Generalizations of the Proximal Method of Multipliers in Convex Optimization

R. Tyrrell Rockafellar[1]

Abstract

The proximal method of multipliers, originally introduced as a way of solving convex programming problems with inequality constraints, is a proximally stabilized alternative to the augmented Lagrangian method that is sometimes called the proximal augmented Lagrangian method. It has gained attention as a vehicle for deriving decomposition algorithms for wider formulations of problems in convex optimization than just convex programming. Here those themes are developed further. The basic algorithm is articulated in several seemingly different formats that are equivalent under exact computations, but diverge when minimization steps are executed only approximately. Stopping criteria are demonstrated to maintain convergence to a particular solution despite such approximations. Q-linear convergence is obtained from a metric regularity property of the Lagrangian mapping at the solution that acts as a mildly enhanced condition for local optimality on top of convexity and is generically available, in a sense.

Moreover, all this is brought about with the proximal terms allowed to vary in their underlying metric from one iteration to the next. That generalization enables the results to be translated to the theory of the progressive decoupling algorithm, significantly adding to its versatility and providing linear convergence guarantees in its broad applicability to techniques for problem decomposition.

Keywords: *convex optimization, augmented Lagrangians, multiplier methods, proximal ALM, progressive decoupling, method of partial inverses, proximal point algorithm, variable-metric prox terms, stopping criteria, linear convergence guarantees, metric subregularity, maximal monotone mappings, variational inequalities.*

Version of 3 July 2023

[1]University of Washington, Department of Mathematics, Box 354350, Seattle, WA 98195-4350;
E-mail: *rtr@uw.edu*, URL: sites.math.washington.edu/∼rtr/mypage.html

1 Introduction

Solution techniques in optimization often utilize dual variables along with primal variables. In the presence of convexity, the proximal point algorithm (PPA) for finding a zero of a monotone mapping [9] can have a role in that. For convex programming with inequality constraints, application of the PPA to solving the dual problem was revealed in [10] to produce a *method of multipliers* for solving the primal problem that extends the so-named procedure of Hestenes [3] and Powell [6] for solving nonlinear programming with equation constraints. Application of the PPA to finding a saddle point of the Lagrangian function produced instead in [10] a *proximal* method of multipliers.

Since then, terminology has shifted, and the method of multipliers has commonly come to be called the *augmented Lagrangian method* (ALM). Its uses in many areas of convex optimization are well known, but advances have also been made in understanding its potential in nonconvex optimization, recently for example in [16]. In contrast, the proximal method of multipliers (PMM), also now called the *proximal augmented Lagrangian method*, has received less attention. However, Shefi and Teboulle in [20] (2014) showed it can provide a unifying basis for deriving the decomposition algorithms in convex optimization that have gained prominence over the years. Further work on it has the potential to yield even more of interest in that direction.

This paper is devoted to broadening the scope of the proximal method of multipliers in applications to general finite-dimensional problems of convex optimization, while at the same time improving knowledge about its convergence properties. New features are introduced, such as variable-metric prox terms which add flexibility to the handling of primal and dual variables. Guarantees for global convergence despite inexact minimization are devised that propagate the original ones for classical convex programming in [10] into this much wider territory. The local condition that was utilized in [10] to establish linear convergence is replaced in this by a distinctly weaker condition of metric subregularity. Alternative algorithmic formats are developed that bypass explicit handling of augmented Lagrangian functions when convenient formulas for them aren't at hand. That has the benefit of leading here, through a connection with Spingarn's method of partial inverses [21], to first-time results on linear convergence of the *progressive decoupling algorithm* in [12], [13]. These results have major import for the schemes of problem decomposition in convex optimization that have been based on progressive decompling, such as elaborations of Spingarn's augmented Lagrangian decomposition scheme in [22], although such applications will not be taken up already in this paper.

Duality framework. Our general approach to augmented Lagrangians and the PMM differs from that of Shefi and Teboulle [20] and allied literature, at least on the surface. We adopt the fundamental framework for duality in convex optimization in which a minimization problem in a vector x is imbedded in a family of such problems parameterized by a vector u having reference value $u = 0$. Thus, for a closed proper convex function φ on $I\!R^n \times I\!R^m$, we seek to

$$\text{minimize } \ \varphi(x,u) \ \text{ subject to } \ u = 0. \tag{P}$$

The dual problem in terms of the conjugate convex function φ^* is to

$$\text{maximize } \ -\varphi^*(v,y) \ \text{ subject to } \ v = 0. \tag{D}$$

The Lagrangian function, defined by

$$l(x,y) := \inf_u \{ \varphi(x,u) - y{\cdot}u \}, \tag{1.1}$$

is convex in x and concave in y. It connects primal and dual by

$$\varphi(x,u) = \sup_y \{ l(x,y) + y{\cdot}u \}, \qquad -\varphi^*(v,y) = \inf_x \{ l(x,y) - v{\cdot}x \}. \tag{1.2}$$

Here $y{\cdot}u$ is the inner product associated with the canonical Euclidean norm, which will be denoted ty $|\cdot|$, in distinction to other inner products $\langle\cdot,\cdot\rangle$ and norms $||\cdot||$ that will have a role later.

This parameterized duality framework has a long history in convex analysis going back to [7] and [8]; see also [19, Section 11H]. Classical convex programming has

$$\varphi(x,u)=\begin{cases} f_0(x) \ \text{ if } x\in X \text{ and } \ f_i(x)+u_i \begin{cases} \leq 0 \ \text{ for } \ i=1,\dots,s, \\ =0 \ \text{ for } \ i=s+1,\dots,m, \end{cases} \\ \infty \ \text{ otherwise,} \end{cases} \tag{1.3}$$

where X is a nonempty closed convex set and f_i is a differentiable function on $I\!R^n$ that is convex for $i=0,1,\dots,s$ and affine for $i=s+1,\dots,m$. The case of *Fenchel duality* has

$$\begin{gathered}\varphi(x,u)=f(x)+g(Ax+u), \qquad \varphi^*(v,y)=f^*(v-A^*y)+g^*(y), \\ l(x,y)=f(x)-g^*(y)+y{\cdot}Ax,\end{gathered} \tag{1.4}$$

for closed proper convex functions f, g, and a linear mapping A with adjoint A^*.[2]

Shefi and Teboulle in [20], along with many other contributors to the theory of decomposition techniques in convex optimization, start from the Fenchel-type model of minimizing $f(x)+g(Ax)$ but recast it as minimizing $f(x)+g(z)$ subject to $Ax-z=0$. They then bring in a Lagrange multiplier vector y for that equation. This amounts, in the framework here, to amplifying the primal variable x to $\tilde{x}=(x,z)$ and shifting to a secondary formulation in the pattern of (1.4) with $\tilde{\varphi}(\tilde{x},u)=\tilde{f}(\tilde{x})+\tilde{g}(\tilde{A}\tilde{x}+u)$ and $\tilde{g}=\delta_0$, the indicator of the origin. The augmented Lagrangian coming from that has x and z coupled in a quadratic expression,

$$f(x)+g(z)+y{\cdot}[Ax-z]+\frac{r}{2}|Ax-z|^2,$$

and the challenge is seen as inventing procedures that interatively perform alternate minimizations of some approximating sort in the x and z variables. The PMM is demonstrated in [20] to provide an engine for that. In contrast, our approach to decomposition techniques heads down a different path toward the progressive decompling algorithm in [12], [13]. The PMM developments we undertake here will be able to help with that.

Optimality in our setting of convexity is naturally always global, of course, and can be characterized either by subgradients of φ or by those of l by way of the *Lagrangian mapping*

$$T_l(x,y):=\{\,(v,u)\mid v\in\partial_x l(x,y),\, u\in\partial_y[-l](x,y)\}\tag{1.5}$$

on which the PMM will rely. The subgradients are related to each other in general by

$$(v,y)\in\partial\varphi(x,u) \iff (v,u)\in T_l(x,y) \iff (x,u)\in\partial\varphi^*(v,y).\tag{1.6}$$

Under a constraint qualification, such as asking $\operatorname{ri}\operatorname{dom}\varphi$ to contain an element of form $(x,0)$, a vector $\bar{x}$ furnishes a (globally) optimal solution $(\bar{x},0)$ to (P) if and only if there is a vector $\bar{y}$ that fulfills with $\bar{x}$ the equivalent conditions

$$(0,\bar{y})\in\partial\varphi(\bar{x},0) \iff (0,0)\in T_l(\bar{x},\bar{y}) \iff (\bar{x},0)\in\partial\varphi^*(0,\bar{y})\tag{1.7}$$

[19, 11.48]. Even without a constraint qualification, these conditions guarantee not only the optimality of $\bar{x}$ in (P) but also the optimality of $\bar{y}$ in (D). The T_l version in (1.7) signifies that $(\bar{x},\bar{y})$ is a saddle

[2] The ambiguity in the Lagrangian formula when $f(x)=\infty$ and $g^*(y)=\infty$ is resolved by (1.1) as $l(x,y)=\infty$.

point of $l(x,y)$ with respect to minimizing in x and maximizing in y. It implies that the minimum in (P) equals the maximum in (D), and that the closed convex set

$$Z := \{\,(\bar{x},\bar{y}) \text{ satisfying the conditions (1.7)}\,\}, \tag{1.8}$$

when nonempty, is the product of the set of solutions to (P) and the set of solutions to (D);

$$Z \neq \emptyset \;\; \textit{is assumed here throughout.}$$

For more on this duality, see [8] and [19, Sections 11H, 11I].

Algorithm framework. The Lagrangian mapping T_l is *maximal monotone* [19, 12.27], and the PPA can therefore be applied to it to determine a pair $(\bar{x},\bar{y}) \in Z$ [9, Section 5]. That procedure, in its simplest implementation with parameter values $c_k > 0$, generates a sequence of pairs (x^k, y^k) from an initial pair (x^0, y^0) by

$$(x^{k+1}, y^{k+1}) = (I + c_k T_l)^{-1}(x^k, y^k), \tag{1.9}$$

and that sequence is sure to converge to some $(\bar{x},\bar{y}) \in Z$. This is the PMM in its *basic mapping* format, which reflects the original source of the "proximal method of multipliers" in [10] in the convex programming case in (1.3), ahead of a restatement there that employed augmented Lagrangians. Here we will want to work with not only that kind of restatement but also others where augmented Lagrangians are not in the spotlight. On the one hand, the iterations (1.9) mean that

$$(x^{k+1}, y^{k+1}) = \underset{x,y}{\operatorname{argminmax}}\left\{ l(x,y) + \frac{1}{2c_k}|x - x^k|^2 - \frac{1}{2c_k}|y - y^k|^2 \right\}, \tag{1.10}$$

the unique saddle point with respect to minimizing in x and maximizing in y. This is the PMM in its *saddle point* format. On the other hand, the iterations can be recast with the introduction of auxiliary vectors u^{k+1} as a doubled primal minimization followed by a dual update:

$$(x^{k+1}, u^{k+1}) = \underset{x,u}{\operatorname{argmin}}\left\{ \varphi(x,u) - y^k{\cdot}u + \frac{1}{2c_k}|x - x^k|^2 + \frac{c_k}{2}|u|^2 \right\}, \quad y^{k+1} = y^k - c_k u^{k+1}. \tag{1.11}$$

This is the PMM in its *doubled minimization* format. The *augmented* Lagrangians associated with problem (P), which enter the picture next, are defined in general by

$$l_c(x,y) = \min_u \left\{ \varphi(x,u) - y{\cdot}u + \frac{c}{2}|u|^2 \right\}, \tag{1.12}$$

but through Fenchel duality they can also be expressed by

$$l_c(x,y) = \max_{y'} \left\{ l(x,y') - \frac{1}{2c}|y' - y|^2 \right\}. \tag{1.13}$$

These functions are convex in x, concave in y, and differentiable in y for any x in the $\mathbb{R}^n$-projection of the convex set $\operatorname{dom}\varphi \subset \mathbb{R}^n \times \mathbb{R}^m$. The PMM in *augmented Lagrangian* format takes

$$x^{k+1} = \underset{x}{\operatorname{argmin}}\left\{ l_{c_k}(x, y^k) + \frac{1}{2c_k}|x - x^k|^2 \right\}, \qquad y^{k+1} = y^k + c_k \nabla_y l_{c_k}(x^{k+1}, y^k), \tag{1.14}$$

with gradient vectors $\nabla_y l_c(x,y)$ being the unique u minimizers in (1.12).

