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), pages 271–291
[Here with some subsequent fixes of miswritten formulas]

PROGRESSIVE DECOUPLING OF LINKAGES IN MONOTONE VARIATIONAL INEQUALITIES AND CONVEX OPTIMIZATION

R. Tyrrell Rockafellar[1]

Abstract

A procedure called the progressive decoupling algorithm is introduced for solving variational inequality problems with monotonicity in which the iterative relaxation of linkages can simplify computations. It derives from the proximal point algorithm in a manner similar to Spingarn's method of partial inverses but deals differently with parameters and is able thereby to fit the format of the progressive hedging algorithm in convex stochastic programming and its recent extension to monotone stochastic variational inequalities. Applications to decomposable structure and splitting structure are provided. Connections with problems of convex optimization are laid out with Lagrangian variational inequalities and iterations involving augmented Lagrangians as a special feature.

Keywords: *variational inequalities, optimality conditions, linkage structure, progressive decoupling, progressive hedging, problem decomposition, splitting, maximal monotone mappings, proximal point algorithm, method of partial inverses, augmented Lagrangians, proximal method of multipliers, ADMM*

December 22, 2017

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

1 Introduction

Generalized equation problems in the form of variational inequalities offer a versatile way of modeling optimality and equilibrium in complex situations. Many large-scale applications in that mode involve a linkage structure which, if it could be decoupled, would enable computation of a solution to proceed more easily. This is addressed here on a fundamental level where the linkage can progressively be decoupled through "multipliers" which get updated in a manner reminiscent of augmented Lagrangian techniques in optimization. The procedure is motivated in form and spirit by the *progressive hedging algorithm* in convex stochastic optimization [14] and its recent extension to the decomposition of multistage stochastic variational inequality problems [13]. It is intimately related to Spingarn's *method of partial inverses* [17] for maximal monotone mappings, which in turn came out of the *proximal point algorithm* [9], but differs in a crucial change of variables in the derivation, which adds flexibility. It specializes to applications in problem decomposition and splitting and, in the setting of convex optimization, can be articulated with minimization subproblems. Extensions to nonconvex optimization and nonmonotone variational inequalites will be offered in [12].

Let H be a (real) Hilbert space with inner product $\langle x, y\rangle$. In particular H could be $I\!R^n$ with $\langle x, y\rangle = x{\cdot}y$, but there are finite-dimensional situations where a different inner product would be advantageous, such as in a stochastic framework. The *variational inequality* problem in H for a nonempty closed convex set $K \subset H$ and a mapping $F : K \to H$ aims to

$$\text{find } \bar{x} \text{ such that } \; -F(\bar{x}) \in N_K(\bar{x}), \tag{1.1}$$

where $N_K(x)$ is the normal cone to K at x:

$$v \in N_K(x) \quad\Longleftrightarrow\quad x \in K \;\text{ and }\; \langle v, x' - x\rangle \le 0 \;\; \forall x' \in K. \tag{1.2}$$

Our focus is on exploiting special structure in problem (1.1) that may lead to "decoupling." To that end we take K to be an intersection $C \cap S \neq 0$ in which C captures "basic" requirements while C embodies "linkages." (Examples of linkages will be provided shortly.) We assume throughout that

$$C \text{ is a nonempty closed convex subset of } H \text{ and } S \text{ is a (closed) linear subspace of } H. \tag{1.3}$$

In support of the workability of this structure, we suppose that the projection mapping P onto S,

$$P : H \to H \;\text{ with }\; P(x) = \;\text{[the point in } S \text{ nearest to } x], \tag{1.4}$$

is relatively convenient to execute numerically. The same executability then holds for the projection mapping $P^{\perp}$ onto the orthogonal complement $S^{\perp}$ of S,

$$P^{\perp} : H \to H \;\text{ with }\; P^{\perp}(x) = \text{[the point in } S^{\perp} \text{ nearest to } x] = [I - P](x). \tag{1.5}$$

We assume further that, instead of just a mapping F on $K = C \cap S$, we have a *continuous* mapping $F : C \to H$ that is *monotone* relative to C, which means

$$\langle F(x') - F(x), x' - x\rangle \ge 0 \;\text{ for all }\; x, x' \in C. \tag{1.6}$$

Linkage variational inequality problems: coupled and decoupled. *The linkage problem in coupled form with respect to S, C and $F : C \to H$ is to*

$$\textit{find } \bar{x} \textit{ such that } \; -F(\bar{x}) \in N_{C\cap S}(\bar{x}), \tag{1.7}$$

whereas the associated linkage problem in decoupled form is to

$$\textit{find } \bar{x} \in S \textit{ such that } -\overline{F}(\bar{x}) \in N_C(\bar{x}) \textit{ for } \overline{F}(x) = F(x) - \bar{y} \textit{ in the case of some } \bar{y} \in S^{\perp}. \tag{1.8}$$

The tight connection between these two problem formulations, amounting to a virtual equivalence, is described by a *decoupling principle*:

$$\text{any solution } \bar{x} \text{ to (1.8) solves (1.7), and conversely under a constraint qualification.} \tag{1.9}$$

This relationship is based on the calculus of normal cones to convex sets through the general rule that $N_{C\cap S}(x) \supset N_C(x) + N_S(x) = \{\, v + y \,|\, v \in N_C(x),\ y \in N_S(x)\}$, with equality holding under various conditions (known as constraint qualifications),[2] along with a fact about subspaces:

$$y \in N_S(x) \iff x \in S,\ y \in S^{\perp} \iff x \in N_{S^{\perp}}(y). \tag{1.10}$$

For the existence of a solution to (1.7) an easy criterion is the compactness of C. If F is *strongly* monotone relative to C, meaning that

$$\exists\, \mu > 0 \ \text{ with } \ \langle F(x') - F(x), x' - x\rangle \geq \mu ||x' - x||^2 \ \text{ for all } \ x, x' \in C, \tag{1.11}$$

existence and *uniqueness* of a solution to (1.7) are assured and likewise for $-\bar{F}(\bar{x}) \in N_C(\bar{x})$ in (1.8), even for unbounded C.

According to the decoupling principle (1.9), any means we might come up with for solving the decoupled problem (1.8) will serve also for solving the coupled problem (1.7). That may offer advantages because the decoupled problem, although tacitly requiring $\bar{y}$ to be determined in tandem with $\bar{x}$, has a variational inequality that is much simpler in being just over C instead of $C \cap S$. An even bigger attraction is the path it offers to avoid having to deal head-on with the condition $\bar{x} \in S$. How so? Suppose we had $\bar{y}$ such that a solution to (1.8) is sure to exist, and suppose F is such that there can only be one $\bar{x}$ satisfying $-F(\bar{x}) \in N_C(\bar{x})$, as under strong monotonicity of F. Then that $\bar{x}$ would have to belong to S. The condition $\bar{x} \in S$ would have been taken care of automatically!

This leaves a lot of ifs, but in fact we will be occupied here with presenting a procedure in which in iteration $\nu = 1, 2, \ldots$, there are elements $x^\nu \in S$, $y^\nu \in S^\perp$, and an associated modification F^ν of F such that the variational inequality $-F^\nu(x) \in N_C(x)$ has a unique solution due to strong monotonicity; calculating that solution leads to updates from x^ν and y^ν to $x^{\nu+1} \in S$ and $y^{\nu+1} \in S^\perp$ which guarantee convergence to some $\bar{x}$ and $\bar{y}$ satisfying (1.8) (as long as at least one such pair does exist). This method, which we call the *progressive decoupling algorithm*, will be laid out in Section 2.

For now, to help with a better understanding of what decoupling can mean in practice, we examine two kinds of problem structure.

