the splitting Algorithm 3.1.

Thus, convergence properties of Algorithm 3.1 are now given by Theorem 2.2, including the new rate of convergence results.

3.2. Proximal Alternating Directions Method

Consider the following structured variational inequality:

$$\text{Find } z^* \in \Omega \text{ such that } \langle \Phi(z^*), z - z^* \rangle \ge 0 \quad \forall z \in \Omega, \tag{47}$$

where

$$\Omega = \{(\xi, \zeta) \in K_1 \times K_2 \mid A\xi + B\zeta - b = 0\},$$
$$\Phi(\xi, \zeta) = (f(\xi), g(\zeta)), \quad f : K_1 \to \mathbb{R}^n, \quad g : K_2 \to \mathbb{R}^m,$$

with $K_1 \subseteq \mathbb{R}^n$, $K_2 \subseteq \mathbb{R}^m$ being closed convex sets, f and g being monotone functions, $A : \mathbb{R}^n \to \mathbb{R}^\ell$ and $B : \mathbb{R}^m \to \mathbb{R}^\ell$ being linear operators (matrices of appropriate dimensions), $b \in \mathbb{R}^\ell$.

We denote the solution set of (47) by $SOL(\Phi, \Omega)$. For a variational inequality associated to any other pair of a set and a mapping, the notation would be analogous. As is well known, associating a Lagrange multiplier to the equality constraint in the definition of Ω, (47) is equivalent to finding

$$(\xi^*, \zeta^*, \lambda^*) \in SOL(\Psi, Z), \tag{48}$$

where

$$Z = K_1 \times K_2 \times \mathbb{R}^\ell, \quad \Psi : Z \to \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^\ell, \quad \Psi(\xi, \zeta, \lambda) = \begin{pmatrix} f(\xi) - A^\top \lambda \\ g(\zeta) - B^\top \lambda \\ A\xi + B\zeta - b \end{pmatrix}. \tag{49}$$

The proximal alternating directions method (PADM) [17] (see also [13, 10] for related techniques) is the following iterative procedure.

Algorithm 3.2. Given $(\xi^k, \zeta^k, \lambda^k) \in K_1 \times K_2 \times \mathbb{R}^\ell$, $k \geq 0$,

1. Choose symmetric positive definite matrices H_k , R_k and S_k, of the dimensions $\ell \times \ell$, $n \times n$ and $m \times m$, respectively.
2. Find $\xi^{k+1} \in SOL(f_k, K_1)$, where
$$f_k(\xi) = f(\xi) - A^\top[\lambda^k - H_k(A\xi + B\zeta^k - b)] + R_k(\xi - \xi^k).$$
3. Find $\zeta^{k+1} \in SOL(g_k, K_2)$, where
$$g_k(\zeta) = g(\zeta) - B^\top[\lambda^k - H_k(A\xi^{k+1} + B\zeta - b)] + S_k(\zeta - \zeta^k).$$
4. Set $\lambda^{k+1} = \lambda^k - H_k(A\xi^{k+1} + B\zeta^{k+1} - b)$, $k := k = 1$ and go to Step 1.

First observe that problem (48) is equivalent to the variational inclusion

$$0 \in T(z) = F(\zeta, \xi, \lambda) \times [G(\zeta, \xi, \lambda) + H(\xi, \lambda)], \tag{50}$$

with

$$F : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^\ell \to \mathbb{R}^m, \qquad F(\zeta, \xi, \lambda) = g(\zeta) - B^\top \lambda + N_{K_2}(\zeta),$$

$$G : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^\ell \to \mathbb{R}^n \times \mathbb{R}^\ell, \qquad G(\zeta, \xi, \lambda) = \begin{pmatrix} f(\xi) - A^\top \lambda \\ A\xi + B\zeta - b \end{pmatrix},$$

$$H : \mathbb{R}^n \times \mathbb{R}^\ell \to \mathbb{R}^n \times \mathbb{R}^\ell, \qquad H(\xi, \lambda) = \begin{pmatrix} N_{K_1}(\xi) \\ 0 \end{pmatrix},$$

where $N_{K_1}(\cdot)$ and $N_{K2}(\cdot)$ are the normal cones to the sets K_1 and K_2, respectively.

