Communications in Optimization Theory

Available online at http://cot.mathres.org

ON CONSTRUCTION OF SPLITTING CONTRACTION ALGORITHMS IN A PREDICTION-CORRECTION FRAMEWORK FOR SEPARABLE CONVEX OPTIMIZATION

BINGSHENG HE[1], XIAOMING YUAN[2,*]

[1]Department of Mathematics, Nanjing University, Nanjing, China
[2]Department of Mathematics, The University of Hong Kong, Hong Kong

Dedicated to the memory of Late Professor Roland Glowinski (1937-2022)

Abstract. In the past decade, we had developed a series of splitting contraction algorithms for separable convex optimization problems, at the root of the alternating direction method of multipliers. Convergence of these algorithms was studied under specific model-tailored conditions, while these conditions can be conceptually abstracted as two generic conditions when these algorithms are all unified as a prediction-correction framework. In this paper, in turn, we showcase a constructive way for specifying the generic convergence-guaranteeing conditions, via which new splitting contraction algorithms can be generated automatically. It becomes possible to design more application-tailored splitting contraction algorithms by specifying the prediction-correction framework, while proving their convergence is a routine.

1. INTRODUCTION

To understand the convergence rate of the well-known alternating direction method of multipliers (ADMM) proposed in [9], we had initiated an analytic technique in [19] and further used it to conduct convergence analysis for a series of ADMM-based splitting contraction algorithms for separable convex programming problems in our previous works such as [15, 16, 17, 18, 21]. As summarized in [12] (see also [13, 22]), the key of this analytic technique is to represent the ADMM or its variants in [15, 16, 17, 18, 21] by a prediction-correction framework in the context of the variational inequality reformulation of the convex programming model under discussion, and then to analyze the contraction property of the iterative scheme to derive the

convergence. In particular, various specific conditions for ensuring convergence of the algorithms in [15, 16, 17, 18, 21] can be conceptually abstracted as two generic conditions. Since the ADMM is an application of the Douglas-Rachford splitting method in [5, 23] and convergence of the mentioned algorithms can be proved by the contraction property, we call them "splitting contraction algorithms" uniformly. Our previous result in [3] showing the divergence of the direct extension of the ADMM to multiple block separable convex programming problems was also based on this analytical technique.

In this paper, we revisit the unified framework in [12] (see also [13, 18, 22]), including the prediction-correction representation of various splitting contraction algorithms and the generic convergence-guaranteeing conditions, and showcase how to constructively specify the generic convergence-guaranteeing conditions. Through the procedure, new splitting contraction algorithms can be automatically generated via specifying the prediction-correction framework. The constructive way provides flexibility to designing various application-tailored splitting contraction algorithms for given specific separable convex optimization models, while proving their convergence is a routine.

To understand the convergence rate of the well-known alternating direction method of multipliers (ADMM) proposed in [9], we had initiated an analytic technique in [19] and further used it to conduct convergence analysis for a series of ADMM-based splitting contraction algorithms for separable convex programming problems in our previous works such as [15, 16, 17, 18, 21]. As summarized in [12] (see also [13, 22]), the key of this analytic technique is to represent the ADMM or its variants in [15, 16, 17, 18, 21] by a prediction-correction framework in the context of the variational inequality reformulation of the convex programming model under discussion, and then to analyze the contraction property of the iterative scheme to derive the convergence. In particular, various specific conditions for ensuring convergence of the algorithms in [15, 16, 17, 18, 21] can be conceptually abstracted as two generic conditions. Since the ADMM is an application of the Douglas-Rachford splitting method in [5, 23] and convergence of the mentioned algorithms can be proved by the contraction property, we call them "splitting contraction algorithms" uniformly. Our previous result in [3] showing the divergence of the direct extension of the ADMM to multiple block separable convex programming problems was also based on this analytical technique.

In this paper, we revisit the unified framework in [12] (see also [13, 18, 22]), including the prediction-correction representation of various splitting contraction algorithms and the generic convergence-guaranteeing conditions, and showcase how to constructively specify the generic convergence-guaranteeing conditions. Through the procedure, new splitting contraction algorithms can be automatically generated via specifying the prediction-correction framework. The constructive way provides flexibility to designing various application-tailored splitting contraction algorithms for given specific separable convex optimization models, while proving their convergence is a routine.

2. UNIFIED FRAMEWORK

We consider the canonical convex minimization problem with linear constraints

$$\min\{\theta(u) \,|\, \mathscr{A}u = b, \, u \in \mathscr{U}\} \tag{2.1}$$

where $\theta : \Re^n \to \Re$ is a closed, proper, and convex but not necessarily smooth function, $\mathscr{U} \subseteq \Re^n$ is a closed convex set, $\mathscr{A} \in \Re^{m \times n}$, and $b \in \Re^m$. For further analysis, we recall the prediction-correction representation of splitting contraction algorithms in the existing literatures [15, 16, 17, 18, 22] in the context of variational inequalities, and the generic convergence-guaranteeing conditions.

2.1. **Variational inequality reformulation.** The Lagrangian function of (2.1) is

$$L(u,\lambda) = \theta(u) - \lambda^T (\mathscr{A}u - b) \tag{2.2}$$

with $\lambda \in \Re^m$ the Lagrange multiplier. A pair of $(u^*, \lambda^*) \in \mathscr{U} \times \Re^m$ is called a saddle point of $L(u,\lambda)$ if

$$L_{\lambda \in \Re^m}(u^*, \lambda) \leq L(u^*, \lambda^*) \leq L_{u \in \mathscr{U}}(u, \lambda^*).$$

As analyzed in, e.g. [13, 15, 16, 17, 19, 21, 22], (u^*, λ^*) can be characterized as a solution point of the following variational inequality (VI):

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall w \in \Omega. \tag{2.3}$$

where

$$w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -\mathscr{A}^T \lambda \\ \mathscr{A}u - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathscr{U} \times \Re^m. \tag{2.4}$$

We denote by Ω^* the set of solution points of the VI (2.3)-(2.4).

2.2. **Prediction-correction framework and convergence conditions.** In the context of the VI (2.3)-(2.4), the splitting contraction algorithms proposed in [15, 16, 17, 18, 22] can all be uniformly represented by a prediction-correction framework, and their specific convergence-guaranteeing conditions can be conceptually abstracted as two generic conditions. Below, note that $v = w$, or v could be a subvector of w, and $\mathscr{V}^*$ is denoted as the set of all v^* whose corresponding $w^* \in \Omega^*$. Dimensionality of the matrices Q and M is conformed with that of v.

Prediction-correction framework for the VI (2.3)-(2.4)

[Prediction Step.] With given v^k, find a vector $\tilde{w}^k \in \Omega$ such that

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \forall w \in \Omega, \tag{2.5}$$

where the matrix Q is not necessarily symmetric but $Q^T + Q$ is assumed to be positive definite.

[Correction Step.] Find a nonsingular matrix M and update v by

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{2.6}$$

Convergence conditions

For the matrices Q and M used in (2.5) and (2.6), respectively, there exists a matrix $H \succ 0$ such that

$$HM = Q, \tag{2.7a}$$

and

$$G := Q^T + Q - M^T H M \succ 0. \tag{2.7b}$$

Remark 2.1. Note that the matrices Q and M take specific forms for the splitting contraction algorithms in [15, 16, 17, 18, 22] when different separable cases of the model (2.1) are studied. With the specifically given Q and M, it is easy to verify whether or not the conditions (2.7) are

satisfied. How to construct the matrices H and M with a given Q to ensure the convergence conditions (2.7), however, has not yet been discussed in the literature. This reverse procedure leads to the specification of a concrete splitting contraction algorithm for the model (2.1) via the prediction-correction framework (2.5)-(2.6). A trivial case is when Q is known to be even symmetric and positive definite. Then, we can simple take $M = \alpha I$ with $\alpha \in (0,2)$ and then $H = \frac{1}{\alpha}Q$, which certainly ensures the conditions (2.7). This case is indeed an application of the classic proximal point algorithm in [24], as studied in [2, 11, 14, 20]. We would emphasize that the corresponding matrix Q is generally asymmetric when many algorithms including those in [15, 16, 17, 18, 22] are represented by the prediction-correction framework (2.5)-(2.6). Hence, we focus on the nontrivial case where Q is asymmetric to discuss how to constructively specify the prediction-correction framework (2.5)-(2.6) and the convergence conditions (2.7).

2.3. **Convergence.** Convergence of the prediction-correction framework (2.5)-(2.6) under the conditions (2.7) has been shown in our previous works such as [12, 13, 18]. The main contraction property ensuring the convergence is summarized in the following theorem; we include the proof for completeness.