The equivalence of the four PMM formats holds if the steps are executed exactly but slips when stopping criteria for inexactness are brought in. The degree of inexactness must be controlled so as not to discrupt the convergence of the iterates (x^k, y^k) to some $(\bar{x}, \bar{y}) \in Z$, but different approaches to execution raise different demands. The basic theory of the PPA in [9] and its extensions in [15] offer criteria that replace the equation in (1.9) by bounds on the distance between the vectors on the left and right sides. Practicality depends, however, on having an estimate of that distance in terms of some readily computable quantity. Conditions that allow for approximate minimization in the versions of the algorithm in (1.11) and (1.14) can serve for this. Originally in [10], such conditions were provided for (1.14), but not for (1.11) — that being unnecessary because the augmented Lagrangians associated with the convex programming case in (1.3) have simple expressions. Sometimes, though, the definition in (1.12) doesn't yield a convenient expression, because the minimization can't be carried out in closed form. The minimization there may then need to be handled inexactly, too. The doubled minimization PMM format in (1.11) can then be helpful as a recourse to the augmented Lagrangian format.

Developing effective stopping criteria within our general framework is tops on our agenda, but it will be important to carry this out with added flexibility in the proximal terms in the minimization. More will be said about this shortly. Another key task is improving the understanding of when *linear* convergence is exhibited and what the rate of that may depend on.

Metric regularity role. For the original PMM in [10], Q-linear convergence of (x^k, y^k) to $(\bar{x}, \bar{y})$ was established through a condition that entailed $(\bar{x}, \bar{y})$ having to be the *sole* element of Z in particular, because that was the state of knowledge then for linear convergence of the underlying PPA algorithm. A subsequent PPA innovation of Luque [4] at least got Q-linear convergence to the solution set when it wasn't a singleton. Luque's condition had a handicap that curtailed applicability, but in [15] we came up with a relaxation of it that works just as well. That relaxed condition, translated to the situation here, asks the Lagrangian mapping T_l to have the following property at $(\bar{x}, \bar{y}) \in Z$:

$$\begin{array}{c} \exists\, a \in (0,\infty) \text{ such that, for } (x,y) \text{ and } (v,u) \text{ near to } (\bar{x},\bar{y}) \text{ and } (0,0), \\ (x,y) \in T_l^{-1}(v,u) \quad \Longrightarrow \quad \text{dist}\Big((x,y), T_l^{-1}(0,0)\Big) \leq a\Big|(v,u)\Big|, \end{array} \tag{1.15}$$

where by "near to" we mean of course "in some sufficiently small neighborhoods of." This is the *metric subregularity* of T_l relative to having $(0,0) \in T_l(\bar{x}, \bar{y})$. The corresponding *modulus of subregularity*,

$$\text{subreg}\Big(T_l; (\bar{x},\bar{y}) \,|\, (0,0)\Big) := \liminf \text{ of the } a \text{ values in (1.15) as neighborhoods dwindle,} \tag{1.16}$$

will be seen to keenly affect the *rate* of Q-linear convergence that the procedure is able to realize. But without the limit pair $(\bar{x}, \bar{y})$ being the sole element of Z, the Q-linear convergence will not quite be that of $(x^k, y^k) \to (\bar{x}, \bar{y})$, but that of $\text{dist}((x^k, y^k), Z) \to 0$.

Subregularity is a key condition in the theory of metric regularity properties of set-valued mappings laid out in the book [1]. Its manifestation for T_l relative to having $(\bar{x}, \bar{y}) \in T_l^{-1}(0,0)$ is a subform of *metric regularity*, which itself requires

$$\begin{array}{c} \exists\, a \in (0,\infty) \text{ such that, for } (x,y) \text{ and } (v',u') \text{ near to } (\bar{x},\bar{y}) \text{ and } (0,0), \\ (v,u) \in T_l(x,y) \quad \Longrightarrow \quad \text{dist}\Big((x,y), T_l^{-1}(v',u')\Big) \leq a\Big|(v.u) - (v',u')\Big|. \end{array} \tag{1.17}$$

That metric regularity of T_l is equivalent in turn to something called the Aubin property of T_l^{-1} relative to having $(\bar{x}, \bar{y}) \in T_l^{-1}(0,0)$, which describes a Lipshitz-like behavior of the set-valued mapping T_l^{-1} that entails it being nonempty-valued near $(0,0)$. The *strong* metric regularity of T_l relative to having $(0,0) \in T_l(\bar{x}, \bar{y})$ adds to this the single-valuedness of T_l^{-1} around $(0,0)$, thereby making that inverse mapping be Lipschitz continuous. For the convex programming case of T_l in (1.3), such strong metric regularity holds when (1) the gradients $\nabla f_i(\bar{x})$ of the active constraints are linearly independent and

(2) the Hessian $\nabla^2_{xx} l(\bar{x},\bar{y})$ is positive-definite relative to the subspace orthogonal to those vectors [1, 2G.9]. For nonlinear programming even without the convexity, these are the powerful conditions that guarantee local optimality along with primal-dual solution stability and are familiar in the convergence analysis of solution algorithms. Indeed, metric regularity properties on all levels are increasingly being realized as furnishing a range of stability-oriented enhancements of sufficient conditions for local optimality that can serve in the justification of numerical methodology.

The observation that the metric subregularity of T_l^{-1} in (1.15) fits neatly into this spectrum as the least demanding of properties is important to appreciating that our identification of it as key support for linear convergence of the PMM is a fitting development in this larger picture.

Variable-metric motivations. Reasons will now be given for why we want to extend the PMM and its convergence guarantees to implementations that deploy proximal terms less rigidly than in the past. A simple move toward such flexibility would be allowing the c_k in the proximal term for the x argument to be replaced in the iterations (1.10), (1.11) and (1.14) by a different value $b_k > 0$, so that the procedure would depend on two different sequences of parameters. An incentive for that comes from comparing the iterations (1.14) with those of the corresponding "nonproximal" method of multipliers that arises applying the PPA to maximizing $-\psi(0,y)$ in the dual problem (D). The iterations of that come out as

$$x^{k+1} \in \operatorname*{argmin}_x l_{c_k}(x,y^k), \qquad y^{k+1} = y^k - c_k \nabla_y l_{c_k}(x^{k+1},y^k), \tag{1.18}$$

and give the corresponding ALM to compare with the PMM. The version of (1.14) in which the c_k for the proximal term in x is replaced by a possibly much higher value b_k would bridge between the two procedures. It could be interpreted as introducing just a bit of strong convexity in order to stabilize the minimum and make it be uniquely attained.

An incentive to variability from another direction comes from contemplating the replacement of the c_k in the proximal term in x in (1.11) by $b_k = c_k^{-1}$, which would transform $\frac{1}{2c_k}|x-x^k|^2 + \frac{c_k}{2}|u|^2$ into $\frac{c_k}{2}|(x,u)-(x^k,0)|^2$. The desirability of this may seem obscure, but it is the key maneuver behind the progressive decoupling algorithm in [12], [13]. That scheme, with numerous specializations, concerns a problem in a space viewed as the product of two complementary subspaces which in principle could be coordinatized into vector pairs (x,u), as here, but may already be presented in a different application-oriented coordinate system. Progressive decoupling has, so far, only been justified for c_k *always having the same value* c and moreover *only with exact minimization in every iteration*. Establishing convergence with stopping criteria for inexact minimization and adjustable c_k, which is one of our goals in this paper, will improve it greatly.

The replacement of the c_k attached to the x variable in (1.11) by a different b_k severs the bond between the iterations in (1.10), (1.11) and (1.14), and the ones in (1.9), thereby undermining the convergence support provided by the theory behind (1.9). Fortunately, a connection can be restored by passing to a version of (1.9) in which the canonical Euclidean norm is replaced by different Euclidean norms that can vary with the iterations. The variable-metric version of the PPA developed recently in [15] and improved in [18] will be shown here to facilitate that.[3] Proximal terms based on alternative norms are anyway known to be useful in computation and, in particular, are prominent in the work of Shefi and Teboulle [20] on using the PMM as a tool for deriving decomposition algorithms. The norms there are fixed, and the proximal parameters as well, but the advances in this paper should enable those elements to vary without disrupting convergence.

[3]An earlier variable-metric PPA of Parente, Lotito and Solodov [5], despite its notable innovations, treats approximation in a completely different manner that rules out the possibility of deriving implementable stopping criteria in our context. This is explained in [18].

In admitting general proximal terms, provision is made in particular for preconditioning of one form or another. Perhaps the right preconditioning might be guessed from the start, and there would be no need then for metrics to vary. Beyond preconditioning, however, there are more basic issues of contrasted scaling that might be helpful between primal and dual components or various primal subcomponents. It seems good from that angle to allow changes from one iteration to the next as a foundation perhaps to future algorithmic developments in which adjustments can be imposed from the information being produced in ongoing computations.

Other prospectives. For practicality, we have chosen to focus here entirely on convex optimization, but applications to nonconvex optimization can be glimpsed on the horizon. The proximal point algorithm itself is capable of operating in a local sense, without requiring global maximal monotonicity of a mapping. We have shown in [16] that this enables the ordinary method of multipliers (ALM) to solve nonconvex problems of "generalized nonlinear programming" when initiated near enough to a point satisfying a new *variationally sufficient* condition for local optimality. That direction of development is also in the offing for the PMM as applied to generalized nonlinear programming, which would extend and complement the 2022 results of Ismailov and Solodov in [2] for standard nonlinear programming. Equally, there is the prospect of tackling nonconvex linkage problems by the progressive decoupling algorithm through such localization in building on the patterns developed here.

Plan of the paper. The central PMM results will be formulated in Section 2 and proved in Section 3. They will be applied then in Section 4 to obtain linear convergence of the progressive decoupling algorithm in a broader context than just the minimization problem (P). Section 5 will identify a situation where both PMM and progressive decoupling are sure to converge Q-linearly and it will establish that they can be expected to exhibit that convergence generically.