It is evident that T defined above satisfies the Assumptions A1–A4 of Section 2. Taking $x = \zeta$ and $y = (\xi, \lambda)$, we apply Algorithm 2.1 to the operator T defined in (50), with the following choices of the algorithm parameters:

$$c_k = 1, \qquad G_1^k(\cdot, \cdot) = G(\zeta^k, \cdot, \cdot), \tag{51}$$

$$Q_k = \begin{pmatrix} R_k^{-1} & 0 \\ 0 & H_k \end{pmatrix}, \qquad P_k = (S_k + B^\top H_k B)^{-1}. \tag{52}$$

Observe that

$$G(x^k, y^k) - G_1^k(y^k) = 0, \qquad G(\hat{x}^k, \hat{y}^k) - G_1^k(\hat{y}^k) = (0, B(\hat{\zeta}^k - \zeta^k)). \tag{53}$$

We next show that VMHPDM Algorithm 2.1, with appropriate choice of the parameters, when applied to the operator T above generates the same iterates as PADM Algorithm 3.2.

`Forward-Backward Splitting step` (i.e., (12) with $\varepsilon_k^y = 0$, $e^k = 0$ and other parameters given by (51)–(52)) computes $\hat{y}^k \in \mathbb{R}^n \times \mathbb{R}^\ell$ such that

$$\begin{aligned} \hat{y}^k = (\hat{\xi}^k, \hat{\lambda}^k) &= y^k + Q_k\left[G_1^k(y^k) - G(x^k, y^k) - \hat{h}^k\right] \\ &= y^k - Q_k\hat{h}^k \\ &\in y^k - Q_k(H + G_1^k)(\hat{y}^k) \\ &= \begin{pmatrix} \xi^k - R_k^{-1}(N_{K_1}(\hat{\xi}^k) + f(\hat{\xi}^k) - A^\top\hat{\lambda}^k) \\ \lambda^k - H_k(A\hat{\xi}^k + B\zeta^k - b) \end{pmatrix}, \end{aligned}$$

where

$$\begin{aligned} &(H + G_1^k)(\hat{y}^k) \ni \hat{h}^k = (\hat{h}_1^k, A\hat{\xi}^k + B\zeta^k - b), \\ &N_{K_1}(\hat{\xi}^k) + f(\hat{\xi}^k) - A^\top\hat{\lambda}^k \ni \hat{h}_1^k = R_k(\xi^k - \hat{\xi}^k), \end{aligned} \tag{54}$$

and we have taken into account (53).

`(Exact) Proximal Step` (i.e., (13) with $\varepsilon_k^x = 0$, $r^k = 0$ and other parameters given by (51)–(52)) computes $\hat{\zeta}^k \in K_2$ and $\hat{u}^k \in \mathbb{R}^m$ such that

$$\begin{cases} \hat{u}^k \in F(\hat{x}^k, \hat{y}^k) = g(\hat{\zeta}^k) - B^\top\hat{\lambda}^k + N_{K_2}(\hat{\zeta}^k) \\ \qquad\qquad\quad = g(\hat{\zeta}^k) - B^\top[\lambda^k - H_k(A\hat{\xi}^k + B\zeta^k - b)] + N_{K_2}(\hat{\zeta}^k), \\ 0 = P_k\hat{u}^k + \hat{\zeta}^k - \zeta^k, \end{cases} \tag{55}$$

with

$$\|s^k\|_{Q_k^{-1}}^2 \le \sigma_k^2(\|\hat{\zeta}^k - \zeta^k\|_{P_k^{-1}}^2 + \|(\hat{\xi}^k\hat{\lambda}^k) - (\xi^k, \lambda^k)\|_{Q_k^{-1}}^2), \tag{56}$$

where (using (53) and (54))

$$\hat{w}_k = G(\hat{x}^k, \hat{y}^k) - G_1^k(\hat{y}^k) + \hat{h}^k = \begin{pmatrix} R_k(\xi^k - \hat{\xi}^k) \\ A\hat{\xi}^k + B\hat{\zeta}^k - b \end{pmatrix},$$

$$\begin{aligned} s_k &= Q_k\hat{w}^k + (\hat{\xi}^k, \hat{\lambda}^k) - (\xi^k, \lambda^k) \\ &= \begin{pmatrix} R_k^{-1}R_k(\xi^k - \hat{\xi}^k) + \hat{\xi}^k - \xi^k \\ H_k(A\hat{\xi}^k + B\hat{\zeta}^k - b) - H_k(A\hat{\xi}^k + B\zeta^k - b) \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ H_kB(\hat{\zeta}^k - \zeta^k) \end{pmatrix}. \end{aligned}$$