Theorem 2.2. *Let $\{v^k\}$ be the sequence generated by the prediction-correction framework (2.5)-(2.6) under the conditions (2.7). Then, it holds that*

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.8}$$

Proof. Using $Q = HM$ (see (2.7a)), the prediction step can be written as

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T HM(v^k - \tilde{v}^k), \quad \forall w \in \Omega.$$

Then, it follows from (2.6) that

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H(v^k - v^{k+1}), \quad \forall w \in \Omega.$$

Setting $w = w^*$ in the above inequality, we get

$$(v^k - v^{k+1})^T H(\tilde{v}^k - v^*) \geq \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k), \quad \forall w^* \in \Omega^*.$$

Note that $(\tilde{w}^k - w^*)^T F(\tilde{w}^k) = (\tilde{w}^k - w^*)^T F(w^*)$ and the optimality of w^*. Thus, we have

$$(v^k - v^{k+1})^T H(\tilde{v}^k - v^*) \geq 0, \quad \forall v^* \in \mathcal{V}^*. \tag{2.9}$$

Setting $a = v^k, b = v^{k+1}, c = \tilde{v}^k$ and $d = v^*$ in the identity

$$2(a - b)^T H(c - d) = \left\{ \|a - d\|_H^2 - \|b - d\|_H^2 \right\} - \left\{ \|a - c\|_H^2 - \|b - c\|_H^2 \right\},$$

we know from (2.9) that

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \geq \|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2. \tag{2.10}$$

For the right-hand side of the last inequality, we have

$$
\begin{aligned}
\|v^k - \tilde{v}^k\|_H^2 &- \|v^{k+1} - \tilde{v}^k\|_H^2 \\
&= \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\
&\overset{(2.6)}{=} \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\
&= 2(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\
&= (v^k - \tilde{v}^k)^T (Q^T + Q - M^T HM)(v^k - \tilde{v}^k) \\
&\overset{(2.7b)}{=} \|v^k - \tilde{v}^k\|_G^2. \tag{2.11}
\end{aligned}
$$

Substituting (2.11) in (2.10), the assertion (2.8) is proved. □

2.4. **Naming of matrices.** Because of clear reasons, we call Q and M the prediction matrix and correction matrix, respectively. Moreover, the inequality (2.8) compares the proximity of two consecutive iterates to v^* in the H-norm, and measures their difference (or progress) by the term $\|v^k - \tilde{v}^k\|_G^2$. Hence, we call H and G the norm matrix and profit matrix, respectively.

2.5. **Example.** To see how the prediction-correction framework (2.5)-(2.6) and the convergence conditions (2.7) are related to our previous works such as [3, 15, 16, 17, 18, 19, 21, 22], we take the strictly contractive Peaceman-Rachford splitting method (PRSM) proposed in [15] as an illustrative example.

2.5.1. *SC-PRSM.* We consider a two-block separable convex programming model

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b, x \in \mathscr{X}, y \in \mathscr{Y}\}, \tag{2.12}$$

where $\theta_i : \Re^{n_i} \to \Re$ $(i = 1,2)$ are closed, proper, and convex but not necessarily smooth functions; $\mathscr{X} \subseteq \Re^{n_1}$ and $\mathscr{Y} \subseteq \Re^{n_2}$ are closed convex sets; $A \in \Re^{m \times n_1}$ and $B \in \Re^{m \times n_2}$; $b \in \Re^m$; and $n = n_1 + n_2$. This is a separable case of (2.1) with

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \theta(u) = \theta_1(x) + \theta_2(y), \quad \mathscr{A} = (A, B) \quad \text{and} \quad \mathscr{U} = \mathscr{X} \times \mathscr{Y}.$$

Thus, the VI (2.3)-(2.4) can be specified as the following:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall w \in \Omega. \tag{2.13a}$$

where

$$w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T \lambda \\ -B^T \lambda \\ Ax + By - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathscr{X} \times \mathscr{Y} \times \Re^m. \tag{2.13b}$$

Let the augmented Lagrangian function of (2.12) be

$$\mathscr{L}_\beta^{[2]}(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax + By - b) + \frac{\beta}{2}\|Ax + By - b\|^2, \tag{2.14}$$

with $\lambda \in \Re^m$ the Lagrange multiplier and $\beta > 0$ the penalty parameter. The well-known alternating direction method of multipliers (ADMM) proposed by Glowinski and Marrocco in [9] for solving (2.12) is

$$\begin{cases} x^{k+1} \in \arg\min\{\mathscr{L}_\beta^{[2]}(x, y^k, \lambda^k) \mid x \in \mathscr{X}\}, \\ y^{k+1} \in \arg\min\{\mathscr{L}_\beta^{[2]}(x^{k+1}, y, \lambda^k) \mid y \in \mathscr{Y}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{cases} \tag{2.15}$$

As shown in [7], the ADMM (2.15) is an application of the Douglas-Rachford splitting method in [5, 23] to the dual of (2.12). In [15], we proposed the following strictly contractive PRSM (SC-PRSM):

$$\text{(SC-PRSM)} \begin{cases} x^{k+1} \in \arg\min\{\mathscr{L}_\beta^{[2]}(x, y^k, \lambda^k) \mid x \in \mathscr{X}\}, & \text{(2.16a)} \\[2mm] \lambda^{k+\frac{1}{2}} = \lambda^k - \mu\beta(Ax^{k+1} + By^k - b), & \text{(2.16b)} \\[2mm] y^{k+1} \in \arg\min\{\mathscr{L}_\beta^{[2]}(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \mid y \in \mathscr{Y}\}, & \text{(2.16c)} \\[2mm] \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \mu\beta(Ax^{k+1} + By^{k+1} - b), & \text{(2.16d)} \end{cases}$$

where $\mu \in (0,1)$. It has been shown that if $\mu = 1$ in (2.16), then the resulting scheme (which can be regarded as a symmetric version of the ADMM (2.15)) is an application of the PRSM in [25] to the dual of (2.12) while its convergence is not guaranteed as shown in [4]. For the SC-PRSM (2.16) with $\mu \in (0,1)$, it was shown in [15] that its sequence of iterations is strictly contractive with respect to the solution set and hence its convergence is guaranteed.

2.5.2. *Prediction-correction representation.* The follow lemma is useful and its proof can be found in, e.g. [1].

Lemma 2.3. *Let $\mathscr{X} \subset \mathfrak{R}^n$ be a closed convex set, $\theta(x)$ and $f(x)$ be convex functions and $f(x)$ is differentiable on an open set which contains $\mathscr{X}$. Assume that the solution set of the minimization problem $\min\{\theta(x) + f(x) \mid x \in \mathscr{X}\}$ is nonempty. Then,*

$$x^* \in \arg\min\{\theta(x) + f(x) \mid x \in \mathscr{X}\} \tag{2.17a}$$

if and only if

$$x^* \in \mathscr{X}, \ \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \geq 0, \ \forall x \in \mathscr{X}. \tag{2.17b}$$

Let us define $\tilde{w}^k$ as

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^k - \beta(Ax^{k+1} + By^k - b) \end{pmatrix}, \tag{2.18}$$

where (x^{k+1}, y^{k+1}) is generated by (2.16). Now, we show that the SC-PRSM (2.16) with $\tilde{w}^k$ defined in (2.18) can be represented by the prediction-correction framework (2.5)-(2.6) .

Theorem 2.4. *The SC-PRSM (2.16) corresponds to the prediction-correction framework (2.5)-(2.6) with*

$$Q = \begin{pmatrix} \beta B^T B & -\mu B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \quad and \quad M = \begin{pmatrix} I & 0 \\ -\mu\beta B & 2\mu I_m \end{pmatrix}. \tag{2.19}$$

Proof. It follows from (2.18) that $\tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b)$. Also, the optimality condition of the x-subproblem (2.16a) is

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T(-A^T\tilde{\lambda}^k) \geq 0, \ \forall x \in \mathscr{X}. \tag{2.20}$$

Notice that the objective function of the y-subproblem (2.16c) is

$$\mathscr{L}_\beta^{[2]}(\tilde{x}^k, y, \lambda^{k+\frac{1}{2}}) = \theta_1(\tilde{x}^k) + \theta_2(y) - (\lambda^{k+\frac{1}{2}})^T(A\tilde{x}^k + By - b) + \frac{\beta}{2}\|A\tilde{x}^k + By - b\|^2.$$

Ignoring some constant term in the y-subproblem, we obtain

$$\tilde{y}^k = \arg\min\{\theta_2(y) - (\lambda^{k+\frac{1}{2}})^T By + \frac{\beta}{2}\|A\tilde{x}^k + By - b\|^2 \mid y \in \mathscr{Y}\}.$$

Consequently, according to Lemma 2.3, we have

$$\tilde{y}^k \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \left\{ -B^T \lambda^{k+\frac{1}{2}} + \beta B^T (A\tilde{x}^k + B\tilde{y}^k - b) \right\} \geq 0, \ \forall y \in \mathcal{Y}.$$

Using $\tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b)$, we get

$$\lambda^{k+\frac{1}{2}} = \lambda^k - \mu(\lambda^k - \tilde{\lambda}^k) = \tilde{\lambda}^k + (1-\mu)(\lambda^k - \tilde{\lambda}^k),$$

and

$$\beta(A\tilde{x}^k + By^k - b) = (\tilde{\lambda}^k - \lambda^k).$$

Finally, the optimality condition of the y-subproblem can be written as $\tilde{y}^k \in \mathcal{Y}$, and

$$\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \left\{ -B^T \tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) - \mu B^T(\tilde{\lambda}^k - \lambda^k) \right\} \geq 0, \ \forall y \in \mathcal{Y}. \quad (2.21)$$

According to the definition of $\tilde{w}^k$ in (2.18), and $\tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b)$, we have

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. \quad (2.22)$$

Combining (2.20), (2.21) and (2.22), and using the notation of (2.13), we obtain

$$\tilde{w}^k \in \Omega, \ \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \forall w \in \Omega, \quad (2.23)$$

which corresponds to the prediction step (2.5) with Q defined in (2.19).

Moreover, note that λ^{k+1} in (2.16d) can be represented as

$$\begin{aligned} \lambda^{k+1} &= [\lambda^k - \mu(\lambda^k - \tilde{\lambda}^k)] - \mu[-\beta B(y^k - \tilde{y}^k) + \beta(Ax^{k+1} + By^k - b)] \\ &= \lambda^k - [-\mu\beta B(y^k - \tilde{y}^k) + 2\mu(\lambda^k - \tilde{\lambda}^k)]. \end{aligned} \quad (2.24)$$

Thus, together with $y^{k+1} = \tilde{y}^k$, we have

$$\begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I & 0 \\ -\mu\beta B & 2\mu I_m \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}.$$

This can be rewritten as the compact form

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \quad (2.25)$$

which corresponds to the correction step (2.6) with M defined in (2.19). $\square$

Hence, we can represent the SC-PRSM (2.16) as the prediction-correction framework (2.5)-(2.6) with the corresponding matrices Q and M defined in (2.19).