Decoupling to exploit decomposable structure. Let $H = H_1 \times \cdots \times H_q$ for Hilbert spaces H_j and likewise $C = C_1 \times \cdots \times C_q$ with $C_j \subset H_j$. Let the mapping F take $(x_1, \ldots, x_q) \in C_1 \times \cdots \times C_q$ to $(F_1(x_1), \ldots, F_q(x_q))$ for mappings $F_j : C_j \to H_j$. Then in problem (1.7) with respect to a subspace $S \subset H$, the components of the vector $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_q)$ are linked by the requirement that $\bar{x} \in S$, but in problem (1.8), which incorporates a vector $\bar{y} = (\bar{y}_1, \ldots, \bar{y}_q) \in S^\perp$, the overall variational inequality decomposes into separate variational inequalities in the components:

$$\text{for each } j, \text{ find } \bar{x}_j \text{ such that } -\bar{F}_j(\bar{x}_j) \in N_{C_j}(\bar{x}_j), \text{ where } \bar{F}_j(x_j) = F_j(x_j) - \bar{y}_j. \tag{1.12}$$

[2]It suffices for instance to have $S \cap \operatorname{int} C \neq$ or just $0 \in \operatorname{int}[C + S]$. When H is finite-dimensional, interiors can be replaced by relative interiors. Moreover the relative interiority criterion can be dropped in the case of C being polyhedral.

Multistage stochastic variational inequalities in the form developed recently in [16] exhibit this structure and motivate for our efforts. In that setting each index j corresponds to a “scenario” in the evolution of information on which decisions can be based. The scenarios in a multistage discrete setting can be represented by a “scenario tree” that branches in response to the resolution of uncertainties as time goes on. If the scenario eventually to be followed could be known at the outset, the decisions could be optimized deterministically, which would be relatively easy. In reality, though, any decision associated with stage k can only be based on information already available in stage k and not on any information yet to come. In other words, decisions associated with two different scenarios have to coincide until those scenarios branch away from each other. That constraint, known as *nonantipativity*, constitutes a linkage structure expressible by a subspace S. In (1.12) the modification by vectors $\bar{y}_j$ allows the decisions to be determined by solving deterministic problems for the individual scenarios after all.

Decoupling to exploit splitting structure. The term “splitting” is typically applied to approaches to solving a generalized equation like $0 \in T_1(w) + \cdots + T_q(w)$ in a space H_0 by iteratively working with the mappings T_j individually. With our focus on variational inequalities, we pose it as trying to

$$\text{find } \bar{w} \text{ such that } -[F_1(\bar{w}) + \cdots + F_q(\bar{w})] \in N_{C_1 \cap \cdots \cap C_q}(\bar{w}) \tag{1.13}$$

for closed convex subsets $C_j \subset H_0$ and continuous monotone mappings $F_j : C_j \to H_0$. We then rely on the fact that [3]

$$\begin{array}{l} N_{C_1 \cap \cdots \cap C_q}(w) \ \subset \ N_{C_1}(w) + \cdots + N_{C_q}(w) \text{ always, and} \\ \text{equality holds } \forall w \text{ under a constraint qualification.} \end{array} \tag{1.14}$$

Equality in this relationship allows the problem in (1.13) to be translated into the framework of decomposable structure by taking $H_1 = \cdots = H_q = H_0$ and selecting the complementary pair of subspaces to be

$$\begin{array}{l} S = \{\, x = (x_1, \dots, x_q) \,|\, \exists\, w \in H_0 \text{ such that } x_1 = \cdots = x_q = w\}, \\ S^\perp = \{\, (y = (y_1, \dots, y_q) \,|\, y_j \in H_0,\ y_1 + \cdots + y_q = 0\}. \end{array} \tag{1.15}$$

Then the problem in (1.13) can be identified with the coupled variational inequality (1.7), whereas the corresponding decoupled variational inequality (1.8) comes out as

$$\begin{array}{c} \text{find } \bar{w} \text{ such that } \exists \bar{y}_j \text{ with } \bar{y}_1 + \cdots \bar{y}_q = 0 \text{ yielding} \\ -\overline{F}_j(\bar{w}) \in N_{C_j}(\bar{w}) \text{ for } j = 1, \dots, q \text{ with } \overline{F}_j(w) = F_j(w) - \bar{y}_j. \end{array} \tag{1.16}$$

The attractive idea again, by way of the decoupling principle, is the existence of vectors $\bar{y}_j$ such that the variational inequalities

$$\text{find } \bar{x}_j \text{ such that } -\bar{F}(\bar{x}_j) \in N_{C_j}(\bar{x}_j) \text{ for } j = 1, \dots, q \tag{1.17}$$

in (1.16) are able to have a common solution $\bar{x}_1 = \cdots = \bar{x}_q$. When each F_j is strongly monotone, so that uniqueness of solutions in (1.17) is assured, these variational inequalities can be solved independently and yet all give the same unique answer, which is the desired $\bar{w}$.

[3] In finite dimensions a useful constraint qualification is the nonemptiness of the intersection of the relative interiors of the sets C_j. A constraint qualification that works in infinite as well as finite dimensions is the existence of a point in one of the sets that belongs to the interior of every one of the other sets.

2 Algorithm derivation and validation

An approach due to Spingarn [17], his method of partial inverses, is already known for solving problems of the kind

$$\text{find } \bar{x} \in S \text{ such that there exists } \bar{y} \in S^{\perp} \text{ with } \bar{y} \in T(\bar{x}) \tag{2.1}$$

for a maximal monotone mapping $T : H \rightrightarrows H$. Our problem (1.8) is the case of this in which $T = F + N_C$, which is indeed maximal monotone under our assumptions on C and F. Spingarn's method, like the algorithm we are about to describe, involves a single "tuning" parameter, which enters however in a manner that prevents the iterations from following the pattern in the progressive hedging algorithm for solving stochastic variational inequality problems [13]. The iterations described next do follow that motivating pattern.

Progressive Decoupling Algorithm (with parameter $r > 0$)**.** *In iteration ν, having $x^\nu \in S$ and $y^\nu \in S^\perp$, determine $\hat{x}^\nu \in H$ as the unique solution to the variational inequality*

$$-F^\nu(\hat{x}^\nu) \in N_C(\hat{x}^\nu), \quad \textit{where } F^\nu(x) = F(x) - y^\nu + r[x - x^\nu]. \tag{2.2}$$

Then get $x^{\nu+1} \in S$ and $y^{\nu+1} \in S^\perp$ from

$$x^{\nu+1} = P(\hat{x}^\nu), \qquad y^{\nu+1} = y^\nu - rP^\perp(\hat{x}^\nu) = y^\nu - r[\hat{x}^\nu - x^{\nu+1}]. \tag{2.3}$$

Recall here that P and $P^\perp$ are the projection mappings onto the subspaces S and $S^\perp$. The existence and uniqueness of the solution to the variational inequality in (2.2) is a consequence of F^ν being strongly monotone — because of the monotonicity of F in (1.6) and the strong monotonicity with respect to x of the proximal term $r[x - x^\nu]$.

Observe that although the iterates x^ν belong to the subspace S, they do not necessarily lie in the set C. On the other hand, the iterates $\hat{x}^\nu$ are in C but not necessarily in S. Thus, there is a jockeying back and forth between the two kinds of feasibility on the way to ultimately identifying a solution $\bar{x} \in C \cap S$.

Convergence Theorem. *Suppose that the linkage problem (1.8) is solvable.*[4] *Then the iterations (2.2)–(2.3) generate pairs (x^ν, y^ν) that converge (in the weak topology of H) to a pair $(\bar{x}, \bar{y})$ solving (1.8), with $\bar{x}$ then solving (1.7). Moreover, this will happen with the values*

$$||x^\nu - \bar{x}||^2 + \frac{1}{r^2}||y^\nu - \bar{y}||^2 \tag{2.4}$$

decreasing from one interation to the next unless they have reached 0.