2 Extended algorithmic platform and statement of central results

Motivations for replacing the algorithm's proximal terms in the x-argument have already been explained, but those terms in the u argument can benefit from that as well. Accordingly we look at alternative norms and inner products induced by symmetric positive-definite matrices C through

$$\langle u',u\rangle_C = u'{\cdot}Cu \quad ||u||_C=\sqrt{\langle u,u\rangle_C},\ \text{ so that }\ ||u||_C^2=u{\cdot}Cu,\quad \nabla||u||_C^2=2Cu. \tag{2.1}$$

The canonical Euclidean norm $|\cdot|$ corresponds, of course, to $C=I$. Distances with respect to $|\cdot|$, $||\cdot||_C$ and $||\cdot||_{C^{-1}}$ are related in terms of the maximum and minimum eigenvalues of C by

$$\begin{gathered} ||C^{-1}||^{-1}{\cdot}|u|^2\le||u||_C^2\le||C||{\cdot}|u|^2, \qquad ||C||^{-1}{\cdot}||u||_C^2\le|u|^2\le||C^{-1}||{\cdot}||u||_C^2, \\ \text{where }\ ||C||=\operatorname{maxeig}(C),\quad ||C^{-1}||^{-1}=\operatorname{mineig}(C),\quad \text{and} \\ [1/\operatorname{maxeig}(C)]\,||u||_C\ \le\ ||u||_{C^{-1}}\ \le\ [1/\operatorname{mineig}(C)]\,||u||_C. \end{gathered} \tag{2.2}$$

(The validity of these useful relationships is elementary in the case of C being a diagonal matrix, but they hold then in general for symmetric positive-definite C through diagnalization under a change to a different orthonormal basis.) We expand the notion of augmented Lagrangian by associating with choices of C the functions

$$\begin{gathered} l_C(x,y)=\inf_u\Big\{\,\varphi(x,u)-y{\cdot}u+\tfrac{1}{2}||u||_C^2\Big\}\ \text{ with} \\ \nabla_y l_C(x,y)=-\operatorname{argmin}_u\Big\{\,\varphi(x,u)-y{\cdot}u+\tfrac{1}{2}||u||_C^2\Big\}. \end{gathered} \tag{2.3}$$

These reduce to the previous augmented Lagrangians in the notation l_c when $C=cI$.

This generalization of augmented Lagrangian allows the diagonal matrix $C = cI$ to be replaced, for instance, by an arbitrary diagonal matrix with positive entries. That would let different components of u be guided in computations by different proximal parameters, in an elaboration of the idea already suggested of allowing different parameters for x and u, as a sort of progressive rescaling feature. The advantage of accommodating nondiagonal C will be made clear later, in Section 4.

The augmented Lagrangian formula in (2.2) translates through Fenchel duality to an equivalent formula involving the Lagrangian l itself that generalizes (1.13),

$$l_C(x, y) = \sup_{y'} \Big\{ l(x, y') - \tfrac{1}{2} ||y' - y||^2_{C^{-1}} \Big\}, \tag{2.4}$$

because the quadratic functions $u \mapsto \frac{1}{2} u{\cdot}Cu$ and $y \mapsto \frac{1}{2} y{\cdot}C^{-1}y$ are conjugate to each other, as are the functions $\varphi(x, \cdot)$ and $-l(x, \cdot)$.

With this in hand, we are ready to formulate the extension of the proximal method from its formats in the introduction to ones that not only permit proximal terms of greater generality, but also let them vary from one iteration to the next. The parametric elements of the algorithm in their dependence on the iteration k will be a choice of

$$c_k \geq 1 \ \text{ such that } \ c_k \to c_\infty \leq \infty, \tag{2.5}$$

along with symmetric, positive-definite matrices $B_k \in I\!R^{n\times n}$ and $C_k \in I\!R^{m\times m}$ satisfying with respect to the relations

$$\begin{array}{c} \beta_k^{-1} ||x||_{B_{k-1}} \leq ||x||_{B_k} \leq \beta_k ||x||_{B_{k-1}} \ \text{ and } \ \beta_0^{-1} |x| \leq ||x||_{B_0} \leq \beta_0 |x|, \\ \gamma_k^{-1} ||y||^{-1}_{C_{k-1}} \leq ||y||_{C_k^{-1}} \leq \gamma_k ||y||_{C_{k-1}^{-1}} \ \text{ and } \ \gamma_0^{-1} |x| \leq ||x||_{C_0^{-1}} \leq \gamma_0 |x|, \\ \text{where } \beta_k \geq 1,\ \gamma_k \geq 1, \text{ with } \ \Pi_{k=0}^{\infty} \beta_k =: \beta < \infty,\ \Pi_{k=0}^{\infty} \gamma_k =: \gamma < \infty. \end{array} \tag{2.6}$$

The matrix sequences in question could in particular, of course, just be constant, and then (2.6) would be satisfied with $\beta_k \equiv 1$ and $\gamma_k \equiv 1$. But in any case, only the existence of β_k and γ_k will be needed, not their particular values. Or, from the opposite angle, one can interpret β_k and γ_k as given in advance for the sake of controlling the generation of B_k and C_k from B_{k-1} and C_{k-1} as the algorithm proceeds. Either way, the relations in (2.6) have the advantage of implying

$$\lim_{k\to\infty} B_k = B_\infty, \quad \lim_{k\to\infty} C_k = C_\infty, \ \text{ again symmetric and positive-definite} \tag{2.7}$$

as proven in [18, Section 2]. This keeps the shifts in metric from getting out of hand.

The iterations will generate from an initial pair (x^0, y^0) a sequence of pairs (x^k, y^k) by rules that not only broaden those in the different formats presented in Section 1, but also permit inexactness in the calculations, as signaled below by "$\approx$." More about controlling these approximations will be brought in shortly. For reasons of exposition, the order of the format types will now be the opposite of that in Section 1.

Algorithm 1: extended PMM in augmented Lagrangian format.

$$\begin{aligned} x^{k+1} &\approx \operatorname*{argmin}_x \Big\{ l_{c_k C_k}(x, y^k) + \frac{1}{2c_k} ||x - x^k||^2_{B_k} \Big\} \\ y^{k+1} &= y^k + c_k C_k \nabla_y l_{c_k C_k}(x^{k+1}, y^k), \end{aligned} \tag{2.8}$$

Having $c_k C_k$ determine the augmented Lagrangian here might seem odd. Couldn't c_k be subsumed into C_k? There are two answers. The first is that possibly $c_k \to \infty$ in (2.5), whereas the C_k sequence is bounded through (2.7). The second is that c_k enters also in the proximal term in the x-argument. The effect of these features will be seen later in the statement of convergence results.

For connecting with the motivations in Section 1, note for example that the choices $C_k = I$, $B_k = b_k c_k I$, would not only replace $l_{c_k C_k}$ by the earlier l_{c_k}, but also turn the x-proximal term in (2.8) into $(b_k/2)|x - x^k|^2$. The limit of $b_k c_k$ must be positive and finite, to be consistent with having $B_k \to B_\infty$ in (2.7).

Algorithm 2: extended PMM in doubled minimization format.

$$\begin{aligned}(x^{k+1}, u^{k+1}) &\approx \underset{x,u}{\operatorname{argmin}}\left\{ \varphi(x,u) - u{\cdot}y^k + \frac{1}{2c_k}||x - x^k||^2_{B_k} + \frac{c_k}{2}||u - u^k||^2_{C_k} \right\} \\ y^{k+1} &= y^k - c_k C_k u^{k+1},\end{aligned} \tag{2.9}$$

Here, for instance, by taking $C_k = I$ but $B_k = c_k^2 I$, the augmented Lagrangian would be just l_{c_k} while the x-proximal term would be $(c_k/2)|x - x^k|^2$, which was indicated in Section 1 as tying in with schemes of problem decomposition in [12], [13]. The extension introduced here will furnish more potential in that direction.

Although Algorithms 1 and 2 are central for the applications we envision, the other two formats in Section 1 have variable-metric counterparts which primarily will serve in the theoretical picture.

Algorithm 3: extended PMM in saddle point format.

$$(x^{k+1}, y^{k+1}) \approx \underset{x,y}{\operatorname{argminmax}}\left\{ l(x,y) + \frac{1}{2c_k}||x - x^k||^2_{B_k} - \frac{1}{2c_k}||y - y^k||^2_{C_k^{-1}} \right\}. \tag{2.10}$$

Algorithm 4: extended PMM in basic mapping format.

$$(x^{k+1}, y^{k+1}) \approx (I + c_k D_k^{-1} T_l)^{-1}(x^k, y^k) \;\; \text{for} \;\; D_k := \begin{bmatrix} B_k & 0 \\ 0 & C_k^{-1} \end{bmatrix}. \tag{2.11}$$

The fourth version is "basic" because the others will be derived from it by way of the variable-metric proximal point algorithm in [15], as improved in [18], and the following fact, the proof of which will be given in Section 3 along with those of the other theorems stated in this section.

Theorem 2.1 (equivalence with exact computations)**.** *Algorithms 1, 2, 3 and 4 with "≈" replaced by "=" produce the same sequence of pairs (x^k, y^k) from the initial (x^0, y^0).*

The aim, of course, is to get the whichever-way-generated sequence of pairs (x^k, y^k) to eventually approach the primal-dual solution set Z in (1.8), and even better, to converge to a particular solution pair in Z.

With computations only approximate, the sequences produced in the four PMM formats can be different, because the style of approximation has to be tailored to the the type of optimization being carried out in the iterations. The appoximation in Algorithm 4 will be the key to applying the convergence theory in [18]. As a practical matter, the approximations in Algorithms 1 and 2 are of keenest interest, but rules articulated for them have to support the requirements of Algorithms 4 and the approximations in Algorithm 3. Those rules, in the form of stopping criteria for the optimization steps, will appeal to the distance between the origin and the subgradient set for the expression being optimized. Exact optimization would correspond to that distance being 0. Convenient upper bounds for the distances in question can serve in place of the distances themselves.

It will best, for the sake of theory, to express the subgradient distances as measured in the norms associated with B_k and C_k along with D_k in (2.11) and

$$E_k = \begin{bmatrix} B_k & 0 \\ 0 & C_k \end{bmatrix}, \tag{2.12}$$

and with their inverses, which are related by

$$\begin{array}{ll} ||(x,y)||^2_{D_k} = ||x||^2_{B_k} + ||y||^2_{C_k^{-1}}, & ||(v,u)||^2_{D_k^{-1}} = ||v||^2_{B_k^{-1}} + ||u||^2_{C_k}, \\ ||(x,u)||^2_{E_k} = ||x||^2_{B_k} + ||u||^2_{C_k}, & ||(v,y)||^2_{E_k^{-1}} = ||x||^2_{B_k^{-1}} + ||u||^2_{C_k^{-1}}. \end{array} \tag{2.13}$$

Such variable-metric distances can in turn be estimated by distances under the norm $|\cdot|$ through the classical bounds in (2.2):

$$\begin{array}{c} \sqrt{\text{mineig}(C_k)}\cdot \text{dist} \le \text{dist}_{C_k} \le \sqrt{\text{maxeig}(C_k)}\cdot \text{dist}, \\ \dfrac{1}{\sqrt{\text{mineig}(C_k)}}\cdot \text{dist} \le \text{dist}_{C_k^{-1}} \le \dfrac{1}{\sqrt{\text{maxeig}(C_k)}}\cdot \text{dist}\,. \end{array} \tag{2.14}$$

and likewise for the other metrics. In that way, workable stopping criteria in ordinary distances can be made available as alternatives, but these are generally stricter in the sense of requiring tighter accuracy. The stopping criteria will all draw on a choice of error parameters

$$\varepsilon_k \in (0,1) \text{ such that } \sum\nolimits_{k=0}^{\infty} \varepsilon_k < \infty. \tag{2.15}$$

Theorem 2.2 (global convergence with inexact computations)**.** *Regardless of the initial* (x^0, y^0), *the sequence of pairs* (x^k, y^k) *generated by Algorithms 1, 2 and 3 will converge to some particular solution pair* $(\bar{x}, \bar{y}) \in Z$ *in (1.8) as long as the stopping criteria for the approximations are taken to be the following.*

(a) *For Algorithm 1, where an approximate minimizer must be calculated for the strongly convex function*

$$\lambda^k(x) := l_{c_k C_k}(x, y^k) + \frac{1}{2c_k}||x - x^k||^2_{B_k} \tag{2.16}$$

and exact computation would require $0 \in \partial\lambda^k(x^{k+1})$, *the stopping criterion is*

$$c_k \, \text{dist}_{B_k^{-1}}\Big(0, \partial\lambda^k(x^{k+1})\Big) \le \varepsilon_k \tag{2.17a}$$

or the stricter alternative

$$\frac{c_k}{\sqrt{\text{mineig}(B_k)}}\cdot \text{dist}\Big(0, \partial\lambda^k(x)\Big) \le \varepsilon_k. \tag{2.17b}$$