In the above, instead of the approximation condition (14) in the VMHPDM framework we use the stronger version (16), corresponding to (56), so that we can define the new iterates using the stepsize $\tau_k a_k = c_k$ in (15), i.e., `Iterates Update:` Stop if $\hat{\xi}^k = \xi^k$, $\hat{\zeta}^k = \zeta^k$ and $\hat{\lambda}^k = \lambda^k$, otherwise

$$\zeta^{k+1} = \zeta^k - P_k \hat{u}^k = \hat{\zeta}^k,$$
$$\xi^{k+1} = \xi^k - R_k^{-1} R_k(\xi^k - \hat{\xi}^k) = \hat{\xi}^k,$$
$$\lambda^{k+1} = \lambda^k - H_k(A\hat{\xi}^k + B\hat{\zeta}^k - b).$$

We next show that the iterates defined above are the same as those given by PADM Algorithm 3.2.

From the second relation in (54), we have that

$$\begin{aligned} N_{K_1}(\hat{\xi}^k) &\ni -\left(f(\hat{\xi}^k) - A^\top \hat{\lambda}^k + R_k(\hat{\xi}^k - \xi^k)\right) \\ &= -\left(f(\hat{\xi}^k) - A^\top[\lambda^k - H_k(A\hat{\xi}^k + B\zeta^k - b)] + R_k(\hat{\xi}^k - \xi^k)\right). \end{aligned}$$

By the definition of f_k in PADM Algorithm 3.2, this gives

$$-f_k(\hat{\xi}^k) \in N_{K_1}(\hat{\xi}^k),$$

i.e., $\xi^{k+1} = \hat{\xi}^k \in SOL(f_k, K_1)$.

Similarly, from (55) and (52), we have that

$$\begin{aligned} N_{K_2}(\hat{\zeta}^k) &\ni \hat{u}^k - \left(g(\hat{\zeta}^k) - B^\top[\lambda^k - H_k(A\hat{\xi}^k + B\zeta^k - b)]\right) \\ &= P_k^{-1}(\zeta^k - \hat{\zeta}^k) - \left(g(\hat{\zeta}^k) - B^\top[\lambda^k - H_k(A\hat{\xi}^k + B\zeta^k - b)]\right) \\ &= -\Big(g(\hat{\zeta}^k) - B^\top[\lambda^k - H_k(A\hat{\xi}^k + B\hat{\zeta}^k - b)] \\ &\qquad +(P_k^{-1} - B^\top H_k B)(\hat{\zeta}^k - \zeta^k)\Big) \\ &= -\left(g(\hat{\zeta}^k) - B^\top[\lambda^k - H_k(A\xi^{k+1} + B\hat{\zeta}^k - b)] + S_k(\hat{\zeta}^k - \zeta^k)\right) \end{aligned}$$

By the definition of g_k in PADM Algorithm 3.2, this gives

$$-g_k(\hat{\zeta}^k) \in N_{K_2}(\hat{\zeta}^k),$$

i.e., $\zeta^{k+1} = \hat{\zeta}^k \in SOL(g_k, K_2)$.

Finally, $\lambda^{k+1} = \lambda^k - H_k(A\xi^{k+1} + B\zeta^{k+1} - b)$ is the same as the update of the multiplier in PADM Algorithm 3.2.

Furthermore, the approximation condition (56) takes the form

$$\begin{aligned} &\|H_k B(\hat{\zeta}^k - \zeta^k)\|^2_{H_k^{-1}} \\ \leq\ & \sigma_k^2 \left(\|\hat{\zeta}^k - \zeta^k\|^2_{S_k} + \|\hat{\zeta}^k - \zeta^k\|^2_{B^\top H_k B} + \|(\hat{\xi}^k \hat{\lambda}^k) - (\xi^k, \lambda^k)\|^2_{Q_k^{-1}}\right) \end{aligned}$$

or, equivalently,

$$\begin{aligned} &\langle \hat{\zeta}^k - \zeta^k, B^\top H_k B(\hat{\zeta}^k - \zeta^k)\rangle \\ \leq\ & \frac{\sigma_k^2}{1 - \sigma_k^2}\left(\|\hat{\zeta}^k - \zeta^k\|^2_{S_k} + \|(\hat{\xi}^k \hat{\lambda}^k) - (\xi^k, \lambda^k)\|^2_{Q_k^{-1}}\right). \end{aligned} \tag{57}$$