2.5.3. *Convergence conditions.* To verify that the matrices Q and M defined in (2.19) satisfy the conditions (2.7) and hence ensure the convergence of the SC-PRSM (2.16), we see that

$$M^{-1} = \begin{pmatrix} I & 0 \\ \frac{1}{2}\beta B & \frac{1}{2\mu} I_m \end{pmatrix}.$$

Hence, if we define

$$H = QM^{-1} = \begin{pmatrix} \beta B^T B & -\mu B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} I & 0 \\ \frac{1}{2}\beta B & \frac{1}{2\mu} I_m \end{pmatrix} = \frac{1}{2} \begin{pmatrix} (2-\mu)\beta B^T B & -B^T \\ -B & \frac{1}{\mu\beta} I_m \end{pmatrix},$$

then H is positive definite for any $\mu \in (0,1)$ when B is full column rank, and the condition (2.7a) is satisfied.

To check the condition (2.7b), i.e., the positiveness of $G := Q^T + Q - M^T HM$, we have

$$M^T HM = M^T Q = \begin{pmatrix} I & -\mu\beta B^T \\ 0 & 2\mu I_m \end{pmatrix} \begin{pmatrix} \beta B^T B & -\mu B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} = \begin{pmatrix} (1+\mu)\beta B^T B & -2\mu B^T \\ -2\mu B & \frac{2\mu}{\beta} I_m \end{pmatrix}.$$

Using (2.19) and the above equation, we have

$$G = Q^T + Q - M^T HM = (1-\mu) \begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{2}{\beta} I_m \end{pmatrix},$$

which implies that G is positive definite for any $\mu \in (0, 1)$ when B is full column rank. Hence, the convergence conditions (2.7) are satisfied for the strictly contractive PRSM (2.16).

3. CONSTRUCTION OF NEW SPLITTING CONTRACTION ALGORITHMS

As shown before, the prediction-correction framework (2.5)-(2.6), along with the generic convergence-guaranteeing conditions (2.7), represents a unified and abstract roadmap to convergence analysis for various splitting contraction algorithms in [15, 16, 17, 18, 22]. In this section, in turn, we focus on how to specify the prediction-correction framework (2.5)-(2.6) and the conditions (2.7). More specifically, with a given Q, we construct H and M to satisfy the conditions (2.7), and each pair of H and M can automatically generate a specific splitting contraction algorithm with provable convergence through the prediction-correction framework (2.5)-(2.6). The constructive way is a principle of designing new application-tailored splitting contraction algorithms when concrete applications of the canonical convex programming model (2.1) are considered.

To design a new algorithm, it is rare to start from scratch; it is more often to start from a given coarse scheme that might be imperfect in theoretical or numerical aspects. Our discussion starts from the scenario where the matrix Q is already determined by, e.g., a given coarse iterative scheme such as (4.5), (5.4), or (5.9), which will be delineated in the next two sections. With a given Q, our recept essentially only requires determining the norm matrix H and the correction matrix M for the prediction-correction framework (2.5)-(2.6), while only the correction matrix M needs to be specified to implement the resulting splitting contraction algorithm. Certainly, we prefer to construct such M that can alleviate the resulting implementation/computation. We further assume that the given matrix Q satisfies

$$Q^T + Q \succ 0.$$

Below, we give two specific principles of constructing the matrices H and M which can ensure the conditions (2.7).

3.1. Construction from the condition (2.7a). Note that the condition (2.7a) can be rewritten as

$$H = QM^{-1}. \tag{3.1}$$

Since the norm matrix H is required to be symmetric and positive definite, the condition (3.1) implies that H should be representable in form of

$$H = QD^{-1}Q^T, \tag{3.2}$$

in which the matrix D is a undetermined positive definite matrix. Indeed, by comparing (3.1) with (3.2), we know that $M^{-1} = D^{-1}Q^T$ and thus

$$M = Q^{-T}D. \tag{3.3}$$

Hence, although the matrix D in (3.2) is still unknown, choosing M as (3.3) can ensure the condition (2.7a).

Now, we investigate the restriction on D to ensure the condition (2.7b) with the matrix M given as (3.3). Notice that

$$M^T H M = \left(DQ^{-1}\right)\left(QD^{-1}Q^T\right)\left(Q^{-T}D\right) = D. \tag{3.4}$$

With (3.4), then the condition (2.7b) is reduced to

$$G := Q^T + Q - M^T H M = Q^T + Q - D \succ 0. \tag{3.5}$$

Hence, to ensure the condition (2.7b), the only restriction on the positive definite matrix D in (3.2) is

$$0 \prec D \prec Q^T + Q, \tag{3.6}$$

In other words, whenever Q is given and it satisfies $Q^T + Q \succ 0$, then both H and M can be constructed via the following steps:

$$\left\{\begin{array}{rcl} HM & = & Q, \\ M^T H M & = & D. \end{array}\right. \quad \Longleftrightarrow \quad \left\{\begin{array}{rcl} HM & = & Q, \\ Q^T M & = & D. \end{array}\right. \quad \Longleftrightarrow \quad \left\{\begin{array}{rcl} H & = & QD^{-1}Q^T, \\ M & = & Q^{-T}D. \end{array}\right. \tag{3.7}$$

Through this construction, both the conditions (2.6) and (2.7b) are guaranteed to be satisfied. Note that once the matrix D is chosen according to (3.6), the matrices H, M and G are all uniquely determined. Then, with the specified matrix M in (3.3), the correction step (2.6) and thus the prediction-correction framework (2.5)-(2.6) is also specified as a concrete contraction splitting algorithm for the VI(2.3)-(2.4).

3.2. Construction from the condition (2.7b).
Alternatively, we can start from the condition (2.7b) to construct the norm matrix H and the correction matrix M. Again, with a given Q satisfying $Q^T + Q \succ 0$, we can choose the profit matrix G such that

$$0 \prec G \prec Q^T + Q. \tag{3.8}$$

Denote

$$\Delta = Q^T + Q - G, \tag{3.9}$$

which is positive definite. According to (2.7b), we know that the matrices H and M should satisfy

$$M^T H M = \Delta.$$

Recall the condition (2.7a): $HM = Q$. Thus, with a chosen G satisfying (3.8), H and M can be constructed vis the following steps:

$$\left\{\begin{array}{rcl} M^T H M & = & \Delta, \\ HM & = & Q. \end{array}\right. \quad \Leftrightarrow \quad \left\{\begin{array}{rcl} Q^T M & = & \Delta, \\ HM & = & Q. \end{array}\right. \quad \Leftrightarrow \quad \left\{\begin{array}{rcl} M & = & Q^{-T}\Delta, \\ H & = & Q\Delta^{-1}Q^T. \end{array}\right. \tag{3.10}$$

Then, with the constructed matrix M in (3.10), the correction step (2.6) and thus the prediction-correction framework (2.5)-(2.6) can also be specified as a concrete splitting contraction algorithm for the VI(2.3)-(2.4). Again, with a given G satisfying (3.8), the matrices H and M are both uniquely determined.

3.3. **Choices of matrices.** It is interesting to observe that the proposed two construction strategies can be related via the relationship

$$D \succ 0, \quad G \succ 0, \quad \text{and} \quad D + G = Q^T + Q. \tag{3.11}$$

Hence, once D is chosen for the construction strategy in Section 3.1, the corresponding G given by (3.11) can be used for the construction strategy in Section 3.2, and vice versa.

Our discussions above show that, with a given prediction matrix Q satisfying $Q^T + Q \succ 0$, once D (Resp., G) is chosen according to (3.6) (Resp., (3.8)), the matrices H and M can be determined as analyzed and thus a concrete contraction splitting algorithm can be specified for the convex programming problem (2.1) via the prediction-correction framework (2.5)-(2.6). Technically, there are infinitely many such choices subject to (3.11). For example, we can choose

$$D = \alpha[Q^T + Q] \quad \text{and} \quad G = (1 - \alpha)[Q^T + Q], \quad \alpha \in (0, 1).$$

We will elaborate on the choice $D = G = \frac{1}{2}[Q^T + Q]$ in Section 4.3.3.

3.4. **Implementation of the correction step (2.6).** Note that the correction step (2.6) can be rewritten as

$$Q^T(v^{k+1} - v^k) = Q^T M(v^k - \tilde{v}^k).$$

To implement the correction step (2.6) with the constructed two choices for M, i.e., $M = Q^{-T}D$ in (3.3) and $M = Q^{-T}\Delta$ in (3.10), we need to solve one of the following systems of equations:

$$Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k), \tag{3.12}$$

and

$$Q^T(v^{k+1} - v^k) = \Delta(\tilde{v}^k - v^k). \tag{3.13}$$

Hence, although D and G (thus Δ) can be chosen arbitrarily with the only constraint (3.6) or (3.8), it is preferred to choose some model-tailored ones that can favor solving the systems of equations (3.12) or (3.13) more efficiently. Irrational choices that make the correction step (2.6) complicated should be generally avoided. Some examples will be discussed in the next sections when specific cases of the canonical convex programming model (2.1) are considered.

4. APPLICATION TO THREE-BLOCK SEPARABLE CONVEX OPTIMIZATION

In this section, we apply the strategies proposed in Sections 3.1, 3.2 and 3.3 to a separable convex optimization problem, and showcase how to construct the norm matrix H and the correction matrix M when the matrix Q is given.