(b) *For Algorithm 2, where an approximate minimizer must be calculated for the strongly convex function*

$$\varphi^k(x,u) := \varphi(x,u) - \langle y^k, u\rangle + \frac{1}{2c_k}||x - x^k||^2_{B_k} + \frac{c_k}{2}||u||^2_{C_k} \tag{2.18}$$

and exact computation would require $(0,0) \in \partial\varphi^k(x^{k+1}, u^{k+1})$, *the stopping criterion is*

$$2c_k \, \text{dist}_{E_k^{-1}}\Big((0,0), \partial\varphi^k(x^{k+1}, u^{k+1})\Big) \le \varepsilon_k \tag{2.19a}$$

or the stricter alternative

$$\frac{2c_k}{\sqrt{\text{mineig}(E_k)}}\cdot \text{dist}\Big((0,0), \partial\varphi^k(x^{k+1}, u^{k+1})\Big) \le \varepsilon_k. \tag{2.19b}$$

(c) *For Algorithm 3, where an approximate saddle point must be calculated for the strongly convex-concave function*

$$l^k(x,y) := l(x,y) + \frac{1}{2c_k}||x - x^k||^2_{B_k} - \frac{1}{2c_k}||y - y^k||^2_{C_k^{-1}} \tag{2.20}$$

and exact computations would require $(0,0) \in T_{l^k}(x^{k+1}, y^{k+1})$ *for the mapping* T_{l^k} *associated with* l^k *as* T_l *with* l *in (1.5), the stopping criterion is*

$$c_k \operatorname{dist}_{D_k^{-1}}\Big((0,0), T_{l^k}(x^{k+1}, y^{k+1})\Big) \leq \varepsilon_k \tag{2.21a}$$

or the stricter alternative

$$\frac{c_k}{\sqrt{\operatorname{mineig}(D_k)}} \cdot \operatorname{dist}\Big((0,0), T_{l^k}(x^{k+1}, y^{k+1})\Big) \leq \varepsilon_k. \tag{2.21b}$$

The next question to answer is whether linear convergence will be obtained. A sequence of positive values α_k is said to converge Q-linearly to 0 at a rate r if $\limsup_{k\to\infty}(\alpha_{k+1}/\alpha_k) \leq r < 1$, and this is said to be *superlinear* convergence if $r = 0$.[4] Here we'll be concerned with whether the distance of (x^k, y^k) from the solution set Z converges Q-linearly to 0. The rate of that convergence will depend on the norm in which that distance is measured. The norm associated with the limit matrix D_∞ will be the best.

The metric subregularity property in (1.15) now comes on stage, but adapted to a changed standard for distances associated with D_∞. The new version requires that

$$\begin{array}{c} \exists\, a \in (0,\infty) \ \text{ such that, for } (x,y) \text{ and } (v,u) \text{ near to } (\bar{x},\bar{y}) \text{ and } (0,0), \\ (v,u) \in T_l(x,y) \quad \Longrightarrow \quad \operatorname{dist}_{D_\infty}\Big((x,y), T_l^{-1}(0,0)\Big) \ \leq \ a||(v,u)||_{D_\infty^{-1}}. \end{array} \tag{2.22}$$

This is equivalent to the original condition in (1.15) through the various bounds relating the norms, *but the range of values of a can be different.* Instead of the modulus in (1.16), we will want to appeal to the D_∞ version of that modulus, namely

$$\operatorname{subreg}_{D_\infty}\Big(T_l; (\bar{x},\bar{y}) \,|\, (0,0)\Big) := \liminf \text{ of the } a \text{ values in (2.22) as neighborhoods dwindle.} \tag{2.23}$$

Theorem 2.3 (linear convergence)**.** *Tighten the stopping criteria (2.17ab), (2.19ab) and (2.21ab), by having*

$$\begin{array}{rl} \varepsilon_k \textit{ replaced on the right sides by: } & \varepsilon_k \min\Big\{1, ||(x^{k+1}, y^{k+1}) - (x^k, y^k)||_{D_k}\Big\} \\ \textit{or by its lower estimate: } & \varepsilon_k \min\left\{1, \dfrac{|(x^{k+1}, y^{k+1}) - (x^k, y^k)|}{\sqrt{\operatorname{mineig}(D_k)}}\right\}. \end{array} \tag{2.24}$$

For the iterates (x^k, y^k) *generated thereby as* $D_k \to D_\infty$*, suppose the limit* $(\bar{x},\bar{y})$ *guaranteed by Theorem 2.2 is such that the metric subregularity in (1.15) or equivalently (2.22) holds there, and let* a_∞ *denote the associated modulus in (2.23). Then*

$$\operatorname{dist}_{D_\infty}((x^k, y^k), Z) \ \textit{ converges Q-linearly to 0 at the rate } \ r = \frac{1}{\sqrt{1 + (c_\infty/a_\infty)^2}}, \tag{2.25}$$

this being superlinear convergence when $c_\infty = \infty$ *or* $a_\infty = 0$.[5]

Of course, Q-linear convergence to 0 of the distance of (x^k, y^k) from $(\bar{x},\bar{y})$ is also of interest as a slightly stronger property, when Z is not just the singleton $\{(\bar{x},\bar{y})\}$. Advanced results about the

[4] The possibility of $\alpha_k = 0$ can be covered through a work-around that for every $r' > r$ requires $\alpha_{k+1} \leq r'\alpha_k$ when k is sufficiently large.

[5] For the sake of these extremes, the expression for r would better be taken in the form $a_\infty/\sqrt{a_\infty^2 + c_\infty^2}$, but the form in (2.25) is superior otherwise for emphasizing that the rate only depends on the ratio between c_∞ and a_∞ when they are finite.

proximal point algorithm in [15] provide a supplementary condition which, with further tightening of the stopping criteria, leads to such direct linear convergence in the *fixed-metric* setting. That feature was not replicated in the variable-metric setting in [15] or its improved version in [18], because it seems unable to cope with the additional uncertainties of repeatedly shifting geometry.

3 Derivation from theory of the proximal point algorithm

The convergence results in Section 2 for the PMM will be derived next from results about the proximal point algorithm (PPA), including recent developments in [15] and [18]. To facilitate this and at the same time open the door to applications outside of optimization, we begin with a translation of current framework of functions and their subgradients into one of mappings and their partial inverses.

In place of the set-valued mappings $\partial\varphi$ and T_l in (1.5) and (1.6), which are maximal monotone and in a partial inverse relationship to each other, we consider a general pair of maximal monotone mappings in that relationship: $M : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m$ and $T : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m$ such that

$$(v,u) \in T(x,y) \iff (v,y) \in M(x,u). \tag{3.1}$$

In a generalization of the problem of trying to find $\bar{x}$ and $\bar{y}$ that satisfy the optimality condition in (1.7), we seek to

$$\begin{array}{l} \text{determine } (\bar{x},\bar{y}) \text{ with } (0,\bar{y}) \in M(\bar{x},0), \\ \text{or equivalently, } (\bar{x},\bar{y}) \in T^{-1}(0,0) =: Z. \end{array} \tag{3.2}$$

We aim to solve this by applying the PPA to T. In our variable-metric framework, with its proximal matrices D_k and parameters c_k, that procedure means starting from a choice of (x^0, y^0) and iterating with

$$(x^{k+1}, y^{k+1}) \approx P_k(x^k, y^k) := (I + c_k D_k^{-1} T)^{-1}(x^k, y^k) \tag{3.3}$$

with the approximation being governed by error parameters ε_k as in (2.20) through

$$||(x^{k+1}, y^{k+1}) - P_k(x^k, y^k)||_{D_k} \le \begin{cases} \text{(a)} \ \ \varepsilon_k, \ \text{ or} \\ \text{(b)} \ \ \varepsilon_k \min\left\{1, ||(x^{k+1}, y^{k+1}) - (x^k, y^k)||_{D_k}\right\}. \end{cases} \tag{3.4}$$

The mapping P_k is single-valued and nonexpansive with respect to the D_k metric, because $D_k^{-1}T$ is maximal monotone with respect to the inner product associated with that metric — as was noted in [15, (4.3)]. Moreover, its set of fixed points is the solution set Z in (3.2), and therefore

$$\text{dist}_{D_k}(P_k(x^k, y^k), T^{-1}(0,0)) \le \text{dist}_{D_k}((x^k, y^k), T^{-1}(0,0)),$$

in fact with strict inequality[6] unless (x^k, y^k) is already in $T^{-1}(0,0)$. If the matrices D_k were all the same, and $(x^{k+1}, y^{k+1}) = P_k(x^k, y^k)$ exactly in (3.3), the sequence of iterates would get ever closer to the solution set right from the start. But with D_k varying and approximations permitted, things are more complicated.

We nonetheless have at our disposal, in support of this solution approach, the following picture of global convergence and potential linear convergence, with the latter relying on the metric subregularity property we have been looking at. Now, though, that is a property of our general mapping T, not just the Lagrangian mapping T_l:

$$\begin{array}{c} \exists\, a \in (0,\infty) \text{ such that, for } (x,y) \text{ and } (v,u) \text{ near to } (\bar{x},\bar{y}) \text{ and } (0,0), \\ (v,u) \in T(x,y) \implies \text{dist}_{D_\infty}\Big((x,y), T^{-1}(0,0)\Big) \le a||(v,u)||_{D_\infty^{-1}} \end{array} \tag{3.5}$$

[6]Because the nonexpansivity is actually "firm nonexpansvity."

with its modulus

$$\operatorname{subreg}_{D_\infty}\Big(T;(\bar{x},\bar{y})\,|\,(0,0)\Big) := \liminf \text{ of } a \text{ values in (3.5) as neighborhoods dwindle.} \tag{3.6}$$

Theorem 3.1 (convergence in the PPA format)**.** *Under the stopping criterion (3.4a), the sequence of iterates (x^k,y^k) generated from any initial (x^0,y^0) will converge to some particular $(\bar{x},\bar{y})$ in the primal-dual solution set $Z := T^{-1}(0,0)$. If the stricter stopping criterion (3.4b) is used and the property in (3.5) holds for $(\bar{x},\bar{y})$, then with a_∞ denoting the subregularity modulus (3.6),*

$$\operatorname{dist}_{D_\infty}((x^k,y^k),Z) \text{ converges Q-linearly to 0 at the rate } r = \frac{1}{\sqrt{1+(c_\infty/a_\infty)^2}}, \tag{3.7}$$

this being superlinear convergence when $c_\infty = \infty$.