Since convergence of (x^ν, y^ν) to some solution $(\bar{x}, \bar{y})$ is only guaranteed in the weak topology of H, the values in (2.4) might not necessarily tend to 0, since that would correspond to convergence of (x^ν, y^ν) to $(\bar{x}, \bar{y})$ in the strong topology of H. Of course when H is finite-dimensional, weak convergence and strong convergence coincide. Note from (2.3) that in the case of strong convergence one also has $\hat{x}^\nu$ converging to $\bar{x}$. Thus the difference between the iterates $x^\nu \in S$ and $\hat{x}^\nu \in C$ vanishes in the limit.

Spingarn derived his method of partial inverses from the proximal point algorithm [9] for solving a generalized equation $0 \in T(\bar{x})$ for a maximal monotone mapping $T : H \rightrightarrows H$. The derivation in [13] of the progressive hedging algorithm for stochastic variational inequalities, likewise proceeded right from the proximal point algorithm, although by a different argument. The progressive decoupling algorithm,

[4] As noted in the decomposition principle (1.7), this follows under indicated conditions from the solvability of the linkage problem (1.5), for which criteria have also been mentioned.

its generalization beyond stochastics, could be established similarly by that argument. However, our strategy here will instead be to derive the progressive decoupling algorithm by showing that, for fixed r, it can be identified with the execution of Spingarn's algorithm on an r-dependent modification of T that is still maximal monotone. This shortcut has the benefit of indicating how details about the manner of convergence, going beyond even the claims in the theorem above, can be carried over from Spingarn's results, gleaned in turn from the proximal point algorithm, without having to be developed independently.

Because changes of variables will have a role in our demonstration, it will help to begin by stating Spingarn's algorithm in alternative notation with ξ and η in place of x and y, and a maximal monotone mapping $T' : H \rightrightarrows H$ different from our targeted T. In this notation the problem to be solved, replacing (2.1), is

$$\text{find } \bar{\xi} \in S,\ \bar{\eta} \in S^\perp, \text{ such that } \bar{\eta} \in T'(\bar{\xi}). \tag{2.5}$$

Utilizing the projection mappings P and $P^\perp$, we can formulate the steps in Spingarn's algorithm in the following manner (with his parameter $c > 0$ taken to be $1/r$).

Method of Partial Inverses [17]. *In iteration ν, having $\xi^\nu \in S$ and $\eta^\nu \in S^\perp$, determine $\hat{\xi}^\nu \in H$ and $\hat{\eta}^\nu \in H$ as the unique solutions to*

$$[rP + P^\perp](\hat{\eta}^\nu) \in T'\Big([P + rP^\perp](\hat{\xi}^\nu)\Big), \quad \hat{\xi}^\nu + \hat{\eta}^\nu = \xi^\nu + \eta^\nu. \tag{2.6}$$

Then update by

$$\xi^{\nu+1} = P(\hat{\xi}^\nu), \qquad \eta^{\nu+1} = P^\perp(\hat{\eta}^\nu). \tag{2.7}$$

Spingarn proved about this procedure that, if problem (2.5) is solvable, the sequence of pairs (ξ^ν, η^ν) will converge in the weak topology of H to some solution pair $(\bar{\xi}, \bar{\eta})$ and do so with

$$\begin{aligned} ||\xi^{\nu+1} - \bar{\xi}||^2 + ||\eta^{\nu+1} - \bar{\eta}||^2 &= ||\xi^\nu - \bar{\xi}||^2 + ||\eta^\nu - \bar{\eta}||^2 - ||P^\perp(\hat{\xi}^\nu)||^2 - ||P(\hat{\eta}^\nu)||^2 - 2\langle \hat{\xi}^\nu - \bar{\xi}, \hat{\eta}^\nu - \bar{\eta}\rangle \\ &\le ||\xi^\nu - \bar{\xi}||^2 + ||\eta^\nu - \bar{\eta}||^2 \end{aligned} \tag{2.8}$$

On the other hand, if problem (2.5) is not solvable, then $||\xi^\nu|| + ||\eta^\nu)|| \to \infty$. Observe here that $\langle \hat{\xi}^\nu - \bar{\xi}, \hat{\eta}^\nu - \bar{\eta}\rangle \ge 0$ by the monotonicity of T', so the amount of decrease in (2.8) is positive unless $P_\perp(\hat{\xi}^\nu) = 0$ and $P(\hat{\eta}^\nu) = 0$, which would mean that $(\hat{\xi}^\nu, \hat{\eta}^\nu)$ is already itself a solution pair.

Although Spingarn's method of partial inverses may seem very different from our progressive decoupling algorithm, a close connection emerges from applying it to the maximal monotone mapping T' obtained from $T = F + N_C$ by

$$T' = [P + r^{-1}P^\perp] \circ T \circ [P + r^{-1}P^\perp].$$

In (2.6), the condition to be solved for $\hat{\xi}^\nu$ then takes the form

$$[rP + P^\perp](\eta^\nu + \xi^\nu - \hat{\xi}^\nu) \in \Big([P + r^{-1}P^\perp] \circ T\Big)\Big([P + r^{-1}P^\perp][P + rP^\perp](\hat{\xi}^\nu)\Big) = [P + r^{-1}P^\perp]\Big(T(\hat{\xi}^\nu)\Big).$$

On applying $P + rP^\perp$ to both sides, this becomes

$$r(\eta^\nu + \xi^\nu - \hat{\xi}^\nu) \in T(\hat{\xi}^\nu) = F(\hat{\xi}^\nu) + N_C(\hat{\xi}^\nu),$$

or in other words

$$-F^\nu(\hat{\xi}^\nu) \in N_C(\hat{\xi}^\nu) \text{ for } F^\nu(\hat{\xi}^\nu) = F(\hat{\xi}^\nu) - r\eta^\nu + r[\hat{\xi}^\nu - \xi^\nu]. \tag{2.9}$$

The update according to (2.7) is then

$$\xi^{\nu+1} = P(\hat{\xi}^\nu), \qquad \eta^{\nu+1} = P^\perp(\hat{\eta}^\nu) = P^\perp(\eta^\nu + \xi^\nu - \hat{\xi}^\nu) = \eta^\nu - P^\perp(\hat{\xi}^\nu), \tag{2.10}$$

inasmuch as $\xi^\nu \in S$ and $\eta^\nu \in S^\perp$.

The steps in (2.2) and (2.3) can be identified with our steps in (2.9) and (2.10) by taking $x^\nu = \xi^\nu$ and $\hat{x}^\nu = \hat{\xi}^\nu$ but $y^\nu = r\eta^\nu$. Spingarn's convergence properties for ξ^ν and η^ν translate then into those that we have claimed for x^ν and y^ν. The progressive decoupling algorithm is thereby validated along with its convergence theorem.

Note from this derivation that *the progressive decoupling algorithm and the method of partial inverses coincide for $T = F + N_C$ for the parameter value $r = 1$, but not otherwise.*

Convergence rates and stopping criteria. Conditions for linear-type convergence of the method of partial inverses were developed by Spingarn in [17] from the corresponding conditions for such convergence of the proximal point algorithm in [9] and [6]. They could now be translated to the progressive decoupling algorithm by way of the changes of variables in the preceding derivation. Criteria for inexact solutions of the subproblems (1.16) could be transported down the same path.

A topic for future work could be to replace the proximal point algorithm, as the hidden engine in the procedure, by its extension due to Eckstein and Bersekas [2]; see also Pennanen [8] for more possibilities in that direction. The connection between Spingarn's algorithm and other methodology, such as the Douglas-Rachford algorithm, is fully explained in [2].