4.1. **Model.** We consider the three-block separable convex optimization model with linear constraints

$$\min\{\theta_1(x) + \theta_2(y) + \theta_3(z) \,|\, Ax + By + Cz = b, x \in \mathscr{X}, y \in \mathscr{Y}, z \in \mathscr{Z}\}, \tag{4.1}$$

where $\theta_i : \Re^{n_i} \to \Re$ $(i = 1, 2, 3)$ are closed, proper, and convex but not necessarily smooth functions; $\mathscr{X} \subseteq \Re^{n_1}$, $\mathscr{Y} \subseteq \Re^{n_2}$ and $\mathscr{Z} \subseteq \Re^{n_3}$ are closed convex sets; $A \in \Re^{m \times n_1}$, $B \in \Re^{m \times n_2}$ and $C \in \Re^{m \times n_3}$; $b \in \Re^m$; and $n_1 + n_2 + n_3 = n$. Clearly, it is a special case of the canonical convex programming problem (2.1), and the VI (2.3)-(2.4) can be specified as the following:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall w \in \Omega, \tag{4.2}$$

where

$$
w = \begin{pmatrix} x \\ y \\ z \\ \lambda \end{pmatrix}, \quad u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ -C^T\lambda \\ Ax+By+Cz-b \end{pmatrix}, \tag{4.3a}
$$

with

$$
\theta(u) = \theta_1(x) + \theta_2(y) + \theta_3(z), \qquad \Omega = \mathscr{X} \times \mathscr{Y} \times \mathscr{Z} \times \mathfrak{R}^m. \tag{4.3b}
$$

Let the augmented Lagrangian function of the model (4.1) be

$$
\mathscr{L}_\beta^{[3]}(x,y,z,\lambda) = \theta_1(x) + \theta_2(y) + \theta_3(z) - \lambda^T(Ax+By+Cz-b) + \frac{\beta}{2}\|Ax+By+Cz-b\|^2 \tag{4.4}
$$

with $\lambda \in \mathfrak{R}^m$ the Lagrange multiplier and $\beta > 0$ the penalty parameter. With the success of the ADMM (see, e.g., [6, 8, 10]), it is natural to consider directly extending the ADMM (2.15) and splitting the augmented Lagrangian function $\mathscr{L}_\beta^{[3]}(x,y,z,\lambda)$ in (4.4) three times by the Gauss-Seidel manner. That is, consider the scheme

$$
\begin{cases}
x^{k+1} \in \arg\min\{\mathscr{L}_\beta^{[3]}(x,y^k,z^k,\lambda^k) \mid x \in \mathscr{X}\}, \\
y^{k+1} \in \arg\min\{\mathscr{L}_\beta^{[3]}(x^{k+1},y,z^k,\lambda^k) \mid y \in \mathscr{Y}\}, \\
z^{k+1} \in \arg\min\{\mathscr{L}_\beta^{[3]}(x^{k+1},y^{k+1},z,\lambda^k) \mid z \in \mathscr{Z}\}, \\
\lambda^{k+1} = \lambda^k - (Ax^{k+1} + By^{k+1} + Cz^{k+1} - b).
\end{cases} \tag{4.5}
$$

However, the splitting scheme (4.5) is coarse in sense of that its convergence is not guaranteed as shown in [3]. Thus, to render convergence, either the scheme (4.5) should be appropriately adjusted or stronger conditions on functions/coefficient matrices/penalty parameters should be additionally assumed. To develop algorithms at the root of the ADMM (2.15), we can consider correcting the output of (4.5) by certain correction steps as what we did in [16, 17, 18, 22]. Below we show how to apply the construction strategies in Section 3 to generate splitting contraction algorithms by taking advantage of the coarse splitting scheme (4.5).

4.2. **Discerning the prediction matrix** Q. Our construction starts from the coarse splitting scheme (4.5) which can be rewritten as the prediction step (2.5) and hence the corresponding prediction matrix Q can be discerned. For this purpose, we first consider the subproblems related to the primal variables in (4.5), and rewrite them as $\tilde{u}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{z}^k) \in \mathscr{X} \times \mathscr{Y} \times \mathscr{Z}$. Namely, we have

$$
\begin{cases}
\tilde{x}^k \in \arg\min\{\mathscr{L}_\beta^{[3]}(x,y^k,z^k,\lambda^k) \mid x \in \mathscr{X}\}, \\
\tilde{y}^k \in \arg\min\{\mathscr{L}_\beta^{[3]}(\tilde{x}^k,y,z^k,\lambda^k) \mid y \in \mathscr{Y}\}, \\
\tilde{z}^k \in \arg\min\{\mathscr{L}_\beta^{[3]}(\tilde{x}^k,\tilde{y}^k,z,\lambda^k) \mid z \in \mathscr{Z}\}.
\end{cases} \tag{4.6}
$$

Ignoring some constant terms, we can rewrite the formula above as

$$
\begin{cases}
\tilde{x}^k \in \arg\min\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax+By^k+Cz^k-b\|^2 \mid x \in \mathscr{X}\}, \\
\tilde{y}^k \in \arg\min\{\theta_2(y) - y^T B^T \lambda^k + \frac{\beta}{2}\|A\tilde{x}^k+By+Cz^k-b\|^2 \mid y \in \mathscr{Y}\}, \\
\tilde{z}^k \in \arg\min\{\theta_3(z) - z^T C^T \lambda^k + \frac{\beta}{2}\|A\tilde{x}^k+B\tilde{y}^k+Cz-b\|^2 \mid z \in \mathscr{Z}\}.
\end{cases} \tag{4.7}
$$

Then, according to Lemma 2.3, we have $\tilde{u}^k \in \mathscr{U}$ and

$$
\left\{
\begin{array}{l}
\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{ -A^T \lambda^k \\
\qquad\qquad\qquad + \beta A^T (A\tilde{x}^k + By^k + Cz^k - b) \} \geq 0, \quad \forall x \in \mathscr{X}, \\
\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{ -B^T \lambda^k \\
\qquad\qquad\qquad + \beta B^T (A\tilde{x}^k + B\tilde{y}^k + Cz^k - b) \} \geq 0, \quad \forall y \in \mathscr{Y}, \\
\theta_3(z) - \theta_3(\tilde{z}^k) + (z - \tilde{z}^k)^T \{ -C^T \lambda^k \\
\qquad\qquad\qquad + \beta C^T (A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b) \} \geq 0, \quad \forall z \in \mathscr{Z}.
\end{array}
\right.
\tag{4.8}
$$

Defining

$$
\tilde{\lambda}^k = \lambda^k - \beta \left(A\tilde{x}^k + By^k + Cz^k - b \right),
\tag{4.9}
$$

and using the VI form (4.3), we have $\tilde{w}^k \in \Omega$ and

$$
\left\{
\begin{array}{l}
\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{ \underline{-A^T \tilde{\lambda}^k} \} \geq 0, \quad \forall x \in \mathscr{X}, \\
\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{ \underline{-B^T \tilde{\lambda}^k} + \beta B^T B(\tilde{y}^k - y^k) \} \geq 0, \quad \forall y \in \mathscr{Y}, \\
\theta_3(z) - \theta_3(\tilde{z}^k) + (z - \tilde{z}^k)^T \left\{ \begin{array}{l} \underline{-C^T \tilde{\lambda}^k} + \beta C^T B(\tilde{y}^k - y^k) \\ \beta C^T C(\tilde{z}^k - z^k) \end{array} \right\} \geq 0, \quad \forall z \in \mathscr{Z}, \\
(\lambda - \tilde{\lambda}^k)^T \left\{ \begin{array}{l} \underline{(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b)} \\ -B(\tilde{y}^k - y^k) - C(\tilde{z}^k - z^k) + \dfrac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) \end{array} \right\} \geq 0, \quad \forall \lambda \in \Lambda.
\end{array}
\right.
\tag{4.10}
$$

The sum of the underline parts of (4.10) is exactly $F(\tilde{w}^k)$, where $F(\cdot)$ is defined in (4.3). Thus, we have

$$
\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega,
\tag{4.11}
$$

where the prediction matrix is

$$
Q = \begin{pmatrix} \beta B^T B & 0 & 0 \\ \beta C^T B & \beta C^T C & 0 \\ -B & -C & \dfrac{1}{\beta} I_m \end{pmatrix}.
\tag{4.12}
$$

Moreover, for the prediction matrix Q in (4.12) which is determined by the coarse splitting scheme (4.5), we have

$$
Q^T + Q = \begin{pmatrix} 2\beta B^T B & \beta B^T C & -B^T \\ \beta C^T B & 2\beta C^T C & -C^T \\ -B & -C & \frac{2}{\beta} I_m \end{pmatrix},
\tag{4.13}
$$

which is positive definite whenever B and C are full column rank.

4.3. Constructing the correction matrix M.

With the prediction matrix Q given in (4.12), the prediction-correction framework (2.5)-(2.6) can be specified as a concrete algorithm for the model (4.1) once the correction step (2.6) is specified. Now, we showcase how to specify the correction step (2.6) by the construction strategies discussed in Sections 3.1, 3.2 and 3.3. Note that $v = (y, z, \lambda)$ below.

4.3.1. *Construction 1.* Based on (4.12) and (4.13), and following the strategy in Section 3.1, we can choose

$$D = \begin{pmatrix} \nu\beta B^T B & 0 & 0 \\ 0 & \nu\beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} \tag{4.14}$$

with $0 < \nu < 1$, which is positive definite whenever B and C are both full column rank. Recall the correction matrix M in (3.3). Then, a concrete splitting contraction algorithm for (4.1) can be generated as below.

Algorithm 1 for the model (4.1)

[Prediction Step.] Obtain $(\tilde{x}^k, \tilde{y}^k, \tilde{z}^k)$ via the direct extension of the ADMM (4.6) and define $\tilde{\lambda}^k$ by (4.9).