Proof. This specializes [18, Theorems 2.1 and 2.2] (improving on [15, Theorems 4.1 and 4.2]) to our product-space setting with T being maximal monotone globally, instead of just locally. Our assumption (2.2) about the B_k and C_k matrices enters into this, since the result being specialized requires here that

$$\begin{array}{l} \alpha_k^{-1}||(x,y)||_{D_{k-1}} \leq ||(x,y)||_{D_k} \leq \alpha_k||(x,y)||_{D_{k-1}} \\ \quad \text{and } \ \alpha_0^{-1}|x| \leq ||(x,y)||_{D_0} \leq \alpha_0|x|, \\ \quad \text{with } \ \alpha_k \geq 1 \ \text{ such that } \ \Pi_{k=0}^{\infty}\alpha_k < \infty. \end{array}$$

That holds via (2.6) with $\alpha_k = \beta_k\gamma_k$, because $\alpha_k \geq \max\{\beta_k,\gamma_k\} \geq 1$ when $\beta_k \geq 1$, $\gamma_k \geq 1$, so that

$$\begin{aligned} ||(x,y)||^2_{D_k} &= ||x||^2_{B_k} + ||y||_{C_k^{-1}} \leq \beta_k^2||x||^2_{B_{k-1}} + \gamma_k^2||y||^2_{C_{k-1}^{-1}} \\ &\leq \alpha_k^2\Big(||x||^2_{B_{k-1}} + ||y||^2_{C_{k-1}^{-1}}\Big) = \alpha_k^2||(x,y)||^2_{D_{k-1}}, \end{aligned}$$

and in parallel

$$\begin{aligned} ||(x,y)||^2_{D_k} &= ||x||^2_{B_k} + ||y||_{C_k^{-1}} \geq \beta_k^{-2}||x||^2_{B_{k-1}} + \gamma_k^{-2}||y||^2_{C_{k-1}^{-1}} \\ &\geq \alpha_k^{-2}\Big(||x||^2_{B_{k-1}} + ||y||^2_{C_{k-1}^{-1}}\Big) = \alpha_k^{-2}||(x,y)||^2_{D_{k-1}}. \end{aligned}$$

Furthermore $\Pi_{k=0}^{\infty}\alpha_k = \lim\limits_{j\to\infty}(\Pi_{k=0}^{j}\beta_k)(\Pi_{k=0}^{j}\gamma_k) = (\Pi_{k=0}^{\infty}\beta_k)(\Pi_{k=0}^{\infty}\gamma_k) < \infty.$ □

The iterations in (3.3) can be expressed equivalently in other ways, which will be important later in relating them to the algorithmic formats in Section 2. Working first with the exact versions, we observe that

$$\begin{array}{l} (x^{k+1},y^{k+1}) = (I + c_kD_k^{-1}T)^{-1}(x^k,y^k) \\ \iff \ (x^k,y^k) \in (I + c_kD_k^{-1}T)(x^{k+1},y^{k+1}) \\ \iff \ (x^k,y^k) - (x^{k+1},y^{k+1}) \in c_kD_k^{-1}T(x^{k+1},y^{k+1}) \\ \iff \ c_k^{-1}D_k(x^k - x^{k+1}, y^k - y^{k+1}) \in T(x^{k+1},y^{k+1}) \\ \iff \ (c_k^{-1}B_k(x^k - x^{k+1}), c_k^{-1}C_k^{-1}(y^k - y^{k+1})) \in T(x^{k+1},y^{k+1}) \end{array} \tag{3.8}$$

and therefore

$$\iff \ (0,0) \in T(x^{k+1},y^{k+1}) + (c_k^{-1}B_k(x^{k+1} - x^k), c_k^{-1}C_k^{-1}(y^{k+1} - y^k)). \tag{3.9}$$

Through (3.1) that can be converted to

$$(c_k^{-1}B_k(x^k - x^{k+1}), y^{k+1}) \in M(x^{k+1}, c_k^{-1}C_k^{-1}(y^k - y^{k+1})) \tag{3.10}$$

and identified with

$$(0,0) \in M(x^{k+1}, c_k^{-1} C_k^{-1}(y^k - y^{k+1})) + (c_k^{-1} B_k(x^{k+1} - x^k), -y^{k+1}). \tag{3.11}$$

This brings us to two different versions that allow for inexact computations in generating a sequence of pairs (x^k, y^k) from some initial (x^0, y^0).

Algorithm A: partial-inverse PPA in primal-dual format.

$$\begin{array}{l} (0,0) \approx\in T^k(x^{k+1}, y^{k+1}), \quad \textit{where} \\ T^k(x,y) := T(x^{k+1}, y^{k+1}) + (c_k^{-1} B_k(x^{k+1} - x^k), c_k^{-1} C_k^{-1}(y^{k+1} - y^k)). \end{array} \tag{3.12}$$

Algorithm B: partial-inverse PPA in fully primal format.

$$\begin{array}{l} y^{k+1} = y^k - c_k C_k u^{k+1} \ \ \textit{with} \ \ (0,0) \approx\in M^k(x^{k+1}, u^{k+1}), \quad \textit{where} \\ M^k(x,u) := M(x,u) + (c_k^{-1} B_k(x - x^k), c_k C_k u - y^k). \end{array} \tag{3.13}$$

The approximate membership in each case must be refined into a variable-metric distance condition, and we do that next.

Theorem 3.2 (stopping criteria for the partial-inverse PPA)**.** *For Algorithm A, the left side in (3.4) can be replaced by the right side in the upper estimate that*

$$||(x^{k+1}, y^{k+1}) - P_k(x^k, y^k)||_{D_k} \ \le \ c_k \,\mathrm{dist}_{D_k^{-1}}\Big(0, T^k(x^{k+1}, y^{k+1})\Big). \tag{3.14}$$

For Algorithm B, the left side in (3.4) can be replaced by the right side in the upper estimate that

$$\begin{array}{c} ||(x^{k+1}, y^{k+1}) - P_k(x^k, y^k)||_{D_k} \ \le \ c_k ||v||_{B_k^{-1}} + 2||z||_{C_k^{-1}} \\ \textit{for some} \ \ (v,z) \in M^k(x^{k+1}, u^{k+1}), \end{array} \tag{3.15}$$

or instead the upper estimate that

$$||(x^{k+1}, y^{k+1}) - P_k(x^k, y^k)||_{D_k} \ \le \ 2c_k \,\mathrm{dist}_{E_k^{-1}}\Big(0, M^k(x^{k+1}, u^{k+1})\Big). \tag{3.16}$$

Proof. To establish (3.14), it's enough to show that the left side is bounded from above by $c_k||(v,u)||_{D_k^{-1}}$ when $(v,u) \in T^k(x^{k+1}, y^{k+1})$, or in other words, when $(0,0)$ is replaced by (v,u) on the left side (3.9). From the formula for D_k, we can elaborate this as

$$\begin{array}{l} (v,u) \in T(x^{k+1}, y^{k+1}) + c_k^{-1} D_k(x^{k+1} - x^k, y^{k+1} - y^k)] \\ \quad \Longleftrightarrow \quad c_k D_k^{-1}(v,u) \in c_k D_k^{-1} T(x^{k+1}, y^{k+1}) + [(x^{k+1}, y^{k+1}) - (x^k, y^k)] \\ \quad \Longleftrightarrow \quad (x^k, y^k) + c_k D_k^{-1}(v,u) \in [I + c_k D_k^{-1} T](x^{k+1}, y^{k+1}) \\ \quad \Longleftrightarrow \quad (x^{k+1}, y^{k+1}) = P_k\left((x^k, y^k) + c_k D_k^{-1}(v,u)\right). \end{array} \tag{3.17}$$

Drawing on the nonexpansivity of P_k in the D_k metric, we obtain from the equation at the end of (3.17) that

$$\begin{array}{l} ||(x^{k+1}, y^{k+1}) - P_k(x^k, y^k)||_{D_k} = ||P_k((x^k, y^k) + c_k D_k^{-1}(v,u)) - P_k(x^k, y^k)||_{D_k} \\ \quad \le ||[(x^k, y^k) + c_k D_k^{-1}(v,u)] - (x^k, y^k)||_{D_k} = ||c_k D_k^{-1}(v,u)||_{D_k} \\ \quad = c_k \langle D_k^{-1}(v,u), D_k D_k^{-1}(v,u)\rangle^{\frac{1}{2}} = c_k ||(v,u)||_{D_k^{-1}}, \end{array}$$

as needed for (3.14).

Following a similar path toward establishing (3.15), in which $u^{k+1} = c_k^{-1} C_k^{-1}(y^k - y^{k+1})$, we put (v, z) in place of $(0, 0)$ on the left side of (3.11) and proceed with this as

$$\begin{aligned}
&(v - c_k^{-1} B_k(x^{k+1} - x^k), z + y^{k+1}) \in M(x^{k+1}, c_k^{-1} C_k^{-1}(y^k - y^{k+1})) \\
&\iff \quad (v - c_k^{-1} B_k(x^{k+1} - x^k), c_k^{-1} C_k^{-1}(y^k - y^{k+1})) \in T(x^{k+1}, y^{k+1} + z) \\
&\iff \quad (v, 0) + c_k^{-1} D_k((x^k, y^k + z) - (x^{k+1}, y^{k+1} + z)) \in T(x^{k+1}, y^{k+1} + z) \\
&\iff \quad c_k D_k^{-1}(v, 0) + (x^k, y^k + z) \in (I + c_k D_k^{-1} T)(x^{k+1}, y^{k+1} + z) \\
&\iff \quad (x^{k+1}, y^{k+1} + z) = P_k\Big((x^k, y^k + z) + c_k D_k^{-1}(v, 0)\Big) = P_k\Big((x^k, y^k) + (c_k B_k^{-1} v, z)\Big).
\end{aligned}$$

With this expression for $(x^{k+1}, y^{k+1} + z)$, the nonexpansivity of P_k in the D_k metric can be invoked again to see that

$$\begin{aligned}
||(x^{k+1}, y^{k+1} + z) - P_k(x^k, y^k)||_{D_k} &= ||P_k\Big((x^k, y^k) + (c_k B_k^{-1} v, z)\Big) - P_k(x^k, y^k)||_{D_k} \\
&\leq ||(c_k B_k^{-1} v, z)||_{D_k}.
\end{aligned}$$

Since $||(x^{k+1}, y^{k+1} + z) - P_k(x^k, y^k)||_{D_k} \geq ||(x^{k+1}, y^{k+1}) - P_k(x^k, y^k)||_{D_k} - ||(0, z)||_{D_k}$, this gives us

$$||(x^{k+1}, y^{k+1}) - P_k(x^k, y^k)||_{D_k} \leq ||(c_k B_k^{-1} v, z)||_{D_k} + ||(0, z)||_{D_k}. \tag{3.18}$$

There are two ways to go now. One way is to use $||(c_k B_k^{-1} v, z)||_{D_k} \leq ||(c_k B_k^{-1} v, 0)||_{D_k} + ||(0, z)||_{D_k}$ and observe that the formula for D_k makes the first term on the right reduce to $c_k ||v||_{B_k^{-1}}$ and the second term reduce to $||z||_{C_k^{-1}}$. That leads to the upper bound in (3.15). The other way is to recall that $c_k \geq 1$ was assumed in (2.5) and combine that with the fact that $||(0, z)||_{D_k} \leq ||(w, z)||_{D_k}$ for any w to see that the right side of (3.18) is bounded above by $2c_k ||(B_k^{-1} v, z)||_{D_k}$, which the formula for D_k turns into $2c_k ||(v, z)||_{E_k^{-1}}$. That produces the alternative upper bound in (3.16). □

Our attention turns now to verifying the theorems of Section 2 as the application of these general results to the case of T being the mapping T_l defined in (1.5).

Proof of Theorem 2.1. The PPA iterations (3.3) for T_l are those of Algorithm 4 in (2.11). In their translation through (3.7) and (3.8) to the form (3.12) in Algorithm A with T_l^k, they turn into the iterations of Algorithm 3 in (2.10), because

$$\begin{aligned}
&T_l(x^{k+1}, y^{k+1}) + (c_k^{-1} B_k(x^{k+1} - x^k), c_k^{-1} C_k^{-1}(y^{k+1} - y^k)) \\
&\quad = (\partial_x l^k(x^{k+1}, y^{k+1}), \partial_y[-l^k](x^{k+1}, y^{k+1})) \\
&\quad = T_{l^k}(x^{k+1}, y^{k+1}) \ \text{ for } \ l^k(x, y) = l(x, y) + \tfrac{1}{2c_k} ||x - x^k||^2_{B_k} - \tfrac{1}{2c_k} ||y - y^k||^2_{C_k^{-1}}
\end{aligned} \tag{3.19}$$

by way of the rules of subgradient calculus and the gradient formula in (2.1). Moreover the condition $(0, 0) \in T_{l^k}(x^{k+1}, y^{k+1})$ corresponds to (x^{k+1}, y^{k+1}) being a saddle point of $l^k(x, y)$ with respect to minimizing in x and maximizing in y.