Let μ_l and μ_u be respectively the lower and upper bounds for the eigenvalues of H_k and S_k. Then

$$\langle \hat{\zeta}^k - \zeta^k, B^\top H_k B(\hat{\zeta}^k - \zeta^k)\rangle \leq \mu_u \|B\|^2 \|\hat{\zeta}^k - \zeta^k\|^2$$

and

$$\mu_l \|\hat{\zeta}^k - \zeta^k\|^2 \leq \|\hat{\zeta}^k - \zeta^k\|^2_{S_k}.$$

Hence, the inequality (57) (equivalently, the approximation condition (56)) would be satisfied if

$$\frac{\sigma_k^2}{1-\sigma_k^2} \geq \frac{\mu_u\|B\|^2}{\mu_l},$$

and this can be ensured by choosing

$$1 > \sigma_k^2 \geq \frac{\mu_u\|B\|^2}{\mu_u\|B\|^2 + \mu_l}.$$

This completes the demonstration of the fact that PADM Algorithm 3.2 is a special case of VMHPDM Algorithm 2.1. Thus, convergence properties of Algorithm 3.2 are now given by Theorem 2.2, including the new rate of convergence results.

References

[1] H. H. Bauschke: Projection algorithms: results and open problems, in: Inherently Parallel Algorithms in Feasibility and Optimization and their Applications, D. Butnariu, Y. Censor, S. Reich (eds.), Studies in Computational Mathematics 8, Elsevier, Amsterdam (2001) 11–22.

[2] J. F. Bonnans, J. Ch. Gilbert, C. Lemaréchal, C. Sagastizábal: A family of variable metric proximal methods, Math. Program., Ser. A 68 (1995) 15–47.

[3] R. S. Burachik, A. N. Iusem, B. F. Svaiter: Enlargement of monotone operators with applications to variational inequalities, Set-Valued Anal. 5 (1997) 159–180.

[4] R. S. Burachik, C. A. Sagastizábal, B. F. Svaiter: ε-Enlargements of maximal monotone operators: Theory and applications, in: Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, M. Fukushima, L. Qi (eds.), Kluwer, Boston (1999) 25–44.

[5] R. S. Burachik, C. A. Sagastizábal, S. Scheimberg: An inexact method of partial inverses and a parallel bundle method, Optim. Methods Softw. 21 (2006) 385–400.

[6] J. V. Burke, M. Qian: A variable metric proximal point algorithm for monotone operators, SIAM J. Control Optimization 37 (1997) 353–375.

[7] J. V. Burke, M. Qian: On the local super-linear convergence of a matrix secant implementation of the variable metric proximal point algorithm for monotone operators, in: Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, M. Fukushima, L. Qi (eds.), Kluwer, Boston (1999) 317–334.

[8] J. V. Burke, M. Qian: On the superlinear convergence of the variable metric proximal point algorithm using Broyden and BFGS matrix secant updating, Math. Program., Ser. A 88 (2000) 157–181.

[9] X. Chen, M. Fukushima: Proximal quasi-Newton methods for nondifferentiable convex optimization, Math. Program., Ser. A 85 (1999) 313–334.

[10] X. Chen, M. Teboulle: A proximal-based decomposition method for convex minimization problems, Math. Program., Ser. A 64 (1994) 81–101.

[11] X. Chen, R. T. Rockafellar: Convergence rates in forward-backward splitting, SIAM J. Optim. 7 (1997) 421–444.

[12] J. Eckstein, D. P. Bertsekas: On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, Math. Program., Ser. A 55 (1992) 293–318.

[13] J. Eckstein: Some saddle-function splitting methods for convex programming, Optim. Methods Softw. 4 (1994) 75–83.

[14] J. Eckstein, M. C. Ferris: Operator splitting methods for monotone affine variational inequalities, with a parallel application to optimal control, INFORMS J. Comput. 10 (1998) 218–235.

[15] D. Gabay: Applications of the method of multipliers to variational inequalities, in: Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems, M. Fortin, R. Glowinski (eds.), North-Holland, Amsterdam (1983) 229–331.