[Correction Step.] $Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k)$.

For the correction step $Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k)$, we know that

$$Q^T = \begin{pmatrix} \beta B^T B & \beta B^T C & -B^T \\ 0 & \beta C^T C & -C^T \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} = \begin{pmatrix} \beta B^T & 0 & 0 \\ 0 & \beta C^T & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} B & C & -\frac{1}{\beta}I_m \\ 0 & C & -\frac{1}{\beta}I_m \\ 0 & 0 & I_m \end{pmatrix},$$

and

$$D = \begin{pmatrix} \nu\beta B^T B & 0 & 0 \\ 0 & \nu\beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} = \begin{pmatrix} \beta B^T & 0 & 0 \\ 0 & \beta C^T & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} \nu B & 0 & 0 \\ 0 & \nu C & 0 \\ 0 & 0 & I_m \end{pmatrix}.$$

That is, Q^T and D have a common matrix in their factorization forms above. Hence, to implement the correction step (2.6), i.e., $Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k)$, essentially we only need to consider the even easier equation

$$\begin{pmatrix} B & C & -\frac{1}{\beta}I_m \\ 0 & C & -\frac{1}{\beta}I_m \\ 0 & 0 & I_m \end{pmatrix} \begin{pmatrix} y^{k+1} - y^k \\ z^{k+1} - z^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} = \begin{pmatrix} \nu B & 0 & 0 \\ 0 & \nu C & 0 \\ 0 & 0 & I_m \end{pmatrix} \begin{pmatrix} \tilde{y}^k - y^k \\ \tilde{z}^k - z^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix}. \tag{4.16}$$

Moreover, iterations of the specified Algorithm 1 can be executed in terms of (By^k, Cz^k, λ^k) because it is sufficient to keep By^k and Cz^k, rather than y^k and z^k, to execute the $(k+1)$-th iteration. The variables y^k and z^k need to be solved once only at the last iteration. Hence, with the choice of D in (4.14), implementing the resulting correction step (2.6) essentially only requires solving the equation (4.16) in terms of (By^k, Cz^k, λ^k), which is extremely easy.

4.3.2. *Construction 2.* Based on (4.12) and (4.13), and following the strategy in Section 3.2, we can choose

$$G = \begin{pmatrix} \nu\beta B^T B & 0 & 0 \\ 0 & \nu\beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix}, \tag{4.17}$$

with $\nu \in (0,1)$, which can be guaranteed to be positive definite whenever B and C are full column rank. Note that the matrix G in (4.17) is precisely the matrix D defined in (4.14).

Furthermore, we have

$$
\Delta = Q^T + Q - G = \begin{pmatrix} (2-v)\beta B^T B & \beta B^T C & -B^T \\ \beta C^T B & (2-v)\beta C^T C & -C^T \\ -B & -C & \frac{1}{\beta}I_m \end{pmatrix}.
$$

Recall the correction matrix M in (3.10). Then, another contraction splitting algorithm for (4.1) can be generated as below.

Algorithm 2 for the model (4.1)

[Prediction Step.] Obtain $(\tilde{x}^k, \tilde{y}^k, \tilde{z}^k)$ via the direct extension of the ADMM (4.6) and define $\tilde{\lambda}^k$ by (4.9).

[Correction Step.] $Q^T(v^{k+1} - v^k) = \Delta(\tilde{v}^k - v^k)$.

For the correction step $Q^T(v^{k+1} - v^k) = \Delta(\tilde{v}^k - v^k)$, we know that

$$
Q^T = \begin{pmatrix} \beta B^T B & \beta B^T C & -B^T \\ 0 & \beta C^T C & -C^T \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} = \begin{pmatrix} \beta B^T & 0 & 0 \\ 0 & \beta C^T & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} B & C & -\frac{1}{\beta}I_m \\ 0 & C & -\frac{1}{\beta}I_m \\ 0 & 0 & I_m \end{pmatrix},
$$

and

$$
\begin{aligned}
\Delta &= \begin{pmatrix} (2-v)\beta B^T B & \beta B^T C & -B^T \\ \beta C^T B & (2-v)\beta C^T C & -C^T \\ -B & -C & \frac{1}{\beta}I_m \end{pmatrix} \\
&= \begin{pmatrix} \beta B^T & 0 & 0 \\ 0 & \beta C^T & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} (2-v)B & C & -\frac{1}{\beta}I_m \\ B & (2-v)C & -\frac{1}{\beta}I_m \\ -\beta B & -\beta C & I_m \end{pmatrix}.
\end{aligned}
$$

That is, Q^T and Δ have a common matrix in their factorization forms above. Hence, to implement the correction step (2.6), i.e., $Q^T(v^{k+1} - v^k) = \Delta(\tilde{v}^k - v^k)$, essentially we only need to consider the even easier equation

$$
\begin{pmatrix} B & C & -\frac{1}{\beta}I_m \\ 0 & C & -\frac{1}{\beta}I_m \\ 0 & 0 & I_m \end{pmatrix} \begin{pmatrix} y^{k+1} - y^k \\ z^{k+1} - z^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} = \begin{pmatrix} (2-v)B & C & -\frac{1}{\beta}I_m \\ B & (2-v)C & -\frac{1}{\beta}I_m \\ -\beta B & -\beta C & I_m \end{pmatrix} \begin{pmatrix} \tilde{y}^k - y^k \\ \tilde{z}^k - z^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix}.
$$

(4.19)

Similar as (4.16), with the choice of G in (4.17), implementing the resulting correction step (2.6) essentially only requires solving the equation (4.19) in terms of (By^k, Cz^k, λ^k), which is extremely easy.

4.3.3. *Construction 3*. Recall the relationship between the matrices D and G in (3.11), and $Q^T + Q$ given in (4.13). Essentially, the proposed construction strategies in Sections 3.1 and 3.2 take the same matrix

$$
\begin{pmatrix} v\beta B^T B & 0 & 0 \\ 0 & v\beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix}
$$

as D and G, respectively, and then the other one is determined by (3.11). As mentioned in Section 3.3, any other choice of D and G subject to the relationship (3.11) is also eligible. Let us consider the following specific one:

$$D = G = \frac{1}{2}[Q^T + Q] = \begin{pmatrix} \beta B^T B & \frac{1}{2}\beta B^T C & -\frac{1}{2}B^T \\ \frac{1}{2}\beta C^T B & \beta C^T C & -\frac{1}{2}C^T \\ -\frac{1}{2}B & \frac{1}{2} - C & \frac{1}{\beta}I_m \end{pmatrix}, \qquad (4.20)$$

which are both positive definite whenever B and C are full column rank. Recall the correction matrix M in (3.10). Then, one more contraction splitting algorithm for (4.1) can be generated as below.

Algorithm 3 for the model (4.1)

[Prediction Step.] Obtain $(\tilde{x}^k, \tilde{y}^k, \tilde{z}^k)$ via the direct extension of the ADMM (4.6) and
  define $\tilde{\lambda}^k$ by (4.9).

[Correction Step.] $Q^T(v^{k+1} - v^k) = \frac{1}{2}[Q^T + Q](\tilde{v}^k - v^k)$.

For the correction step $Q^T(v^{k+1} - v^k) = \frac{1}{2}[Q^T + Q](\tilde{v}^k - v^k)$, we know that

$$Q^T = \begin{pmatrix} \beta B^T B & \beta B^T C & -B^T \\ 0 & \beta C^T C & -C^T \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} = \begin{pmatrix} \beta B^T & 0 & 0 \\ 0 & \beta C^T & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} B & C & -\frac{1}{\beta}I_m \\ 0 & C & -\frac{1}{\beta}I_m \\ 0 & 0 & I_m \end{pmatrix},$$

and

$$\frac{1}{2}[Q^T + Q] = \begin{pmatrix} \beta B^T B & \frac{1}{2}\beta B^T C & -\frac{1}{2}B^T \\ \frac{1}{2}\beta C^T B & \beta C^T C & -\frac{1}{2}C^T \\ -\frac{1}{2}B & \frac{1}{2} - C & \frac{1}{\beta}I_m \end{pmatrix} = \begin{pmatrix} \beta B^T & 0 & 0 \\ 0 & \beta C^T & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} B & \frac{1}{2}C & -\frac{1}{2\beta}I_m \\ \frac{1}{2}B & C & -\frac{1}{2\beta}I_m \\ -\frac{1}{2}\beta B & -\frac{1}{2}\beta C & I_m \end{pmatrix}.$$

That is, Q^T and $\frac{1}{2}[Q^T + Q]$ have a common matrix in their factorization forms above. Hence, to implement the correction step (2.6), i.e., $Q^T(v^{k+1} - v^k) = \frac{1}{2}[Q^T + Q](\tilde{v}^k - v^k)$, essentially we only need to consider the even easier equation

$$\begin{pmatrix} B & C & -\frac{1}{\beta}I_m \\ 0 & C & -\frac{1}{\beta}I_m \\ 0 & 0 & I_m \end{pmatrix} \begin{pmatrix} y^{k+1} - y^k \\ z^{k+1} - z^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} = \begin{pmatrix} B & \frac{1}{2}C & -\frac{1}{2\beta}I_m \\ \frac{1}{2}B & C & -\frac{1}{2\beta}I_m \\ -\frac{1}{2}\beta B & -\frac{1}{2}\beta C & I_m \end{pmatrix} \begin{pmatrix} \tilde{y}^k - y^k \\ \tilde{z}^k - z^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix}. \quad (4.22)$$

Similar as (4.16) and (4.19), implementing the resulting correction step (2.6) essentially only requires solving the equation (4.22) in terms of (By^k, Cz^k, λ^k), which is extremely easy.