The mapping that replaces the general M in (3.1) when $T = T_l$ is $\partial\varphi$, according to (1.6). The corresponding iterations of Algorithm B in (3.13) then have as $M^k(x^{k+1}, u^{k+1})$ the subgradient set $\partial\varphi^k(x^{k+1}, u^{k+1})$ for the convex function

$$\varphi^k(x, u) = \varphi(x, u) - \langle u, y^k \rangle + \frac{1}{2c_k} ||x - x^k||^2_{B_k} + \frac{c_k}{2} ||u - u^k||^2_{C_k}. \tag{3.20}$$

Since the condition $(0, 0) \in \partial\varphi^k(x^{k+1}, u^{k+1})$ characterizes (x^{k+1}, y^{k+1}) as giving the minimum of φ_k, we conclude that the iterations of Algorithm B align with those of Algorithm 2 in (2.9).

Those iterations in exact form also correspond to those of Algorithm 1 in (2.8) for the reason that

$$\min_u \varphi^k(x,u) = \lambda^k(x) := l_{c_k C_k}(x,y^k) + \frac{1}{2c_k}||x-x^k||^2_{B_k}, \tag{3.21}$$

with strong convexity guaranteeing that the minimum is attained. Getting (x^{k+1}, u^{k+1}) to minimize φ^k comes down to getting x^{k+1} to minimize $\lambda^k(x)$ (which is equivalent to solving $0 \in \partial\lambda^k(x^{k+1})$ and then partnering it with a minimizer u^{k+1} of $\varphi^k(x^{k+1}, u)$. This can be identified again with exact iterations of Algorithm B in (3.13) through the rule of convex analysis that

$$v \in \partial\lambda^k(x) \iff (v,0) \in \partial\varphi^k(x,u) \text{ for some } u \text{ that minimizes } \varphi^k(x,\cdot) \tag{3.22}$$

under the strong convexity of φ_k. Thus, the iterations in all four algorithmic formats in Section 2 coincide when executed exactly. □

Proof of Theorems 2.2 and 2.3. When iterations are inexact, the equivalence of the algorithms in the proof just given no longer holds. The issue to be confronted then is how the inexact versions of Algorithms 1, 2, and 3 with their different approaches to stopping criteria fit with the inexact versions of Algorithms A and B. Obviously Algorithm 3 specializes Algorithm A through (3.19). Likewise, Algorithm 2 specializes Algorithm B with M^k being $\partial\varphi^k$ when estimate in (3.15) is invoked. Algorithm 1 comes in as the version of this in which the minimization in x and u is exact in u for each x, so that z in the estimate (3.14) can be taken to be 0. Then the subgradient formula in (3.22) takes over and turns Algorithm B with that stopping criterion into Algorithm 1. □

4 Application to progressive decoupling

The partial-inverse PPA developed in Section 3 for solving the mapping problem (3.2) greatly extends the original partial-inverse method of Spingarn [21] and its enhancement in the decoupling framework of [12] and [13], which was undertaken to advance the theory of decomposition techniques that included Spingarn's way of decomposing augmented Lagrangians in [22]. It may be hard to see this at first glance, however, because changing coordinates may be essential in making the connection.

The problem addressed by the progressive decoupling algorithm needs, first of all, to be posed in the more abstract setting of a finite-dimensional Hilbert space with norm $||\cdot||$ and inner product $\langle\cdot,\cdot\rangle$. The ingredients are a maximal monotone mapping $M : H \rightrightarrows H$ and a subspace $S \subset H$ with orthogonal complement $S^\perp$. It seeks to

$$\text{determine } \bar{z} \in S \text{ and } \bar{w} \in S^\perp \text{ such that } \bar{w} \in M(\bar{z}). \tag{4.1}$$

The background idea is that the elements of S are "linked" by linear relationships that specify S within H, but a solution might be computed through steps in which the linkages somehow get decoupled "progressively." (This terminology goes back to the progressive hedging algorithm in stochastic programming [11], which is a motivational special case for this PPA application.)

This *linkage problem for M and S* can be identified with the problem in (3.2) by expressing H as the direct product $S \times S^\perp$, so that

$$\begin{aligned}&\text{any } z \in H \text{ can be identified uniquely with a pair } (x,u) \in S \times S^\perp, \text{ and}\\ &\text{any } w \in H \text{ can be identified uniquely with a pair } (v,y) \in S \times S^\perp.\end{aligned} \tag{4.2}$$

Then $w \in M(z)$ comes out as $(v,y) \in M(x,u)$, while vectors $\bar{z} \in S$ and $\bar{w} \in S^\perp$ come out in the form of $(\bar{x}, 0)$ and $(0, \bar{y})$, making (4.1) turn into (3.2). And this can go further. Supposing that

S is n-dimensional and $S^\perp$ is m-dimensional, the dimension of H thus being $N = n + m$, we can introduce an n-vector orthonormal basis for S and an m-vector orthonormal basis for $S^\perp$ under which x and u are coordinatized as $(x_1, \dots, x_n)$ and $(u_1, \dots, u_m)$. The combined orthonormal basis for H coordinatizes its elements z that way as $(x_1, \dots, x_n, u_1, \dots, u_m)$, and in those coordinates the norm and inner product of H turn into the canonical norm and inner product of $I\!R^N$. From that angle it's obvious that that algorithm in Section 3 can be applied to solve the problem in (4.1), with the results in Theorems 3.1 and 3.2 then describing its capabilities.

But here's *the very important thing to have in mind.* In the developments of decomposition technology based on (4.1), H may *already* have a form like $I\!R^N = I\!R^{N_1} \times \cdots \times I\!R^{N_q}$ with its vectors z coordinatized as $(z_1, \dots, z_N)$ or the like. *The subspaces S and $S^\perp$ may, however, not fit easily with that portrayal of H.* Passage from the given representation of $z \in H$ as $(z_1, \dots, z_N)$ to a representation as $(x_1, \dots, x_n, u_1, \dots, u_m)$ in the manner above would then require a wholesale change of coordinates, which could well be inconvenient and trouble-making. This suggests trying to pose the partial-inverse PPA in Section 3 as a procedure that can operate directly in the framework of (4.1), without having to pass through such a switch.

An advanced form of the progressive decoupling algorithm in [12] and [13] emerges here from that effort. It supposes only about the subspace S, for practicality, that the projection mappings

$$P_S = \text{projection onto } S, \qquad P_S^\perp = I - P_S = \text{projection onto } S^\perp, \tag{4.3}$$

are readily computable.[7] In terms of them, problem (4.1) seeks to

$$\text{determine } \bar{z}^\# \in H \text{ such that } \bar{w} \in M(\bar{z}) \text{ for } \bar{z} = P_S(\bar{z}^\#) \text{ and } \bar{w} = P_S^\perp(\bar{z}^\#). \tag{4.4}$$

The solution set in the linkage problem is, from that angle, best seen not as a set of pairs $(\bar{z}, \bar{w}) \in S \times S^\perp$, but rather as

$$Z^\# := \{ \text{ set of all } \bar{z}^\# = \bar{z} + \bar{w} \text{ such that } \bar{z} \text{ and } \bar{w} \text{ solve (4.1)} \}. \tag{4.5}$$

In place of the matrix sequences in Section 2, the solution algorithm in this context makes use of self-adjoint, positive-definite linear mappings $J_k \to J_\infty$ from H onto H that preserve S and $S^\perp$,

$$J_k(S) \subset S, \qquad J_k(S^\perp) \subset S^\perp, \tag{4.6}$$

and satisfy for $z \in S$ and $w \in S^\perp$ the norm bounds[8]

$$\begin{array}{l} \beta_k^{-1}||z||_{J_{k-1}} \le ||z||_{J_k} \le \beta_k ||z||_{J_{k-1}} \text{ and } \beta_0^{-1}||z|| \le ||z||_{J_0} \le \beta_0||z|| \\ \gamma_k^{-1}||w||_{J_{k-1}^{-1}} \le ||w||_{J_k^{-1}} \le \gamma_k ||w||_{J_{k-1}^{-1}} \text{ and } \gamma_0^{-1}||w|| \le ||w||_{J_0^{-1}} \le \gamma_0||w||, \\ \quad \text{with } \beta_k \ge 1, \ \Pi_{k=0}^\infty \beta_k < \infty, \text{ and } \gamma_k \ge 1, \ \Pi_{k=0}^\infty \gamma_k < \infty. \end{array} \tag{4.7}$$

It eventually utilizes as well a sequence of error parameters ε_k as in (2.15).

Progressive decoupling algorithm. *From any initial $z^0 \in S$ and $w^0 \in S^\perp$, generate sequences of vectors $z^k \in S$ and $w^k \in S^\perp$ along with auxiliary vectors $\hat{z}^k \in H$ as follows:*

$$\begin{array}{l} \textit{determine } \hat{z}^{k+1} \approx (M^k)^{-1}(0) \textit{ for } M^k(z) = M(z) + J_k[z - z^k] - w^k, \\ \textit{then update with } z^{k+1} = P_S(\hat{z}^{k+1}), \quad w^{k+1} = w^k - J_k P_S^\perp(\hat{z}^{k+1}). \end{array} \tag{4.8}$$

[7]If either the projection on S or the projection on $S^\perp$ is readily computable, then so too is the other, according to the complementarity in (4.3).

[8]Of course, it could be assumed for simplicity that these bounds hold for all z and w in H, but only this much is needed in our analysis.

Observe that w^{k+1} inherits membership in $S^\perp$ from w^k because $P_S^\perp(\hat{z}^{k+1})$ lies in $S^\perp$, hence $J_k(P_S^\perp(\hat{z}^{k+1}) \in S^\perp$ by (4.6). Also, since M is maximal monotone, and J_k is self-adjoint and positive-definite, thus maximal *strongly* monotone, the mapping M^k in (4.8) is maximal *strongly* monotone and therefore has a single-valued, Lipschitz continuous inverse [19, 12.55]. Thus, $(M^k)^{-1}(0)$ in (4.8) stands for a single vector that is being approximated by $\hat{z}^{k+1}$. It can be described also as the unique z such that $w^k \in M(z) + J_k[z - z^k]$.

The method of partial inverses of Spingarn in [21] is the case of this algorithm in which $J_k \equiv I$. It started the thinking on these lines that led to the version of progressive decoupling in [12] and [13]. That extension allowed $J_k \equiv rI$ for a fixed $r > 0$ and provided insights into how the choice of r might quantitatively affect the way z^k and w^k converge to $\bar{z}$ and $\bar{w}$ solving (4.1). But the steps had to be computed exactly, without the approximation indicated in (4.8). Relaxed computations and the introduction of varying J_k are the new features here. Even $J_k = r_k I$ for a sequence of values r_k was previously out of grasp.