[16] Y. García, M. Lassonde, J. P. Revalski: Extended sums and extended compositions of monotone operators, J. Convex Analysis 13 (2006) 721–738.

[17] B. He, L. Z. Liao, D. Han, H. Yang: A new inexact alternating directions method for monotone variational inequalities, Math. Program., Ser. A 92 (2002) 103–118.

[18] B. Lemaire: The proximal algorithm, in: New Methods of Optimization and Their Industrial Uses, J. P. Penot (ed.), International Series of Numerical Mathematics 87, Birkhäuser, Basel (1989) 73–87.

[19] C. Lemaréchal, C. Sagastizábal: Variable metric bundle methods: from conceptual to implementable forms, Math. Program., Ser. B 76 (1997) 393–410.

[20] P. L. Lions, B. Mercier: Splitting algorithms for the sum of two nonlinear operators, SIAM J. Numer. Anal. 16 (1979) 964–979.

[21] B. Martinet: Régularisation d'inéquations variationnelles par approximations successives, Rev. Franç. Inform. Rech. Opér. 4 (1970) 154–158.

[22] A. Ouorou: Epsilon-proximal decomposition method, Math. Program., Ser. A 99 (2004) 89–108.

[23] L. A. Parente, P. A. Lotito, M. V. Solodov: A class of inexact variable metric proximal point algorithms, SIAM J. Optim. 19 (2008) 240–260.

[24] T. Pennanen: Dualization of generalized equations of maximal monotone type, SIAM J. Optim. 10 (2000) 809–835.

[25] T. Pennanen: A splitting method for composite mappings, Numer. Funct. Anal. Optimization 23 (2002) 875–890.

[26] L. Qi, X. Chen: A preconditioning proximal Newton's method for nondifferentiable convex optimization, Math. Program., Ser. B 76 (1995) 411–430.

[27] S. M. Robinson: Composition duality and maximal monotonicity, Math. Program., Ser. A 85 (1999) 1–13.

[28] R. T. Rockafellar: On the maximality of sums of nonlinear monotone operators, Trans. Amer. Math. Soc. 149 (1970) 75–88.

[29] R. T. Rockafellar: Augmented Lagrangians and applications of the proximal point algorithm in convex programming, Math. Oper. Res. 1 (1976) 97–116.

[30] R. T. Rockafellar: Monotone operators and the proximal point algorithm, SIAM J. Control Optimization 14 (1976) 877–898.

[31] R. T. Rockafellar, J.-B. Wets: Variational Analysis, Springer, New York (1997).

[32] M. V. Solodov: A class of decomposition methods for convex optimization and monotone variational inclusions via the hybrid inexact proximal point framework, Optim. Methods Softw. 19 (2004) 557–575.

[33] M. V. Solodov, B. F. Svaiter: A globally convergent inexact Newton method for systems of monotone equations, in: Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, M. Fukushima, L. Qi (eds.), Kluwer, Boston (1999) 355–369.

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

[35] M. V. Solodov, B. F. Svaiter: A hybrid projection-proximal point algorithm, J. Convex Analysis 6 (1999) 59–70.

[36] M. V. Solodov, B. F. Svaiter: Error bounds for proximal point subproblems and associated inexact proximal point algorithms, Math. Program., Ser. B 88 (2000) 371–389.

[37] M. V. Solodov, B. F. Svaiter: A truly globally convergent Newton-type method for the monotone nonlinear complementarity problem, SIAM J. Optim. 10 (2000) 605–625.

[38] M. V. Solodov, B. F. Svaiter: A unified framework for some inexact proximal point algorithms, Numer. Funct. Anal. Optimization 22 (2001) 1013–1035.

[39] M. V. Solodov, B. F. Svaiter: A new proximal-based globalization strategy for the Josephy-Newton method for variational inequalities, Optim. Methods Softw. 17 (2002) 965–983.

[40] J. E. Spingarn: Applications of the method of partial inverses to convex programming, Math. Program. 32 (1985) 199–223.

[41] P. Tseng: Applications of a splitting algorithm to decomposition in convex programming and variational inequalities, SIAM J. Control Optimization 29 (1991) 119–138.

[42] P. Tseng: Alternating projection-proximal methods for convex programming and variational inequalities, SIAM J. Optim. 7 (1997) 951–965.

[43] P. Tseng: A modified forward-backward splitting method for maximal monotone mappings, SIAM J. Control Optimization 38 (2000) 431–446.