4.3.4. *Summary.* We have discussed three concrete strategies for constructing ADMM-based splitting contraction algorithms for the model (4.1), with the same prediction step determined by the coarse splitting scheme (4.6) with (4.9). These three strategies differ in how to choose the matrices D and G subject to (3.11); accordingly specifications of the correction step (2.6) are different. It is easy to verify that Algorithm 1 corresponds to the algorithm proposed in [16]. Certainly, any other choice of D and G in accordance with (3.11) leads to another splitting contraction algorithm for the model (4.1) whose prediction step remains unchanged as that in Algorithms 1-3, while there are infinitely many such choices.

5. APPLICATION TO MULTIPLE BLOCK SEPARABLE CONVEX OPTIMIZATION

In this section, we extend the analysis in Section 4 to more general and complicated convex programming problems. We recall a recent work of ours [18], and will show that more new algorithms with similar advantages as those in [18] can be presented by following the proposed constructions strategies in Section 3.

5.1. Model.
We consider the generic multiple block convex programming problem with both linear equality and inequality constraints:

$$\min\Big\{\sum_{i=1}^{p}\theta_i(x_i) \mid \sum_{i=1}^{p}A_ix_i = b \text{ (or } \geq b), \ x_i \in \mathscr{X}_i\Big\}, \tag{5.1}$$

where $\theta_i : \mathfrak{R}^{n_i} \to \mathfrak{R}$ $(i = 1,\ldots,p)$ are closed, proper, and convex but not necessarily smooth functions; $\mathscr{X}_i \subseteq \mathfrak{R}^{n_i}$ $(i = 1,\ldots,p)$ are closed convex sets; $A_i \in \mathfrak{R}^{m\times n_i}$ $(i = 1,\ldots,p)$ are given matrices; $b \in \mathfrak{R}^m$; $\sum_{i=1}^{p} n_i = n$; and $p \geq 3$. Note that the more general model (5.1) differs from the special three-block separable convex programming problem (4.1) in that $p \geq 3$ and that both linear equality and inequalities are considered.

Let the Lagrangian function of (5.1) be

$$L(x_1,\ldots,x_p,\lambda) = \sum_{i=1}^{p}\theta_i(x_i) - \lambda^T\Big(\sum_{i=1}^{p}A_ix_i - b\Big), \tag{5.2}$$

with λ the Lagrange multiplier. We know that $\lambda \in \mathfrak{R}^m$ or $\lambda \in \mathfrak{R}^m_+$ if $\sum_{i=1}^{p}A_ix_i = b$ or $\sum_{i=1}^{p}A_ix_i \geq b$ is considered in the model (5.1), respectively. The optimality condition of (5.1) can be written as the VI (2.3)-(2.4) with the following specifications:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall w \in \Omega, \tag{5.3a}$$

where

$$w = \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \lambda \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}, \quad \theta(x) = \sum_{i=1}^{p}\theta_i(x_i), \quad F(w) = \begin{pmatrix} -A_1^T\lambda \\ \vdots \\ -A_p^T\lambda \\ \sum_{i=1}^{p}A_ix_i - b \end{pmatrix}, \tag{5.3b}$$

and

$$\Omega = \prod_{i=1}^{p}\mathscr{X}_i \times \Lambda \quad \text{with} \quad \Lambda = \begin{cases} \mathfrak{R}^m, & \text{if } \sum_{i=1}^{p}A_ix_i = b, \\ \mathfrak{R}^m_+, & \text{if } \sum_{i=1}^{p}A_ix_i \geq b. \end{cases}$$

5.2. Discerning the prediction matrix Q.
To design an algorithm for (5.1), it is natural to consider extending the ADMM (2.15) and splitting the augmented Lagrangian function $L(x_1,\ldots,x_p,\lambda)$ in (5.2) to obtain easier subproblems. In [18], two different ways were suggested and the resulting subproblems were used as prediction steps for the prediction-correction framework (2.5)-(2.6) . We recall these results; accordingly the prediction matrix Q becomes clear.

5.2.1. *Primal-dual order.* We first consider splitting the augmented Lagrangian function $L(x_1,\ldots,x_p,\lambda)$ in the primal-dual order, and thus obtain the following scheme similar as (4.5):

$$
\begin{cases}
\tilde{x}_1^k \in \arg\min\left\{\theta_1(x_1) - x_1^T A_1^T \lambda^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\right\}; \\[2mm]
\tilde{x}_2^k \in \arg\min\left\{\theta_2(x_2) - x_2^T A_2^T \lambda^k + \frac{\beta}{2}\|A_1(\tilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\right\}; \\[2mm]
\qquad\vdots \\[1mm]
\tilde{x}_i^k \in \arg\min\left\{\theta_i(x_i) - x_i^T A_i^T \lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2 \mid x_i \in \mathcal{X}_i\right\}; \\[2mm]
\qquad\vdots \\[1mm]
\tilde{x}_p^k \in \arg\min\left\{\theta_p(x_p) - x_p^T A_p^T \lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{p-1} A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2 \mid x_p \in \mathcal{X}_p\right\}; \\[2mm]
\tilde{\lambda}^k = \arg\max\left\{-\lambda^T\left(\sum_{j=1}^{p} A_j\tilde{x}_j^k - b\right) - \frac{1}{2\beta}\|\lambda - \lambda^k\|^2 \mid \lambda \in \Lambda\right\}.
\end{cases}
$$

$$(5.4)$$

Note that linear inequalities are considered in the model (5.1). Hence, updating the Lagrange multiplier λ in (5.4) involves the constraint Λ. For the same reason in [3], the coarse splitting scheme (5.4) has no guaranteed convergence, but it can be used as a starting point to design a splitting contraction algorithm with provable convergence. As analyzed in [18], the coarse splitting scheme (5.4) can be rewritten as

$$\tilde{w}^k \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (w - \tilde{w}^k)^T Q_{PD}(w^k - \tilde{w}^k), \quad \forall w \in \Omega, \quad (5.5a)$$

where

$$
Q_{PD} = \begin{pmatrix}
\beta A_1^T A_1 & 0 & \cdots & 0 & A_1^T \\[2mm]
\beta A_2^T A_1 & \beta A_2^T A_2 & \ddots & \vdots & A_2^T \\[2mm]
\vdots & & \ddots & 0 & \vdots \\[2mm]
\beta A_p^T A_1 & \beta A_p^T A_2 & \cdots & \beta A_p^T A_p & A_p^T \\[2mm]
0 & 0 & \cdots & 0 & \frac{1}{\beta}I_m
\end{pmatrix}, \quad (5.5b)
$$

which is in form of the prediction step (2.5) with the prediction matrix Q_{PD}. It is easy to see that the matrix $Q_{PD}^T + Q_{PD}$ is positive definite if all A_i's are full column rank.

To simplify the notation, let us further denote

$$
P = \begin{pmatrix}
\sqrt{\beta}A_1 & 0 & \cdots & \cdots & 0 \\[2mm]
0 & \sqrt{\beta}A_2 & \ddots & & \vdots \\[2mm]
\vdots & \ddots & \ddots & \ddots & \vdots \\[2mm]
\vdots & & \ddots & \sqrt{\beta}A_p & 0 \\[2mm]
0 & \cdots & \cdots & 0 & \frac{1}{\sqrt{\beta}}I_m
\end{pmatrix}, \quad
\xi = Pw = \begin{pmatrix}
\sqrt{\beta}A_1 x_1 \\[2mm]
\sqrt{\beta}A_2 x_2 \\[2mm]
\vdots \\[2mm]
\sqrt{\beta}A_p x_p \\[2mm]
\frac{1}{\sqrt{\beta}}\lambda
\end{pmatrix}; \quad (5.6)
$$

and also

$$\Xi = \left\{\xi \mid \xi = Pw, \, w \in \Omega\right\} \quad \text{and} \quad \Xi^* = \left\{\xi^* \mid \xi^* = Pw^*, \, w^* \in \Omega^*\right\}.$$

Then, with P and ξ in (5.6), we can rewrite the VI (5.5a)-(5.5b) as

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (\xi - \tilde{\xi}^k)^T \mathcal{Q}_{PD}(\xi^k - \tilde{\xi}^k), \quad \forall w \in \Omega, \quad (5.7)$$

where

$$Q_{PD} = P^T \mathcal{Q}_{PD} P \quad \text{with} \quad \mathcal{Q}_{PD} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & I_m \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{5.8}$$

5.2.2. *Dual-primal order.* In (5.4), the primal variables are solved first, followed by the dual variable. We can alternatively update the dual variable first. That is, we can consider the following scheme which updates the variables in the dual-primal order:

$$\begin{cases} \tilde{\lambda}^k = \arg\max\left\{ -\lambda^T \left(\sum_{j=1}^p A_j x_j^k - b \right) - \frac{1}{2\beta} \|\lambda - \lambda^k\|^2 \mid \lambda \in \Lambda \right\}; \\ \tilde{x}_1^k \in \arg\min\left\{ \theta_1(x_1) - x_1^T A_1^T \tilde{\lambda}^k + \frac{\beta}{2} \|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathscr{X}_1 \right\}; \\ \tilde{x}_2^k \in \arg\min\left\{ \theta_2(x_2) - x_2^T A_2^T \tilde{\lambda}^k + \frac{\beta}{2} \|A_1(\tilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathscr{X}_2 \right\}; \\ \qquad \vdots \\ \tilde{x}_i^k \in \arg\min\left\{ \theta_i(x_i) - x_i^T A_i^T \tilde{\lambda}^k + \frac{\beta}{2} \|\sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2 \mid x_i \in \mathscr{X}_i \right\}; \\ \qquad \vdots \\ \tilde{x}_p^k \in \arg\min\left\{ \theta_p(x_p) - x_p^T A_p^T \tilde{\lambda}^k + \frac{\beta}{2} \|\sum_{j=1}^{p-1} A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2 \mid x_p \in \mathscr{X}_p \right\}. \end{cases} \tag{5.9}$$

Similarly as (5.4), the coarse splitting scheme (5.9) has no guaranteed convergence, but it can be used as a starting point to design a splitting contraction algorithm with provable convergece. As analyzed in [18], the scheme (5.9) can be rewritten as

$$\tilde{w}^k \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (w - \tilde{w}^k)^T Q_{DP}(w^k - \tilde{w}^k), \ \ \forall w \in \Omega, \tag{5.10a}$$

where

$$Q_{DP} = \begin{pmatrix} \beta A_1^T A_1 & 0 & \cdots & 0 & 0 \\ \beta A_2^T A_1 & \beta A_2^T A_2 & \ddots & \vdots & 0 \\ \vdots & & \ddots & 0 & \vdots \\ \beta A_p^T A_1 & \beta A_p^T A_2 & \cdots & \beta A_p^T A_p & 0 \\ -A_1 & -A_2 & \cdots & -A_p & \frac{1}{\beta} I_m \end{pmatrix}. \tag{5.10b}$$

It is easy to see that the matrix $Q_{DP}^T + Q_{DP}$ is positive definite if all A_i's are full column rank.