Applications to problem decomposition and splitting. The progressive decoupling algorithm can be specialized to the decomposable structure in Section 1, where

$$C = C_1 \times \cdots \times C_q \subset H_1 \times \cdots \times H_q, \qquad F(x_1, \ldots, x_q) = (F_1(x_1), \ldots, F_q(x_q)). \tag{2.11}$$

The decoupled variational inequality for $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_q) \in S$ with respect to $\bar{y} = (\bar{y}_1, \ldots, \bar{y}_q) \in S^\perp$ is decomposable in that case into the separate conditions in (1.12). The iterations of the progressive decoupling algorithm maintain that decomposability and therefore furnish a decomposition procedure.

Progressive Decomposition Algorithm (with parameter $r > 0$)**.** *Let C and F have the decomposable structure in (2.11). In iteration ν, having $x^\nu = (x_1^\nu, \ldots, x_1^\nu) \in S$ and $y^\nu = (y_1^\nu, \ldots, y_1^\nu) \in S^\perp$, determine the components $\hat{x}_j$ of $\hat{x}^\nu = (\hat{x}_1^\nu, \ldots, \hat{x}_1^\nu)$ by*

$$\hat{x}_j^\nu = \textit{unique solution to } \; -F_j^\nu(\hat{x}_j^\nu) \in N_C(\hat{x}_j^\nu), \;\; \textit{where } \; F_j^\nu(x_j) = F_j(x_j) - y_j^\nu + r[\hat{x}_j^\nu - x_j^\nu]. \tag{2.12}$$

Then get $x^{\nu+1} = (x_1^{\nu+1}, \ldots, x_q^{\nu+1}) \in S$ and $y^{\nu+1} = (y_1^{\nu+1}, \ldots, y_q^{\nu+1}) \in S^\perp$ from

$$x^{\nu+1} = P(\hat{x}^\nu), \qquad y_j^{\nu+1} = y_j^\nu - r[\hat{x}_j^\nu - x_j^{\nu+1}] \;\; \textit{for } j = 1, \ldots, q. \tag{2.13}$$

This covers as a special case the progressive hedging algorithm for stochastic variational inequalities in [13], which iteratively decomposes into subproblems involving individual scenarios. Our desire to generalize that procedure beyond its stochastic setting is thereby fulfilled

The progressive decoupling algorithm can specialized also to the splitting structure in Section 1 in order to solve problem (1.13) by way of the rule in (1.14). This corresponds to taking S and $S^\perp$ to be the subspaces in (1.15) when applying the decomposition procedure just obtained.

Progressive Splitting Algorithm (with parameter value $r > 0$)**.** *Let C and F have the structure (2.11) with the spaces $H_1, \ldots, H_q$, all being the same space H_0. In iteration ν, having $w^\nu \in H_0$ and $y_j^\nu \in H_0$ with $y_1^\nu + \cdots + y_q^\nu = 0$, determine the components $\hat{x}_j$ of $\hat{x}^\nu = (\hat{x}_1^\nu, \ldots, \hat{x}_q^\nu)$ by*

$$\hat{x}_j^\nu = \text{unique solution to } \; -F_j^\nu(\hat{x}_j^\nu) \in N_{C_j}(\hat{x}_j^\nu), \;\; \text{where } \; F_j^\nu(x_j) = F_j(x_j) - y_j^\nu + r[x_j - x_j^\nu]. \tag{2.14}$$

Then update by

$$w^{\nu+1} = \frac{1}{q}[\hat{x}_1^\nu + \cdots + \hat{x}_q^\nu], \qquad y_j^{\nu+1} = y_j^\nu - r[\hat{x}_j^\nu - w^{\nu+1}] \;\text{ for }\; j = 1, \ldots, q. \tag{2.15}$$

3 Connections with optimization

Although our focus up to now has on variational inequality conditions, those conditions can well arise from problems of optimization. The iterations of the progressive decoupling algorithm then can be executed in minimization mode.

Linkage variational inequality problems from minimization. Suppose that $F = \nabla f$ for a continuously differentiable function $f : H \to I\!R$ which is convex relative to C, that being equivalent to the monotonicity property in (1.6) for ∇f. Then the basic variational inequality problem in (1.7) poses the first-order condition that is both necessary and sufficient for minimizing $f(x)$ over $C \cap S$, and it can therefore be expressed by

$$\text{find } \; \bar{x} \in \operatorname*{argmin}_{x \in C \cap S} f(x). \tag{3.1}$$

The corresponding problem (1.8) in decoupled form incorporates a multiplier vector $\bar{y}$ for the constraint $\bar{x} \in S$ (equivalently, $P(x) = 0$) and takes the form

$$\text{find } \; \bar{x} \in S \; \text{ such that } \; \bar{x} \in \operatorname*{argmin}_{x \in C} \bar{f}(x), \;\; \text{where } \; \bar{f}(x) = f(x) - \langle \bar{y}, x \rangle \; \text{ for some } \; \bar{y} \in S^\perp. \tag{3.2}$$

In this situation strong mononicity of $F = \nabla f$ relative to C would be equivalent to strong convexity of f relative to C and would ensure that the minimum over C is attained at a unique point. Even if not available for f, strong convexity will assist in the steps of the algorithm.

Progressive Decoupling Algorithm in Optimization Mode (with parameter $r > 0$)**.** *In iteration ν with $x^\nu \in S$ and $y^\nu \in S^\perp$, determine $\hat{x}^\nu \in H$ by solving a strongly convex problem of minimization,*

$$\hat{x}^\nu = \operatorname*{argmin}_{x \in C} f^\nu(x), \;\; \text{where } \; f^\nu(x) = f(x) - \langle y^\nu, x \rangle + \frac{r}{2}||x - x^\nu||^2. \tag{3.3}$$

Then get $x^{\nu+1} \in S$ and $y^{\nu+1} \in S^\perp$ by

$$x^{\nu+1} = P(\hat{x}^\nu), \qquad y^{\nu+1} = y^\nu - rP^\perp(\hat{x}^\nu) = y^\nu - r[\hat{x}^\nu - x^{\nu+1}]. \tag{3.4}$$

The convergence theorem for progressive decoupling applies to this. The proximal term at the end of (3.4), for which the gradient is the term $r[x - x^\nu]$ in the variational inequality expression of progressive decoupling, adds strong convexity that makes sure the minimum in the subproblem is attained and uniquely.

Minimization problems with decomposable structure. Let $H = H_1 \times \cdots \times H_q$ for Hilbert spaces H_j and consider closed convex sets $C_j \subset H_j$ along with continuously differentiable convex functions $f_j : C_j \to I\!R$. The problem

$$\text{minimize } \; f_1(x_1) + \cdots + f_1(x_q) \; \text{ subject to } \; x_j \in C_j, \; (x_1, \ldots, x_q) \in S, \tag{3.5}$$

specializes (3.1) as a coupled problem of optimization for which the decoupled version parallel to (3.2) seeks to

$$\begin{aligned} &\text{find } \; (\bar{x}_1, \ldots, \bar{x}_q) \in S \; \text{ such that } \; \bar{x}_j \in \operatorname*{argmin}_{x_j \in C_j} \bar{f}_j(x_j) \; \text{ for } \; j = 1, \ldots, q, \\ &\text{where } \; \bar{f}_j(x_j) = f_j(x_j) - \langle \bar{y}_j, x_j \rangle \; \text{ with respect to some } \; (\bar{y}_1, \ldots, \bar{y}_q) \in S^\perp. \end{aligned} \tag{3.6}$$

The facts in Section 1 relating the coupled and decoupled problems (1.7) and (1.8) hold for problems (3.5) and (3.6) as a special case and tell us that by solving (3.6) we can get a solution $\bar{x}$ to (3.5). The progressive decoupling algorithm in this setting provides a way of doing that.