Partial inverses of mappings from H to H via the choice of the subspace S can be defined with respect to the representations (4.2), which we can better think of now in the notation

$$z = z_S + z_{S^\perp}, \quad \text{where } \ z_S = P_S(z) \ \text{ and } \ z_{S^\perp} = P_S^\perp(z). \tag{4.9}$$

The partial inverse of the mapping M in this context is the mapping we'll denote here by $M^\#$. It is defined graphically by

$$(z', w') \in \operatorname{gph} M^\# \quad \Longleftrightarrow \quad \begin{cases} \exists\, (z, w) \in \operatorname{gph} M \ \text{ such that} \\ z' = z_S + w_{S^\perp} \ \text{ and } \ w' = w_S + z_{S^\perp} \end{cases} \tag{4.10}$$

in an echo of the relationship between M and T in (3.1) as interpreted via (4.2), with $M^\#$ once more maximal monotone. Note that $z^\# \in Z^\#$ corresponds then to $0 \in M^\#(z^\#)$; thus, the solution set $Z^\#$ in (4.5) is the closed convex set $M^{\#\,-1}(0)$. In these terms,

$$\begin{array}{l} \textit{solving (4.1) as interpreted in (4.4) means finding } \bar{z}^\# \textit{ such that} \\ 0 \in M^\#(\bar{z}^\#), \ \textit{ and then setting } \ \bar{z} = P_S(\bar{z}^\#) \ \textit{ and } \ \bar{w} = P_S^\perp(\bar{z}^\#). \end{array} \tag{4.11}$$

Such $\bar{z}^\# = \bar{z} + \bar{w}$ is to be approximated by iterates $z^{\#k} = z^k + w^k$ with $z^k \in S$ and $w^k \in S^\perp$.

The partial inverses $J_k^\#$ of the mappings J_k, defined likewise the mode of (4.10), will be needed as well. In fact, the targeted procedure in (4.8), when viewed in the light of (4.11), takes the form of generating a sequence of iterates $z^{\#k}$ by

$$z^{\#\,k+1} \approx (I + J_k^{\#-1} M^\#)^{-1}(z^{\#k}), \ \text{ with } \ z^k = P_S(z^{\#k}), \ w^k = P_S^\perp(z^{\#k}). \tag{4.12}$$

Like J_k, the linear mappings $J_k^\#$ are self-adjoint and positive-definite, and so too are the mappings J_k^{-1} and $J_k^{\#-1}$. The inner products and norms associated with all of them in the pattern of

$$\langle z', z \rangle_{J_k} := \langle z', J_k z \rangle, \qquad ||z||_{J_k} = \sqrt{\langle z, z \rangle_{J_k}}, \tag{4.13}$$

can serve, through the interpretation of (4.8) as (4.12), in translating the variable-metric results of Section 3 to the framework of progressive decoupling. For these norms the many avenues for estimation in (2.2) are equally available. Working with them is helped by the relationship between J_k and S in (4.3), which they all inherit. Only the restrictions of J_k to the subspaces S and $S^\perp$, and the inverses of those restrictions, really matter. Thus, in the notation (4.9),

$$||z||^2_{J_k} = ||z_S||^2_{J_k} + ||z_{S^\perp}||^2_{J_k}, \qquad ||z||^2_{J_k^{-1}} = ||z_S||^2_{J_k^{-1}} + ||z_{S^\perp}||^2_{J_k^{-1}}, \tag{4.14}$$

and in consequence

$$||z||^2_{J_k^{\#}} = ||z_S||^2_{J_k} + ||z_{S^\perp}||^2_{J_k^{-1}}, \qquad ||z||^2_{J_k^{\#-1}} = ||z_S||^2_{J_k^{-1}} + ||z_{S^\perp}||^2_{J_k}. \tag{4.15}$$

This is especially welcome because, with $J_k^{\#}$ doing the job in (4.12) that was assigned before to the matrix D_k, the norm associated with it will be a critical ingredient in the stopping criteria we come up with. Particularly important in that respect is that, in the alternative view of the progressive decoupling algorithm (4.8) in (4.12), we have

$$||z^{\#\,k+1} - z^{\#\,k}||_{J_k^{\#}} = \sqrt{||z^{k+1} - z^k||^2_{J_k} + ||w^{k+1} - w^k||^2_{J_k^{-1}}}. \tag{4.16}$$

The criteria for getting convergence of $z^k + w^k$ for some $\bar{x} + \bar{w} \in Z^{\#}$, despite inexactness in (4.12), will benefit from this.

In support of linear convergence, we will need to appeal to the *metric subregularity* of the mapping $M^{\#}$ at the solution element $\bar{z}^{\#} = \bar{z} + \bar{w}$ reached in the limit. The presence or absence of that metric subregularity is independent of the norm used in describing it, but the modulus of property does depend on it. Our focus will be on $||\cdot||_{J_\infty^{\#}}$, because that modulus will enter the achieved rate of linear convergence. The formulation of metric subregularity in this case is that

$$\begin{array}{l} \exists\, a \in (0,\infty) \ \text{ such that, for } z^{\#} \text{ and } w^{\#} \text{ near to } \bar{z}^{\#} \text{ and } 0, \\ w^{\#} \in M^{\#}(z^{\#}) \quad \Longrightarrow \quad \text{dist}_{J_\infty^{\#}}\Big(z^{\#}, M^{\#-1}(0)\Big) \le a||w^{\#}||_{J_\infty^{\#-1}}, \end{array} \tag{4.17}$$

with its modulus being

$$\text{subreg}_{J_\infty^{\#}}\Big(M^{\#}; \bar{z}^{\#} \,|\, 0\Big) := \liminf \text{ of the } a \text{ values in (3.5) as neighborhoods dwindle.} \tag{4.18}$$

Theorem 4.1 (global convergence in progressive decoupling)**.** *The vectors z^k and w^k generated from any initial $z^0 \in S$ and $w^0 \in S^\perp$ by the steps in (4.8) will converge to some particular $\bar{z}$ and $\bar{w}$ solving problem (4.1) if the stopping criterion on $\widehat{z}^{k+1}$ is*

$$2\,\text{dist}_{J_k^{-1}}(0, M^k(\widehat{z}^{k+1})) \le \varepsilon_k. \tag{4.19}$$

If the stopping criterion is tightened by also requiring

$$2\,\text{dist}_{J_k^{-1}}(0, M^k(\widehat{z}^{k+1})) \le \varepsilon_k \sqrt{||z^{k+1} - z^k||^2_{J_k} + ||w^{k+1} - w^k||^2_{J_k^{-1}}}, \tag{4.20}$$

and the metric subregularity property in (4.17) holds for $\bar{z}^{\#} = \bar{z} + \bar{w}$, then the distances

$$d^k = \min\left\{ \sqrt{||z^k - \bar{z}'||^2_{J_k} + ||w^k - \bar{w}'||^2_{J_k^{-1}}} \;\middle|\; \bar{z}' + \bar{w}' \in Z^{\#} \right\} \tag{4.21}$$

converge Q-linearly to 0 at the rate

$$r = \frac{1}{\sqrt{1 + a_\infty^{-2}}} = \frac{a_\infty}{\sqrt{1 + a_\infty^2}} \ \ \textit{with} \ \ a_\infty = \text{subreg}_{D_\infty}\Big(M^{\#}; \bar{z}^{\#} \,\Big|\, 0\Big). \tag{4.22}$$

Proof. This just translates the results in Theorem 3.2 about Algorithm B in the case of $c_k \equiv 1$ into the notation of the algorithm in (4.8) through the notational shifts based on (4.2) that we have already been utilizing. In this translation, J_k corresponds to the matrix E_k. Its restrictions to S and $S^\perp$ corresponds to B_k and C_k, for which our assumption (4.7) with respect to $z \in S$ and $w \in S^\perp$ turns into the conditions imposed in (2.6). At the same time, $J_k^\#$ corresponds to D_k, and all justification for all the claims falls into place from Theorem 3.2 with no need for additional arguments. □

Note that the factor 2 on the left side of (4.19) and (4.20) could simply be absorbed into the choice of ε_k, but we have retained it for transparency in the translation.

The presence of both J_k and J_k^{-1} in the distances (4.21) reveals something important about the character of linear convergence in progressive decoupling. This is especially clear when the solution set is a singleton, so that

$$d^k = \sqrt{||z^k - \bar{z}||^2_{J_k} + ||w^k - \bar{w}||^2_{J_k^{-1}}}, \tag{4.23}$$

and even more so when simply $J_k = r_k I$ with $r_k \to r_\infty$ as $J_k \to J_\infty$ (which excludes having $r_\infty = \infty$), making

$$d^k = \sqrt{r_\infty^2 |z^k - \bar{z}|^2 + r_\infty^{-2} |w^k - \bar{w}|^2}. \tag{4.24}$$

There is an inevitable trade-off between the primal convergence of z^k to $\bar{z}$ and the dual convergence of w^k to $\bar{w}$. A higher value of r_∞ places emphasis on the z^k convergence while a lower value places it on the w^k convergence. Such a tradoff was already perceived in [12, (2.4)] when the r_k sequence was constant, but until now only convergence properties falling short of the one in Theorem 4.1 were pinned down.

5 Subregularity assurances

The results on linear convergence in Theorems 2.2 and 4.1 rely on the property of metric subregularity holding at a solution point reached in the limit but unknown in advance. How realistic is that supposition? We provide now a special but important case where it is automatically satisfied. We go on to explain why it can anyway be regarded, at least from one reasonable perspective, as *typically* satisfied.

A convex function is called *piecewise linear-quadratic* if its effective domain is the union of a finite collection of polyhedral convex sets on which the function is linear-quadratic, i.e., can be expressed by a polynomial of degree no more than 2 [19, 10.20]. A set-valued mapping is called *piecewise polyhedral* if its graph is the union of a finite collection of polyhedral convex sets [19, 9.57].

Theorem 5.1 (guaranteed linear convergence to the solution set)**.**

(a) *If the convex function φ in problem (P) is piecewise linear-quadratic, then the metric subregularity property supporting linear convergence in Theorem 2.2 is sure to be fulfilled.*

(b) *If the mapping M in the linkage problem (4.1) is piecewise polyhedral, then the metric subregularity property supporting linear convergence in Theorem 4.1 is sure to be fulfilled.*

Proof. Starting with (b), we note that in passing from M to its partial inverse $M^\#$ in (4.10), piecewise polyhedrality is maintained, since only a sort of graphical reflection is involved. Theorem 4.1 rests on applying the proximal point algorithm to $M^\#$ to solve the reformulated problem in (4.11) and requires the metric subregularity of $M^\#$ at $\bar{z}^\#$ for $0 \in M^\#(\bar{z}^\#)$. That is univeral for a piecewise polyhedral mapping, since it is equivalent to calmness properties of the inverse of $M^\#$ by [1, 3H.3], that likewise being piecewise polyhedral, and such calmness is universal for such a mapping by [19, 9.57]. But the linkage problem (4.10) subsumes the optimization problem (P) as the case where $M = \partial\varphi$, as explained in Section 4. Furthermore, φ is piecewise linear-quadratic if and only if its subgradient mapping is piecewise polyhedral [19, 12.30]. That way, (b) establishes (a). □

The case of problem (P) being piecewise linear-quadratic covers, in particular, problems of piecewise linear-quadratic programming as formulated in [19, p. 506].

Although the needed subregularity is not always assured, it is anyway generic from a perspective in which the given problem is embedded in a parameterized family of problems. In the optimization framework, these problems have the form

$$(P(\bar{v},\bar{u})) \qquad \text{for given } (\bar{v},\bar{u}), \text{ minimize } \varphi(x,u) - \bar{v}{\cdot}x \text{ subject to } u = \bar{u},$$

with the original (P) corresponding to $(\bar{v},\bar{u}) = (0,0)$. Obviously the PMM algorithm in its different forms in Section 2 can be applied to solve $(P(\bar{v},\bar{u}))$ with only some changes of notation. We can ask then about typical behavior for the family, instead of just behavior in the single instance giving (P).