Also, using P and ξ in (5.6), we can rewrite the VI (5.10a)-(5.10b) as

$$\tilde{w}^k \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (\xi - \tilde{\xi}^k)^T \mathcal{Q}_{DP}(\xi^k - \tilde{\xi}^k), \quad \forall w \in \Omega. \tag{5.11}$$

where

$$Q_{DP} = P^T \mathcal{Q}_{DP} P \quad \text{with} \quad \mathcal{Q}_{DP} = \begin{pmatrix} I_m & 0 & \cdots & 0 & 0 \\ I_m & I_m & \ddots & \vdots & 0 \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & 0 \\ -I_m & -I_m & \cdots & -I_m & I_m \end{pmatrix}. \tag{5.12}$$

5.3. Representation of the prediction-correction framework (2.5)-(2.6) and convergence conditions (2.7).

To show splitting contraction algorithms for the model (5.1) more clearly, we can rewrite the prediction-correction framework (2.5)-(2.6) and convergence conditions (2.7) in the context of the VI (5.3a)-(5.3b) with the notation P and ξ in (5.6).

Prediction-correction framework for the VI (5.3a)-(5.3b)

[Prediction Step.] With given $\xi^k = Pw^k$, find $\tilde{w}^k \in \Omega$ such that

$$\tilde{w}^k \in \Omega, \ \ \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (\xi - \tilde{\xi}^k)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k), \ \ \forall w \in \Omega, \tag{5.13a}$$

where the matrix $\mathcal{Q} \in \mathfrak{R}^{(p+1)m \times (p+1)m}$ is not necessarily symmetric but the matrix $\mathcal{Q}^T + \mathcal{Q}$ is assumed to be positive definite.

[Correction Step.] Find a nonsingular matrix $\mathcal{M}$ and update ξ by

$$\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k). \tag{5.13b}$$

Convergence conditions

For the matrices Q and M used in (5.13a) and (5.13b), respectively, there exists a matrix $\mathcal{H} \succ 0$ such that

$$\mathcal{H}\mathcal{M} = \mathcal{Q}, \tag{5.14a}$$

and

$$\mathcal{G} := \mathcal{Q}^T + \mathcal{Q} - \mathcal{M}^T \mathcal{H} \mathcal{M} \succ 0. \tag{5.14b}$$

Obviously, the VIs (5.7) and (5.11) are specific cases of (5.13a), whose corresponding prediction matrices are $\mathcal{Q}_{PD}$ in (5.8) and $\mathcal{Q}_{DP}$ in (5.12), respectively.

5.4. Specifications of splitting contraction algorithms.

Now, we focus on specifying the correction matrix $\mathcal{M}$ in (5.13b) with $\mathcal{Q}_{PD}$ in (5.8) and $\mathcal{Q}_{DP}$ in (5.12). We will show that the algorithms in [18] can be recovered, while more new algorithms can be designed easily by our proposed construction strategies. With $\mathcal{Q}_{PD}$ and $\mathcal{Q}_{DP}$, their associated correction matrices are denoted by $\mathcal{M}_{PD}$ and $\mathcal{M}_{DP}$, and $\mathcal{G}_{PD}$ and $\mathcal{G}_{DP}$, respectively.

5.4.1. *Some matrices.*

To further simplify the notation to be used, we define the following $p \times p$ block matrices:

$$\mathcal{L} = \begin{pmatrix} I_m & 0 & \cdots & 0 \\ I_m & I_m & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ I_m & I_m & \cdots & I_m \end{pmatrix} \quad \text{and} \quad \mathcal{I} = \begin{pmatrix} I_m & 0 & \cdots & 0 \\ 0 & I_m & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{5.15}$$

We also define the $1 \times p$ block matrix

$$\mathscr{E} = \begin{pmatrix} I_m & I_m & \cdots & I_m \end{pmatrix}. \tag{5.16}$$

It is cleat that

$$\mathscr{L}^T + \mathscr{L} = \mathscr{I} + \mathscr{E}^T \mathscr{E}. \tag{5.17}$$

Furthermore, the matrix $\mathscr{Q}_{PD}$ in (5.8) has the form

$$\mathscr{Q}_{PD} = \begin{pmatrix} \mathscr{L} & \mathscr{E}^T \\ 0 & I_m \end{pmatrix} \qquad \text{and thus} \qquad \mathscr{Q}_{PD}^T + \mathscr{Q}_{PD} = \begin{pmatrix} \mathscr{I} + \mathscr{E}^T \mathscr{E} & \mathscr{E}^T \\ \mathscr{E} & 2I_m \end{pmatrix}.$$

Similarly, the matrix $\mathscr{Q}_{DP}$ in (5.12) has the form

$$\mathscr{Q}_{DP} = \begin{pmatrix} \mathscr{L} & 0 \\ -\mathscr{E} & I_m \end{pmatrix} \qquad \text{and thus} \qquad \mathscr{Q}_{DP}^T + \mathscr{Q}_{DP} = \begin{pmatrix} \mathscr{I} + \mathscr{E}^T \mathscr{E} & -\mathscr{E}^T \\ -\mathscr{E} & 2I_m \end{pmatrix}.$$

To further analyze the correction steps associated with the correction matrices $\mathscr{M}_{PD}$ and $\mathscr{M}_{DP}$, let us take a closer look at the matrices $\mathscr{Q}_{PD}^{-T}$ and $\mathscr{Q}_{DP}^{-T}$. Indeed, we have

$$\mathscr{Q}_{PD}^{-T} = \begin{pmatrix} \mathscr{L}^T & 0 \\ \mathscr{E} & I_m \end{pmatrix}^{-1} = \begin{pmatrix} \mathscr{L}^{-T} & 0 \\ -\mathscr{E}\mathscr{L}^{-T} & I_m \end{pmatrix}. \tag{5.18}$$

and

$$\mathscr{Q}_{DP}^{-T} = \begin{pmatrix} \mathscr{L}^T & -\mathscr{E}^T \\ 0 & I_m \end{pmatrix}^{-1} = \begin{pmatrix} \mathscr{L}^{-T} & \mathscr{L}^{-T}\mathscr{E}^T \\ 0 & I_m \end{pmatrix}. \tag{5.19}$$

Recall the respective definitions $\mathscr{L}$ and $\mathscr{E}$ in (5.15) and (5.16). We have

$$\mathscr{L}^{-T} = \begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix} \tag{5.20}$$

and

$$\mathscr{E}\mathscr{L}^{-T} = \begin{pmatrix} I_m & 0 & \cdots & 0 \end{pmatrix} \quad \text{and} \quad \mathscr{L}^{-T}\mathscr{E}^T = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ I_m \end{pmatrix}. \tag{5.21}$$

Hence, the matrices $\mathscr{Q}_{PD}^{-T}$ in (5.18) and $\mathscr{Q}_{DP}^{-T}$ in (5.19) are both very simple in structure; their entries only consist of blocks of I_m, $-I_m$ and 0.

5.4.2. $\mathcal{M}_{PD}$ *for the primal-dual prediction (5.4).* With (5.4) as the prediction step, the prediction matrix $\mathcal{Q}_{PD}$ is given in (5.8). To construct the corresponding correction matrix $\mathcal{M}_{PD}$, for example, we can choose

$$\mathcal{D}_{PD} = \begin{pmatrix} \nu \mathcal{I} & 0 \\ 0 & I_m \end{pmatrix}, \tag{5.22}$$

with any $\nu \in (0,1)$. Recall (3.11). Thus, (5.22) also means

$$\mathcal{G}_{PD} := (\mathcal{Q}_{PD}^T + \mathcal{Q}_{PD}) - \mathcal{D}_{PD} = \begin{pmatrix} (1-\nu)\mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & I_m \end{pmatrix}.$$

It is clear that both the matrices $\mathcal{D}_{PD}$ and $\mathcal{G}_{PD}$ are positive definite. According to (3.3), (5.18) and (5.22), the correction matrix $\mathcal{M}_{PD}$ can be constructed as

$$\mathcal{M}_{PD} = \mathcal{Q}_{PD}^{-T}\mathcal{D}_{PD} = \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ -\mathcal{E}\mathcal{L}^{-T} & I_m \end{pmatrix} \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} = \begin{pmatrix} \nu\mathcal{L}^{-T} & 0 \\ -\nu\mathcal{E}\mathcal{L}^{-T} & I_m \end{pmatrix}. \tag{5.23}$$

This coincides with the correction step in Section 7 of [18]. Recall (5.20) and (5.21). We know that the correction step (5.13b) with the correction matrix $\mathcal{M}_{PD}$ defined in (5.23) is extremely easy to be implemented. Hence, the implementation of the specified splitting contraction algorithm mainly needs to solve the splitting x_i-subproblems in (5.4).