Progressive Decomposition Algorithm in Optimization Mode (with parameter $r > 0$)**.** *Assume that the minimization problem in (3.1) has the structure in (3.5). In iteration ν, having $x^\nu = (x_1^\nu, \ldots, x_1^\nu) \in S$ and $y^\nu = (y_1^\nu, \ldots, y_1^\nu) \in S^\perp$, determine the components $\hat{x}_j$ of $\hat{x}^\nu = (\hat{x}_1^\nu, \ldots, \hat{x}_1^\nu)$ by solving strongly convex minimization subproblems in parallel for $j = 1, \ldots, q$:*

$$\hat{x}_j^\nu = \operatorname*{argmin}_{x_j \in C_j} f_j^\nu(x_j), \quad \textit{where } \; f_j^\nu(x_j) = f_j(x_j) - \langle y_j^\nu, x_j \rangle + \frac{r}{2} ||x_j - x_j^\nu||^2. \tag{3.7}$$

Then update to

$$x^{\nu+1} = P(\hat{x}^\nu), \qquad y_j^{\nu+1} = y_j^\nu - r[\hat{x}_j^\nu - x_j^{\nu+1}] \; \textit{ for } j = 1, \ldots, q. \tag{3.8}$$

Once again we can contemplate as a particular application the stochastic linkage structure of nonanticipativity which was explained in Section 2 after (1.12). There, the progressive hedging algorithm for solving stochastic variational inequality problems was the featured special case, but now it is the original progressive hedging algorithm in stochastic programming [14]. That procedure, which reduces the solution of a given problem iteratively into solving deterministic optimization subproblems for individual scenarios, was the source of the pattern we wanted to emulate in developing the progressive decoupling algorithm.

Minimization problems with splitting structure. We turn finally to the case of splitting structure in variational inequalities as applied with $F_j = \nabla f_j$. This corresponds in optimization to the problem

$$\text{minimize } \; f_1(w) + \cdots f_q(w) \; \text{ over } \; w \in C_1 \cap \cdots \cap C_q \tag{3.9}$$

and reconstituting it as the problem

$$\text{minimize } \; f_1(x_1) + \cdots f_q(x_q) \; \text{ subject to } \; x_j \in C_j \; \text{ and } \; x_1 = \cdots = x_q, \tag{3.10}$$

with the common value of the x_j's in (3.10) being identified in the end as the w in (3.9). In effect we adopt the structure for decomposition in optimization that we have just been dealing with, but in the case where the spaces H_j are all the same H_0, and with the linkage specialized to the subpace pair S and $S^\perp$ in (1.15).

The computations afforded by applying the progressive decoupling algorithm in optimization mode take the following pattern in this situation.

Progressive Splitting Algorithm in Optimization Mode (with parameter $r > 0$). *In iteration ν, having $w^\nu \in H_0$ and $y_j^\nu \in H_0$ with $y_1^\nu + \cdots + y_q^\nu = 0$, calculate the components $\hat{x}_j$ of $\hat{x}^\nu = (\hat{x}_1^\nu, \ldots, \hat{x}_1^\nu)$ by*

$$\hat{x}_j^\nu = \operatorname*{argmin}_{x_j \in C_j} f^\nu(x_j), \quad \text{where } f^\nu(x_j) = f_j(x_j) - \langle y_j^\nu, x_j \rangle + \frac{r}{2}||x_j - x_j^\nu||^2. \tag{3.11}$$

Then update by

$$w^{\nu+1} = \frac{1}{q}[\hat{x}_1^\nu + \cdots + \hat{x}_q^\nu], \qquad y_j^{\nu+1} = y_j^\nu - r[\hat{x}_j^\nu - w^{\nu+1}] \text{ for } j = 1, \ldots, q. \tag{3.12}$$

The case of this algorithm with $r = 1$ matches a splitting method that Spingarn in [17] derived from his method of partial inverses. The extra flexibility provided by admitting parameter values $r \neq 1$ illustrates an advantage of the proximal decompling algorithm over Spingarn's approach.

A splitting algorithm having a minimization step like (3.11) with general r but a different way of updating was proposed by Mahey et al. [7] in the two-component case, i.e., $q = 2$. It can be extended also to $q > 2$. The survey of Lenoir and Mahey [5] gives many insights into this and other connections with Spingarn's work.

4 Extensions to augmented Lagrangian methodology

Beyond direct applications to problems of minimization, progressive decoupling provides ways of dealing broadly with Lagrange multipliers. Although that topic can be pursued for general variational inequalities, for which the role of Lagrange multipliers was explained in [11], the ideas are easier to understand when viewed in an optimzation context. That is what we will concentrate on here.

For background, we pose a problem in a Hilbert space H_0 of the type

$$\text{minimize } f(x) \text{ subject to } x \in X,\ G(x) \in K, \tag{4.1}$$

where $f : H_0 \to I\!R$ is continuously differentiable, $X \subset H_0$ is nonempty, closed, convex, K is a closed convex cone in another Hilbert space H_0', and $G : H_0 \to H_0'$ is continuously differential. First-order optimality conditions for such a problem can be expressed in terms of the Lagrangian function

$$l(x, y) = f(x) + \langle y, G(x) \rangle \text{ for } x \in X,\ y \in Y, \text{ where } Y = K^* \text{ (polar cone)}, \tag{4.2}$$

in the form

$$-\nabla_x l(\bar{x}, \bar{y}) \in N_X(\bar{x}), \quad \nabla_y l(\bar{x}, \bar{y}) \in N_Y(\bar{y}), \tag{4.3}$$

in which $\bar{y}$ is a Lagrange multiplier associated with $\bar{x}$ with respect to the constraint $G(x) \in K$. The interesting aspect for our purposes here is that the conditions (4.3) can be written as a variational inequality:

$$-F(\bar{x}, \bar{y}) \in N_{X \times Y}(\bar{x}, \bar{y}) \text{ with } F(x, y) = (\nabla_x l(x, y), -\nabla_y l(x, y)). \tag{4.4}$$

Our focus will be on the *convex* case of problem (4.1), by which we mean the case in which the Lagrangian $l(x, y)$ is convex in $x \in X$ for each $y \in Y$. Since $l(x, y)$ is concave (actually affine) in $y \in Y$ for each $x \in X$, this corresponds to l being a convex-concave *saddle function* on $X \times Y$. Then (4.2) is equivalent to the saddle point condition

$$\bar{x} \in \operatorname*{argmin}_{x \in X} l(x, \bar{y}), \qquad \bar{y} \in \operatorname*{argmax}_{y \in Y} l(\bar{x}, y), \tag{4.5}$$

which is always sufficient for the global optimality of $\bar{x}$ as a solution to problem (4.1) and is necessary, with respect to some $\bar{y}$, under a constraint qualification. Most important here is that F is *monotone* in the convex case, which we assume we are in henceforth. We then have a monotone variational inequality to which the approaches we have been developing might be invoked.