To get a better picture of this, consider along with the problems $(P(\bar{v},\bar{u}))$ the similar-looking dual problems

$$(D(\bar{v},\bar{u})) \qquad \text{for given } (\bar{v},\bar{u}), \text{ maximize } -\varphi^*(v,y) + \bar{u}{\cdot}y \text{ subject to } v = \bar{v},$$

and the equivalent of optimality conditions for them that generalizes (1.7):

$$(\bar{v},\bar{y}) \in \partial\varphi(\bar{x},\bar{u}) \iff (\bar{v},\bar{u}) \in T_l(\bar{x},\bar{y}) \iff (\bar{x},\bar{u}) \in \partial\varphi^*(\bar{v},\bar{y}). \tag{5.1}$$

The set Z in (1.8) is accordingly replaced by

$$Z(\bar{v},\bar{u}) := \{\, (\bar{x},\bar{y}) \text{ satisfying (5.1)} \,\} = T_l^{-1}(\bar{v},\bar{u}), \tag{5.2}$$

which, if nonempty, is the product of the set of solutions $\bar{x}$ to $(P(\bar{v},\bar{u}))$ and the set of solutions $\bar{y}$ to $(D(\bar{v},\bar{u}))$. In terms of the nonempty convex sets

$$U = \{\, u \in I\!R^m \mid \exists\, x,\ \varphi(x,u) < \infty\}, \qquad V = \{\, v \in I\!R^n \mid \exists\, y,\ \varphi^*(v,y) < \infty\}, \tag{5.3}$$

it is known from [19, 11.40] that

$$Z(\bar{v},\bar{u}) \text{ is nonempty and bounded} \iff (\bar{v},\bar{u}) \in (\operatorname{int} V) \times (\operatorname{int} U). \tag{5.4}$$

Theorem 5.2 (generic linear convergence in optimization)**.** *Suppose that the convex sets U and V in (5.3) are full-dimensional, or equivalently that, for at least one pair $(\bar{v},\bar{u})$, the solution set $Z(\bar{v},\bar{u})$ in (5.2) is nonempty and bounded.*

Then for almost all $(\bar{v},\bar{u}) \in V \times U$, the solution set $Z(\bar{v},\bar{u})$ will consist of a single pair $(\bar{x},\bar{y})$, and T_l will be metrically subregular there with respect to $(\bar{v},\bar{u})$ as demanded by Theorem 2.3. The sequence of pairs (x^k,y^k) generated by the proximal method of multipliers in solving $(P(\bar{v},\bar{u}))$ and $(D(\bar{v},\bar{u}))$ will therefore converge Q-linearly to that solution $(\bar{x},\bar{y})$ at the rate

$$r = \frac{1}{\sqrt{1+(c_\infty/a_\infty)^2}} \quad \text{with } a_\infty = \operatorname{subreg}_{D_\infty}\Big(T_l;(\bar{x},\bar{y})\,|\,(\bar{v},\bar{u})\Big). \tag{5.6}$$

Proof. The effective range set $\operatorname{rge} T_l$ for the mapping T_l, which is the same as the set of $(\bar{v},\bar{u})$ such that $T_l^{-1}(\bar{v},\bar{u}) \neq \emptyset$, satisfies

$$\operatorname{int}(V \times U) \subset \operatorname{rge} T_l \subset V \times U \tag{5.7}$$

[19, 12.41]. But according to a recently demonstrated property of maximal monotone mappings, for almost every $(\bar{v},\bar{u})$ belonging to $\operatorname{int}(\operatorname{rge} T_l)$ in the sense of Lebesgue measure, there is a unique $(\bar{x},\bar{y}) \in T_l^{-1}(\bar{v},\bar{u})$, and T_l is metrically subregular at $(\bar{x},\bar{y})$ for $(\bar{v},\bar{u})$. That's all we need to justify the claim. □

A parallel result holds for the linkage problem (4.1), which can likewise be viewed with such a $(\bar{v},\bar{u})$ parameterization in the pattern of

$$\begin{array}{l}\text{given } \bar{v} \in S \text{ and } \bar{u} \in S^{\perp}, \text{ determine } \bar{z} \in S \text{ and } \bar{w} \in S^{\perp} \\ \text{such that } \ \bar{w} \in M_{\bar{v},\bar{u}}(\bar{z}), \ \text{ where } \ M_{\bar{v},\bar{u}}(z) = M(z+\bar{u}) - \bar{v}\end{array} \tag{5.8}$$

with associated solution sets

$$Z_{\bar{v},\bar{u}} = \{\, (\bar{z},\bar{w}) \text{ solving}(5.8) \,\}. \tag{5.9}$$

In this case we are looking at a parameterized family of maximal monotone mappings and can ask about generic behavior of the progressive decoupling algorithm in solving the linkage problems associated with them for the fixed subspace S. Interest centers on the set

$$\Omega = \{\, (\bar{v},\bar{u}) \in S \times S^{\perp} \mid Z_{\bar{v},\bar{u}} \neq \emptyset \}. \tag{5.10}$$

Theorem 5.3 (generic linear convergence in progressive decoupling)**.** *Suppose the set Ω in (5.10) has nonempty interior; this corresponds to its full-dimensionality and holds if and only if $Z_{\bar{v},\bar{u}}$ in (5.9) is nonempty and bounded for at least one pair $(\bar{v},\bar{u})$, in which case the set of those pairs constitutes that interior.*

Then for almost all $(\bar{v},\bar{u}) \in \Omega$, the solution set $Z_{\bar{v},\bar{u}}$ will consist of a single pair $(\bar{z},\bar{w})$, and $M^{\#}_{\bar{v},\bar{u}}$ will have the metric subregularity property in (4.17) with modulus in (4.18) as demanded by Theorem 4.1. The sequence of pairs (z^k,w^k) generated by the proximal method of multipliers in solving the corresponding problem in (5.8) will therefore converge Q-linearly to that solution $(\bar{z},\bar{w})$ at the rate

$$r = \frac{1}{\sqrt{1+a_\infty^{-2}}} = \frac{a_\infty}{\sqrt{1+a_\infty^{2}}}. \ \textit{with} \ \ a_\infty = \operatorname{subreg}_{J^{\#}_{\infty}}\Big(M^{\#}; \bar{z}+\bar{w} \,|\, \bar{v}+\bar{u}\Big). \tag{5.11}$$

Proof. By definition, $(\bar{z},\bar{w}) \in Z_{\bar{v},\bar{u}}$ if and only if $\bar{v}+\bar{w} \in M(\bar{z}+\bar{u})$, which is equivalent to $\bar{v}+\bar{u} \in M^{\#}(\bar{z}+\bar{w})$ in (4.10). Thus, $(\bar{v},\bar{u}) \in \Omega$ corresponds to $M^{\#-1}(\bar{v}+\bar{u}) \neq \emptyset$, i.e., $\bar{v}+\bar{u} \in \operatorname{dom} M^{\#-1}$, that being a nearly convex set by maximal monotonicity [19, 12.41]. The interior of $\operatorname{dom} M^{\#-1}$ consists of $\bar{v}+\bar{u}$ such that $M^{\#-1}(\bar{v}+\bar{u})$ is both nonempty and bounded. It corresponds to the interior of Ω (since $\bar{v} \in S$ and $\bar{u} \in S^{\perp}$). The existence of at least one such pair $(\bar{v},\bar{u})$, or the full dimensionality of Ω iand hence that of $\operatorname{dom} M^{\#-1}$, makes that interior be nonempty. Then, as in the proof of Theorem 5.2, we know from [17, Theorem 1] that for almost all elements of $\operatorname{int}\Omega$, which by near convexity is the same as saying for almost all pairs $(\bar{v},\bar{u})$ in Ω itself, $Z_{\bar{v},\bar{u}}$ is a singleton $\{(\bar{z},\bar{w})\}$ and $M^{\#}$ is metrically subregular at $\bar{z}+\bar{w}$ with respect to having $\bar{v}+\bar{u} \in M^{\#}(\bar{z}+\bar{w})$, as claimed. □

Data availability: No data was generated or analyzed.
Competing interests: None.

References

[1] Dontchev, A.D., Rockafellar, R.T., *Implicit Functions and Solution Mappings*, Springer Verlag, second edition: 2014.

[2] Ismailov, A.F., and Solodov, M.V., "Perturbed augmented Lagrangian method framework with applications to proximal and smoothed variants," *J. Optimization Theory Appl.* **193** (2022), 491–522.

[3] Hestenes, M., "Multiplier and gradient methods," *J. Optimization Theory Appl.* **4** (1969), 303–320.

[4] Luque, F.J., "Asymptotic convergence analysis of the proximal point algorithm," *SIAM J. Control Opt.* **22** (1984), 277–293.

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

[6] Powell, M. J. D., "A method for nonlinear optimization in minimization problems," in *Optimization* (R. Fletcher, ed.), Academic Press, 1969, 283–298.

[7] Rockafellar, R. T., *Convex Analysis*, Princeton University Press, 1970.

[8] Rockafellar, R. T., *Conjugate Duality and Optimization*, No. 16 in Conference Board of MathṠciences Series, SIAM Publications, 1974.

[9] Rockafellar, R. T., "Monotone operators and the proximal point algorithm." *SIAM J. Control Opt.* **14** (1976), 877–898.

[10] Rockafellar, R. T., "Augmented Lagrangians and applications of the proximal point algorithm in convex programming." *Math. of Operations Research* **1** (1976), 97–116.

[11] Rockafellar, R. T., and Wets, R. J-B, "Scenarios and policy aggregation in optimization under uncertainty." *Mathematics of Operations Research* **16** (1991), 119–147.

[12] Rockafellar, R. T., "Progressive decoupling of linkages in monotone variational inequalities and convex optimization," *Proceedings of the 10th International Conference on Nonlinear Analysis and Convex Analysis* (Chitose, Japan, 2017), M. Hojo, M. Hoshino, W. Takahashi (eds.), Yokohama Publishers, 2019, 271–291.

[13] Rockafellar, R. T., "Progressive decoupling of linkages in optimization and variational inequalities with elicitable convexity or monotonicity," *Set-Valued and Variational Analysis* **27** (2019), 863–893.

[14] Rockafellar, R. T., "Augmented Lagrangians and hidden convexity in sufficient conditions for local optimality," *Math. Programming* **198** (2023), 159–194 (published online January 2022; doi.org/10.1007/s10107-022-01768-w).

[15] Rockafellar, R. T., "Advances in convergence and scope of the proximal point algorithm," *J. Nonlinear and Convex Analysis* **22** (2021), 2347–2375.

[16] Rockafellar, R. T., "Convergence of augmented Lagrangian methods in extensions beyond nonlinear programming," *Mathematical Programming* **199** (2023), 375–420 (published online June 2022; DOI 10.1007/s10107-022-01832-5).

[17] Rockafellar R. T., "Metric regularity properties of monotone mappings," *Mathematica Serdica* (memorial issue for Asen Dontchev), accepted.

[18] Rockafellar R. T., "Generic linear convergence through metric subregularity in a variable-metric extension of the proximal point algorithm," *Computational Optimization and Applications*, accepted.

[19] Rockafellar, R. T., and Wets, R. J-B, *Variational Analysis*, No. 317 in the series *Grundlehren der Mathematischen Wissenschaften*, Springer-Verlag, 1997.

[20] Sefi,, R., and Teboulle, M., "Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization,"*SIAM J. Optimization* **24** (2014), 269–297.

[21] Spingarn, J., "Partial inverse of a monotone operator," *Applied Mathematics and Optimization* **10** (1983), 247–265.

[22] Spingarn, J., "Applications of the method of partial inverses to convex programming: decomposition," *Mathematical Programming* **32** (1985), 199-121.