5.4.3. $\mathcal{M}_{DP}$ *for the dual-primal prediction (5.9).* With (5.9) as the prediction step, the prediction matrix $\mathcal{Q}_{DP}$ is given in (5.12). To construct the correction matrix $\mathcal{M}_{DP}$, for example, we can choose

$$\mathcal{D}_{DP} = \begin{pmatrix} \nu\mathcal{I} + \mathcal{E}^T\mathcal{E} & -\mathcal{E}^T \\ -\mathcal{E} & I_m \end{pmatrix} \tag{5.24}$$

with $\nu \in (0,1)$. Recall (3.11). Thus, (5.24) also means

$$\mathcal{G}_{DP} := (\mathcal{Q}_{DP}^T + \mathcal{Q}_{DP}) - \mathcal{D}_{DP} = \begin{pmatrix} \nu\mathcal{E} & 0 \\ 0 & I_m \end{pmatrix}.$$

It is clear that both the matrices $\mathcal{D}_{DP}$ and $\mathcal{G}_{DP}$ are positive definite. According to (3.3), (5.19) and (5.24), the correction matrix $\mathcal{M}_{DP}$ can be constructed by

$$\mathcal{M}_{DP} = \mathcal{Q}_{DP}^{-T}\mathcal{D}_{DP} = \begin{pmatrix} \mathcal{L}^{-T} & \mathcal{L}^{-T}\mathcal{E}^T \\ 0 & I_m \end{pmatrix} \begin{pmatrix} \nu\mathcal{I} + \mathcal{E}^T\mathcal{E} & -\mathcal{E}^T \\ -\mathcal{E} & I_m \end{pmatrix} = \begin{pmatrix} \nu\mathcal{L}^{-T} & 0 \\ -\mathcal{E} & I_m \end{pmatrix}. \tag{5.25}$$

This coincides with the correction step in Section 8 of [18]. Recall (5.20) and (5.21). We know that the correction step (5.13b) with the correction matrix $\mathcal{M}_{DP}$ defined in (5.25) is also extremely easy to be implemented. Hence, the implementation of the specified splitting contraction algorithm mainly needs to solve the splitting x_i-subproblems in (5.9).

5.4.4. *More choices.* To specify the prediction-correction framework (5.13) and ensure the convergence conditions (5.14) with a given $\mathscr{Q}$ satisfying $\mathscr{Q}^T + \mathscr{Q} \succ 0$, like (3.11), the matrices $\mathscr{D}$ and $\mathscr{G}$ can be chosen with the only restriction

$$\mathscr{D} \succ 0, \quad \mathscr{G} \succ 0, \quad \text{and} \quad \mathscr{D} + \mathscr{G} = \mathscr{Q}^T + \mathscr{Q}. \tag{5.26}$$

Hence, there are infinitely many ways to construct $\mathscr{H}$ and $\mathscr{M}$ with the given prediction matrix $\mathscr{Q}_{PD}$ in (5.8) or $\mathscr{Q}_{DP}$ in (5.12). For instance, we can choose $\mathscr{G}_{PD}$ and $\mathscr{G}_{DP}$, instead of $\mathscr{D}_{PD}$ and $\mathscr{D}_{DP}$, as the matrices defined in (5.22) and (5.24), respectively; or we can choose

$$\mathscr{D} = \alpha\big[\mathscr{Q}^T + \mathscr{Q}\big] \text{ and } \mathscr{G} = (1-\alpha)\big[\mathscr{Q}^T + \mathscr{Q}\big], \ \alpha \in (0,1).$$

All these choices lead to new splitting contraction algorithms for the model (5.1) with provable convergence that are not covered in [18].

Finally, we would emphasize that, as shown in (5.18)-(5.21), $\mathscr{Q}_{PD}^{-T}$ and $\mathscr{Q}_{DP}^{-T}$ are both very simple in structure, and they do not cause too much additional computation for constructing the correction matrix via $\mathscr{M}_{PD} = \mathscr{Q}_{PD}^{-T}\mathscr{D}_{PD}$ (see (5.23)) or $\mathscr{M}_{DP} = \mathscr{Q}_{DP}^{-T}\mathscr{D}_{DP}$ (see (5.25)). This advantage makes it practical and adaptable to choose more application-tailored $\mathscr{D}_{PD}$ and $\mathscr{D}_{DP}$ for specific applications of the model (5.1).

6. Conclusions

We revisited a unified framework for algorithmic design and convergence analysis that can capture a series of our previous works of desiging/analyzing splitting contraction algorithms for separable convex programming problems, and provided some construction strategies to specify this unified framework. By the proposed strategies, once a matrix (e.g., D or G as mentioned) is chosen, a splitting contraction algorithm with provable convergence can be automatically generated. There are many specification strategies, and the flexibility of choosing such a matrix enables us to design model-tailored/application-tailored splitting contraction algorithms with easy subproblems conveniently. We illustrated how to apply this construction principle to generate easily implementable ADMM-based algorithms for separable convex programming models with linear constraints. The same methodology can be applied to improve some other algorithms with theoretical or numerical disadvantages for other optimization problems.

References

[1] A. Beck, First-Order Methods in Optimization, SIAM, Philadelphia, 2017.
[2] X. J. Cai, G .Y. Gu, B. S. He and X. M. Yuan, A proximal point algorithms revisit on the alternating direction method of multipliers, Sci. China Math. 56 (2013) 2179-2186.
[3] C. H. Chen, B. S. He, Y. Y. Ye and X. M. Yuan, The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent, Mathematical Programming, Series A 155 (2016) 55-79.
[4] E. Corman and X. M. Yuan, A generalized proximal point algorithm and its convergence rate, SIAM J. Optim. 24 (2014) 1614-1638.
[5] J. Douglas and H. H. Rachford, On the numerical solution of the heat conduction problem in 2 and 3 space variables, Trans. Am. Math. Soc. 82 (1956) 421-439.

[6] M. Fortin and R. Glowinski, Augmented Lagrangian methods: Applications to the Numerical Solution of Boundary-value Problems, Elsevier, Stud. Math. Appl. 15, North-Holland, Amsterdam, 1983.

[7] D. Gabay, Applications of the method of multipliers to variational inequalities. In: Fortin, M., Glowinski, R. (eds.) Augmented Lagrange Methods: Applications to the Solution of Boundary-valued Problems, pp. 299-331. North Holland, Amsterdam (1983).

[8] R. Glowinski, Numerical Methods for Nonlinear Variational Problems, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1984.

[9] R. Glowinski and A. Marrocco, A, Approximation par éléments finis d'ordre un et résolution par pénalisation-dualité d'une classe de problémes non linéaires. RAIRO Anal. Numer. R2 (1975) 41-76.

[10] R. Glowinski and P. Le Tallec, Augmented Lagrangian and Operator-splitting Methods in Nonlinear Mechanics, SIAM, Philadelphia, 1989.

[11] G. Y. Gu, B. S. He and X. M. Yuan, Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: a unified approach, Comput. Optim. Appl. 59 (2014) 135-161.

[12] B. S. He, From the projection and contraction methods for variatioonal inequality to the splittingcontraction methods fro convex optimization (in Chinese), Numerical Mathematics, A Journal of Chinese Universities, 38 (2016) 74-96.

[13] B. S. He, My 20 years research on alternating directions method of multipliers (in Chinese), Operations Research Trasactions, 22 (2018) 1-31.

[14] B. S. He, Using a unified framework to design the splitting and contraction methods for convex optimization (in Chinese), Numerical Mathematics, A Journal of Chinese Universities, 44 (2022), 1-35.

[15] B. S. He, H. Liu, Z. R. Wang and X. M. Yuan, A strictly Peaceman-Rachford splitting method for convex programming, SIAM J. Optim. 24 (2014) 1011-1040.

[16] B. S. He, M. Tao and X. M. Yuan, Alternating direction method with Gaussian back substitution for separable convex programming, SIAM J. Optim. 22 (2012) 313-340.

[17] B. S. He, M. Tao and X. M. Yuan, A splitting method for separable convex programming, IMA J. Numer. Anal. 31 (2015) 394-426.

[18] B. S. He, S. J. Xu and X. M. Yuan, Extensions of ADMM for separable convex optimization problems with linear equality or inequality constraints, arXiv:2107.01897v2[math.OC].

[19] B. S. He and X. M. Yuan, On the $O(1/n)$ convergence rate of the alternating direction method, SIAM J. Numer. Anal. 50 (2012) 700-709.

[20] B. S. He and X. M. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective, SIAM J. Imaging Sci. 5 (2012) 119-149.

[21] B. S. He and X. M. Yuan, On non-ergodic convergence rate of Douglas-Rachford alternating directions method of multipliers, Numerische Math. 130 (2015) 567-577.

[22] B. S. He and X. M. Yuan, A class of ADMM-based algorithms for three-block separable convex programming, Comput. Optim. Appl. 70 (2018), 791-826.

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

[24] B. Martinet, Regularisation, d'inéquations variationelles par approximations succesives. Rev. Francaise d'Inform. Recherche Oper. 4 (1970) 154–159.

[25] D. H. Peaceman and H. H. Rachford, The numerical solution of parabolic elliptic differential equations, J. Soc. Ind. Appl. Math. 3 (1955) 28-41.