Along with the Lagrangian l in (4.2) we will also be able to appeal to the associated *augmented* Lagrangian function, which depends on a parameter value $\rho > 0$ and is given by

$$\begin{aligned} l_\rho(x,y) &= \max_{y' \in Y} \Big\{ l(x,y') - \tfrac{1}{2\rho}||y'-y||^2 \Big\} \\ &= f(x) + \langle y, G(x)\rangle + \tfrac{\rho}{2}||G(x)||^2 - \tfrac{1}{2\rho}\mathrm{dist}_Y^2(y+\rho G(x)) \ \text{ over } \ X \times H_0', \end{aligned} \tag{4.6}$$

where

$$\mathrm{dist}_Y(u) = \text{ distance of } u \text{ from } Y \ = \min_{y\in Y} ||u-y||. \tag{4.7}$$

Note that when $Y = H_0'$, which corresponds to $K = \{0\}$ and an equation constraint $G(x) = 0$, the distance term drops off. This is the most familiar version on an augmented Lagrangian, but the theory goes far beyond that.[5] The conditions in (4.3) can be expressed equivalently through the augmented Lagrangian as

$$-\nabla_x l_\rho(\bar{x},\bar{y}) \in N_X(\bar{x}), \qquad \nabla_y l_\rho(\bar{x},\bar{y}) = 0. \tag{4.8}$$

From the definition (4.6), l_ρ not only retains the convexity-concavity of l but also the continuous differentiability, in fact with

$$\nabla_y l_\rho(x,y) = \operatorname*{argmax}_{y'\in Y} \Big\{ l(x,y') - \frac{1}{2\rho}||y'-y||^2 \Big\}. \tag{4.9}$$

The augmented Lagrangian will be important in progressive decoupling for the following reason. Suppose the variational inequality (4.4) is modified by the addition of a proximal term, that is, by passing from $F(x,y)$ to $F(x,y) + r[(x,y) - (x^*,y^*)]$ for some choice of (x^*,y^*). That would correspond to replacing (4.5) by the saddle point condition for

$$l(x,y) + \frac{r}{2}||x-x^*||^2 - \frac{r}{2}||y-y^*||^2 \ \text{ for } \ x \in X, \ y \in Y. \tag{4.10}$$

Due to the strong convexity-concavity in this expression we can calculate the saddle point by first maximizing over $y \in Y$ and then minimizing the "residual" over $x \in X$. But the residual from maximizing in y is the augmented Lagrangian $l_\rho(x,y^*)$ for $\rho = r^{-1}$. After the minimization of that has been carried out to get $\bar{x}$, the $\bar{y}$ component of the saddle point can be obtained from maximizing over $y \in Y$ in (4.10) with x fixed at $\bar{x}$. Thus, the unique saddle point in (4.10) can be determined by

$$\begin{aligned} \bar{x} &= \operatorname*{argmin}_{x\in X}\Big\{ l_\rho(x,y^*) + \tfrac{1}{2\rho}||x-x^*||^2 \Big\} \ \text{ for } \ \rho = r^{-1}, \ \text{ then} \\ \bar{y} &= \operatorname*{argmax}_{y\in Y}\Big\{ l(\bar{x},y) - \tfrac{1}{2\rho}||y-y^*||^2 \Big\} = \mathrm{proj}_Y(y^* + \rho G(\bar{x})), \end{aligned} \tag{4.11}$$

where

$$\mathrm{proj}_Y(u) = \text{nearest point of } Y \text{ to } u. \tag{4.12}$$

With these facts in hand we can proceed with further applications of progressive decoupling.

For monotone variational inequalities of Lagrangian type arising from the constraint model in (4.1), we can contemplate decomposable structure or splitting structure which might be open to progressive

[5] See for example Chapter 11 of the book [15].

decoupling. Interestingly, those two types of structure can enter simultaneously, one in the primal variables and the other in the dual variables.

Lagrangian variational inequalities with decomposable structure. Decomposable structure in the primal variables in (4.1) is introduced in the primal variables by taking the space H_0 with which we began this section to be $H_1 \times \cdots \times H_q$ and letting $x = (x_1, \ldots, x_q)$ with $x_j \in H_j$:

$$\text{minimize } \sum_{j=1}^q f_j(x_j) \text{ subject to } x_j \in X_j,\ \sum_{j=1}^q G_j(x_j) \in K \tag{4.13}$$

with $f_j : H_j \to I\!R$ and $G_j : H_j \to H_0'$ continuously differentiable, and $X_j \subset H_j$ nonempty, closed, convex. The Lagrangian over the product of $X = X_1 \times \cdots X_q$ and Y is then

$$l(x,y) = \sum_{j=1}^q l_j(x_j, y), \text{ where } l_j(x_j, y) = f_j(x_j) + \langle y, G_j(x_j)\rangle. \tag{4.14}$$

As we know, the corresponding variational inequality gives the conditions for a saddle point with respect to minimizing in the primal variables in X and maximizing in the dual variables Y. If there were a separate multiplier element in Y for each j, this would break down into a separate saddle point problem for each j. This suggests applying the ideas about "splitting" to y. For that we can pass to saddle points of

$$\sum_{j=1}^q l_j(x_j, z_j) = \sum_{j=1}^q \Big[f_j(x_j) + \langle z_j, G_j(x_j)\rangle \Big] \text{ for } x_j \in X_j,\ z_j \in Y,\ z_1 = \cdots = z_q. \tag{4.15}$$

To exploit that idea by way of a linkage model, we can take $H = H_1 \times \cdots \times H_q \times \Pi_{j=1}^q H_0'$ and let

$$\begin{aligned} &S = \{\, (x_1, \ldots, x_q, z_1, \ldots, z_q) \,|\, \exists\, y \in H_0' \text{ such that } z_1 = \cdots = z_q = y\}, \\ &S^\perp = \{\, (0, \ldots, 0, w_1, \ldots, w_q) \,|\, w_j \in H_0',\ w_1 + \cdots + w_1 = 0\}. \end{aligned} \tag{4.16}$$

Then, while the variational inequality problem in basic form can be identified with determing a saddle point in (4.15), the corresponding problem in decoupled form corresponds to determining instead a saddle point of

$$\sum_{j=1}^q \Big[l_j(x_j, z_j) - \langle \bar{w}_j, z_j\rangle \Big] = \sum_{j=1}^q \Big[f_j(x_j) + \langle z_j, G_j(x_j) - \bar{w}_j\rangle \Big] \text{ for } x_j \in X_j,\ z_j \in Y. \tag{4.17}$$

How does the progressive decoupling algorithm play out in this situation?

In iteration ν with $(x_1^\nu, \ldots, x_q^\nu, z_1^\nu, \ldots, z_q^\nu) \in S$, i.e., the z_j^ν's all equaling some y^ν, we subtract from the Lagrangian in (4.17) the term $\sum_{j=1}^q \langle w_j^\nu, z_j^\nu\rangle$ in the dual variables while adding the proximal terms $\frac{r}{2}||x_j - x_j^\nu||^2$ in the primal variables and subtracting the proximal terms $\frac{r}{2}||z_j - y^\nu||^2$ in the dual variables, getting an expression that can be written as

$$\begin{aligned} &\sum_{j=1}^q \Big[l_j^\nu(x_j, z_j) + \tfrac{r}{2}||x_j - x_j^\nu||^2 - \tfrac{r}{2}||z_j - y^\nu||^2 \Big] \text{ for } x_j \in X_j,\ z_j \in Y, \\ &\quad \text{where } l_j^\nu(x_j, z_j) = f_j(x_j) + \langle z_j, G_j(x_j) - w_j^\nu\rangle. \end{aligned} \tag{4.18}$$

Note that $l_j^\nu(x_j, z_j)$ is the Lagrangian function for the problem:

$$\text{minimize } f_j(x_j) \text{ subject to } x_j \in X_j,\ G_j(x_j) - w_j^\nu \in K. \tag{4.19}$$

We calculate $(\hat{x}_1^\nu, \ldots, \hat{x}_1^\nu, \hat{z}_1^\nu, \ldots, \hat{z}_q^\nu)$ as the unique saddle point in (4.18), which decomposes into

$$(\hat{x}_j^\nu, \hat{z}_j^\nu) = \underset{x_j \in X_j,\, z_j \in Y}{\text{argminimax}} \Big\{ l_j^\nu(x_j, z_j) + \frac{r}{2}||x_j - x_j^\nu||^2 - \frac{r}{2}||z_j - y^\nu||^2 \Big\}. \tag{4.20}$$

We then update by

$$x_j^{\nu+1} = \hat{x}_j^\nu, \qquad y^{\nu+1} = \frac{1}{q}[\hat{z}_1^\nu + \cdots + \hat{z}_q^\nu], \qquad w_j^{\nu+1} = w_j^\nu - r[z_j^\nu - y^{\nu+1}]. \tag{4.21}$$

An important simplification now enters through the observations made earlier in the context of (4.10) and (4.11) about saddle points when proximal terms are present. The saddle point subproblems in (4.19) can be solved in terms of the augmented Lagrangians

$$l_{j,r^{-1}}^\nu(x_j, y^\nu) = l_j(x_j, y^\nu) + \frac{1}{2r}||G_j(x_j) - w_j^\nu||^2 - \frac{r}{2}\text{dist}_Y^2(y^\nu + r^{-1}[G_j(x_j) - w_j^\nu]) \tag{4.22}$$

for the problems in (4.19) by

$$\hat{x}_j^\nu = \operatorname*{argmin}_{x_j \in X_j} \Big\{ l_{j,r^{-1}}^\nu(x_j, y^\nu) + \frac{r}{2}||x_j - x_j^\nu||^2 \Big\}, \quad \hat{z}_j^\nu = \text{proj}_Y(y^\nu + r^{-1}[G_j(x_j) - w_j^\nu]). \tag{4.23}$$

By converting from r to $\rho = r^{-1}$, we arrive at the procedure laid out next.

Progressive Decomposition Algorithm in Lagrangian Mode (with parameter $r > 0$**.** *For the purpose of solving problem (4.13) by way of its Lagrangian conditions for optimality, proceed as follows. In iteration ν, having $x_j^\nu \in X_j$, $y^\nu \in Y = K^*$ and w_j^ν with $w_1^\nu + \cdots + w_q^\nu = 0$, determine $\hat{x}_j^\nu$ and $\hat{z}_j^\nu$ for $j = 1, \ldots, q$ from the augmented sub-Lagrangians*

$$l_{j,\rho}^\nu(x_j, y^\nu) = f_j(x_j) + \langle y^\nu, G_j(x_j) - w_j^\nu \rangle + \frac{\rho}{2}||G_j(x_j) - w_j^\nu||^2 - \frac{\rho}{2}\text{dist}_Y^2(y^\nu + \rho G_j(x_j)) \tag{4.23}$$

by calculating

$$\hat{x}_j^\nu = \operatorname*{argmin}_{x_j \in X_j} \Big\{ l_{j,\rho}^\nu(x_j, y^\nu) + \frac{1}{2\rho}||x_j - x_j^\nu||^2 \Big\}, \qquad \hat{z}_j^\nu = P_Y(y^\nu + \rho[G_j(\hat{x}_j^\nu) - w_j^\nu]\,). \tag{4.25}$$

Then update by

$$x_j^{\nu+1} = \hat{x}_j^\nu, \qquad y^{\nu+1} = \frac{1}{q}[\hat{z}_1^\nu + \cdots + \hat{z}_q^\nu], \qquad w_j^{\nu+1} = w_j^\nu - \frac{1}{\rho}[\hat{z}_j^\nu - y^{\nu+1}]. \tag{4.27}$$

This is very similar to an algorithm developed by Spingarn in [18], but there the proximal term in the primal variables has $\rho/2$ instead of $1/2\rho$. Also, he only treated classical inequality constraints, namely the case of $K = I\!R^m_-$. It differs from the SALA approach to the same situation described by Lenoir and Mahey in [5], which has no proximal term in the primal variables and focuses on affine equation constraints. Proximal terms in the primal variables in augmented Lagrangian methods go back to [10].

Example of linear equation constraints. As a special case of the decomposable structure in the optimization problem (4.13), suppose that $K = \{0\}$ and $G_j(x_j) = A_j x_j$ for a (continuous) linear mapping A_j. The task then is to

$$\text{minimize} \;\; \textstyle\sum_{j=1}^q f_j(x_j) \;\; \text{subject to} \;\; \sum_{j=1}^q A_j x_j = 0 \;\; \text{with} \;\; x_j \in X_j. \tag{4.27}$$

The augmented Lagrangian decomposition method then has[6]

$$l_{j,\rho}^\nu(x_j, y^\nu) = f_j(x_j) + \langle y^\nu, A_j x_j - w_j^\nu \rangle + \frac{\rho}{2}||A_j x_j - w_j^\nu||^2 \tag{4.28}$$

[6]The quadratic term in this formula was missing in the published version of the paper.

and proceeds by calculating

$$\hat{x}_j^\nu = \underset{x_j \in X_j}{\operatorname{argmin}} \Big\{ l_{j,\rho}^\nu(x_j, y^\nu) + \frac{1}{2\rho}||x_j - x_j^\nu||^2 \Big\}, \qquad \hat{z}_j^\nu = y^\nu + \rho[A_j(\hat{x}_j^\nu) - w_j^\nu]\,) \tag{4.29}$$

and updating by

$$x_j^{\nu+1} = \hat{x}_j^\nu, \quad y^{\nu+1} = \frac{1}{q}[\hat{z}_1^\nu + \cdots + \hat{z}_q^\nu] = y^\nu + \frac{\rho}{q}\sum\nolimits_{j=1}^q A_j \hat{x}_j^\nu, \quad w_j^{\nu+1} = w_j^\nu - \frac{1}{\rho}[\hat{z}_j^\nu - y^{\nu+1}]. \tag{4.30}$$

This example covers in particular a problem formulation in convex optimization which has received very wide attention. By stating it as

$$\text{minimize } \ f_1(u) + f_2(Au) \ \text{ with } \ u \in X_1, \ Au \in X_2, \tag{4.31}$$

we can identify it in the pattern of (4.27) as

$$\text{minimize } \ f_1(x_1) + f_2(x_2) \ \text{ subject to } \ x_1 \in X_1, \ x_2 \in X_2, \ Ax_1 - x_2 = 0, \tag{4.32}$$

which corresponds to $A_1 = A$ and $A_2 = -I$. The iteration (4.28)–(4.29) can be executed then with $q = 2$ and the simplification that $w_2^\nu = -w_1^\nu$.

Problem (4.31) is typically considered without the differentiability assumptions we have imposed on the functions f_i, through our choice of variational inequalities in traditional formulation as the basis for exposition, but this is a minor point. The derivation of our algorithm from Spingarn's method of partial inversed didn't depend on it.

The best-known approaches to solving (4.31) in a decomposable manner are variants of the alternating direction method of multipliers (ADMM) as derived from the Douglas-Rachford algorithm. A history of this method has recently been provided by Glowinski [4]. The literature on it is huge. The article of Lenoir and Mahey [5] can be helpful and also the survey of Bertsekas [1]. Another method proposed for solving (4.30) is due to Chen and Teboulle [3]. The procedure suggested here is simpler, but its comparative efficacy must await testing.

Lagrangian variational inequalities with splitting structure. Consider now an optimization problem of the form

$$\begin{array}{l} \text{minimize } \ f_1(w) + \cdots + f_q(w) \ \text{ over } \ w \in C_1 \cap \cdots \cap C_q \\ \text{where } \ C_j = \{\, w \in X_j \,|\, G_j(w) \in K_j \} \ \text{ for } \ j = 1, \ldots, q, \end{array} \tag{4.33}$$

which builds on the problem in (3.13) by adding constraint details. The trick once more is to expand into the form in (3.14) but then to pass to the Lagrangian. Multiplier vectors y_j for the G_j constraints must belong to the closed convex cones Y_j that are polar to the closed convex cones K_j. The Lagrangian function is therefore

$$\sum\nolimits_{j=1}^q l_j(x_j, y_j) \ \text{ for } \ x_j \in X_j, \ y_j \in Y_j, \ \text{ with } \ x_1 = \cdots = x_q. \tag{4.34}$$

The linkage variational inequality in coupled form in this setting is the Lagrangian variational inequality coming from (4.34) with

$$F(x,y) = (\nabla_{x_1} l_1(x_1,y_1), \ldots, \nabla_{x_q} l_q(x_1,y_q), -\nabla_{y_1} l_1(x_1,y_1), \cdots, -\nabla_{y_q} l_q(x_q,y_q))$$

and the set

$$[X_1 \times \cdots \times X_q \times Y_1 \times \cdots \times Y_q] \cap S \ \text{ for } \ S = \{\, (x,y) \,|\, \exists\, w \ \text{ with } \ x_1 = \cdots = x_q = w\}. \tag{4.35}$$

For the indicated linkage subspace S the complementary subspace is

$$S^{\perp} = \{ (z_1, \dots, z_q, 0, \dots, 0) \,|\, z_1 + \cdots + z_q = 0\}. \tag{4.36}$$

The associated linkage variational inequality in decoupled form thus seeks to

$$\begin{array}{l}\text{find } (\bar{x}_j, \bar{y}_j) \text{ with } (-\nabla_{x_j} l_j(\bar{x}_j, \bar{y}_j), \nabla_{y_j} l_j(\bar{x}_j, \bar{y}_j)) + (\bar{z}_j, 0) \in N_{X_j \times Y_j}(\bar{x}_j, \bar{y}_j),\ j = 1, \dots, q, \\ \text{for some choices of } \bar{z}_j \text{ having } \bar{z}_1 + \cdots + \bar{z}_q = 0 \text{ allowing that actually } \bar{x}_1 = \cdots = \bar{x}_q.\end{array} \tag{4.37}$$

The jth variational inequality in (4.37) is the saddle point optimality condition for the problem of minimizing the "tilted" function $\bar{f}_j(x_j) = f_j(x_j) - \langle \bar{z}_j, x_j\rangle$ subject to $x_j \in X_j$ and $G_j(x_j) \in K_j$. The tilt vectors are aimed at allowing the solutions $\bar{x}_j$ to be the same $\bar{w}$, which then solves (4.33).

The progressive decoupling algorithm introduces proximal terms as in its optimization version in (3.16) but additional proximal terms now also in the dual variables. The tilted Lagrangian $l_j(x_j, y_j) - \langle z_j, x_j\rangle$ behind the variational inequality in (4.36) is replaced in iteration ν by

$$l_j(x_j, y_j) - \langle z_j^\nu, x_j\rangle + \frac{r}{2}||x_j - x_j^\nu||^2 - \frac{r}{2}||y_j - y_j^\nu||^2, \text{ where } x_1^\nu = \cdots = x_q^\nu = \text{some } w^\nu, \tag{4.38}$$

for which the unique saddle point $(\hat{x}_j^\nu, \hat{y}_j^\nu)$ is sought over $X_j \times Y_j$ for each j. The update then is

$$y_j^{\nu+1} = \hat{y}_j^\nu, \qquad w^{\nu+1} = \frac{1}{q}\sum\nolimits_{j=1}^q \hat{x}_j^\nu, \qquad z_j^{\nu+1} = z_j^\nu - r[\hat{x}_j^\nu - w^{\nu+1}]. \tag{4.39}$$

The methodology of augmented Lagrangians can come into play now as a simplification. In maximizing over y_j in the saddle point problem for (4.38) the residual expression is

$$l_{j,r^{-1}}(x_j, y_j^\nu) - \langle z_j^\nu, x_j\rangle + \frac{r}{2}||x_j - w^\nu||^2 \tag{4.40}$$

which then has to be minimized in x_j to get $\hat{x}_j^\nu$. The $\hat{y}_j^\nu$ component of the saddle point is then the projection of $y_j^\nu + r^{-1}G_j(\hat{x}_j^\nu)$ on Y_j. With these simplifications the resulting procedure takes the following form with ρ replacing r^{-1}.

Progressive Splitting Algorithm in Lagrangian Mode (with parameter $\rho > 0$)**.** *For the purpose of solving problem (4.32) by way of its Lagrangian conditions for optimality, proceed as follows. In iteration ν, having w^ν along with $y_j^\nu \in Y_j$ and z_j^ν such that $z_1^\nu + \cdots + z_q^\nu = 0$, determine $(\hat{x}_j^\nu, \hat{y}^\nu)$ by*

$$\hat{x}_j^\nu = \operatorname*{argmin}_{x_j \in X_j}\Big\{ l_{j,\rho}(x_j, y_j^\nu) - \langle z_j^\nu, x_j\rangle + \frac{1}{2\rho}||x_j - w^\nu||^2\Big\}, \qquad \hat{y}_j^\nu = \text{proj}_{Y_j}(y_j^\nu + \rho G_j(\hat{x}_j^\nu)). \tag{4.41}$$

Then update by

$$y_j^{\nu+1} = \hat{y}_j^\nu, \qquad w^{\nu+1} = \frac{1}{q}\sum\nolimits_{j=1}^q \hat{x}_j^\nu, \qquad z_j^{\nu+1} = z_j^\nu - \rho^{-1}[\hat{x}_j^\nu - w^{\nu+1}]. \tag{4.42}$$

References

[1] Bertsekas, D. P., "Incremental aggregated proximal and augmented Lagrangian algorithms." Lab. for Information and Decision Systems Report LIDS-P-3176, Massachusetts Institute of Technology, 2016.

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

[3] Chen, G., and Teboulle, M., "A proximal-based decomposition method for convex minimization problems." *Mathematical Programming* **64** (1994), 81–101.

[4] Glowinski, R., "On alternating direction methods of multipliers: a historical perspective," in *Modeling Simulation and Optimization for Science and Technology*, Computational Methods in Applied Sciences **43** (W. Fitzgibbon et al., eds.), Springer (2014), 59–82.

[5] Lenoir, A., and Mahey, P., "A survey on operator splitting and decomposition of convex programs," *RAIRO —Operations Research* **51** (2017), 17–41.

[6] Luque, F. J. R., "Asymptotic convergence analysis of the proximal point algorithm," *SIAM Journal of Control and Optimizat* **22** (1984), 277–293.

[7] Mahey, P., Oualibouch, S., and Pham, D. T., "Proximal decomposition on the graph of a maximal monotone mapping," *SIAM J. Optimization* **5** (1995), 454–466.

[8] Pennanen, T., "Local convergence of the proximal point algorithm and multiplier methods without monotonicity." *Mathematics of Operations Research* **27** (2002), 170–191.

[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., "Lagrange multipliers and variational inequalities," in *Variational Inequalities and Complementarity Problems*, R. W. Cottle, F. Gianessi, and J.-L. Lions, (eds.), Wiley, 1982, 303–322.

[12] Rockafellar, R. T., "Progressive decoupling of linkages in nonmonotone variational inequalities and nonconvex optimization," in preparation.

[13] Rockafellar, R. T., and Sun, Jie "Solving monotone stochastic variational inequalities and complementarity problems by progressive hedging," *Mathematical Programming B*, submitted.

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

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

[16] Rockafellar, R. T., and Wets, R. J-B, "Stochastic variational inequalities: single-stage to multistage." *Mathematical Programming B* **165** (2017), 331–360.

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

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