Set-valued and Variational Analysis 27 (2019), 863–893
http://dx.doi.org/10.1007/s1128-018-0496-1

PROGRESSIVE DECOUPLING OF LINKAGES IN OPTIMIZATION AND VARIATIONAL INEQUALITIES WITH ELICITABLE CONVEXITY OR MONOTONICITY

R. Tyrrell Rockafellar[1]

Abstract

Algorithms for problem decomposition and splitting in optimization and the solving of variational inequalities have largely depended on assumptions of convexity or monotonicity. Here, a way of "eliciting" convexity or monotonicity is developed which can get around that limitation. It supports a procedure called the progressive decoupling algorithm, which is derived from the proximal point algorithm through passing to a partial inverse, localizing and rescaling. In the optimization setting, elicitability of convexity corresponds to a new and very general kind of second-order sufficient condition for a local minimum. Applications are thereby opened up to problem decomposition and splitting even in nonconvex optimization, moreover with augmented Lagrangians for subproblems assisting in the implementation.

Keywords: *convex/nonconvex optimization, monotone/nonmonotone variational inequalities, linkage problems, progressive decoupling, progressive hedging, problem decomposition, splitting, elicitable convexity, elicitable monotonicity, variational convexity, variational second-order sufficiency, proximal point algorithm, method of partial inverses, proximal methods of multipliers, augmented Lagrangians*

[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

Many techniques have been devised for decomposition or splitting in the solution of variational inequalities and related problems of optimization, but mostly in circumstances of global monotonicity or convexity. Here a procedure called the *progressive decoupling algorithm* will be developed that can transcend such limitations by eliciting convexity or monotonicity, whether globally or locally, in situations where its availability may not be evident. In optimization the availability emerges when near enough to a local minimum that satisfies a new, general kind of second-order sufficiency condition. That echoes the convergence of Newton-type iterations in smooth optimization once they are in some neighborhood of a point where strong conditions for a local minimum hold. However the method in this paper can embrace nonsmoothness and work with problem formulations far more general than nonlinear programming. It can take advantage of augmented Lagrangians despite their reputation for often interfering with decomposition technology, and in so doing it can utilize their convexifying capabilities in the "elicitation."

The progressive decoupling algorithm emulates the pattern of the *progressive hedging algorithm* [28] for solving stochastic programming problems, and more recently stochastic variational inequalities [26], which promotes iterative decomposition into subproblems focused on individual scenarios. For problems outside of the stochastic setting (where "hedging" is no longer an appropriate term), the progressive decoupling algorithm adopts the platform of Spingarn's *method of partial inverses* [31, 32]. Like that, it relies ultimately on properties of the *proximal point algorithm* [22]. But it diverges from Spingarn's method in "localizing," "eliciting," and in the treatment of computational parameters.

The localization ideas come from contributions of Pennanen [16] to the local execution of the proximal point algorithm, although the implementation of those ideas in optimization mode requires further results such as were obtained recently in our paper [25]. The elicitation scheme we develop is completely new. Pennanen in [16] likewise pursued hidden monotonicity of a sort, but his approach was different and not oriented to decoupling linkages.[2]

The novel role of localization and elicitation in the progressive decoupling algorithm has been stressed so far, but new things come out of the procedure even when global monotonicity or convexity is on hand. Results in that context, with specializations to problem decomposition and splitting, were recently laid out in a conference paper [24] with a simplified derivation. Those results will be covered now in a broader framerowk with additional parameters to work with.

Let H be a Hilbert space with inner product $\langle x, y\rangle$ and consider a (generally) set-valued mapping $T : H \rightrightarrows H$. Let S be a (closed) subspace of H and let $S^\perp$ be its orthogonal complement. Fundamentally, we will be concerned in this paper with solving the *linkage problem* for S and T, by which we mean

$$\text{find } \bar{x} \in S \text{ and } \bar{y} \in S^\perp \text{ such that } \bar{y} \in T(\bar{x}). \tag{1.1}$$

The idea is that the condition $\bar{x} \in S$, with S being a subspace specifying linear relationships, represents a "linkage" constraint with respect to which $\bar{y} \in S^\perp$ can act as a kind of multiplier element.

An important case case to keep in mind is

$$T(x) = F(x) + N_C(x) \text{ for } C \subset H \text{ nonempty closed convex, } F : C \to H \text{ continuous,} \tag{1.2}$$

where $N_C : H \rightrightarrows H$ is the normal cone mapping associated with C,

$$v \in N_C(x) \iff x \in C \text{ and } \langle v, x' - x\rangle \le 0 \;\; \forall x' \in C. \tag{1.3}$$

[2] Still, some combination of the two approaches may well turn out to be valuable.

The linkage problem (1.1) then has a close connection with the variational inequality

$$-F(\bar{x}) \in N_{C\cap S}(\bar{x}) \tag{1.4}$$

through the fact that $y \in N_S(x) \iff x \in S,\ y \in S^\perp \iff x \in N_{S^\perp}(y)$, and on the other hand, $N_{C\cap S}(x) \supset N_C(x) + N_S(x)$. We call (1.4) a *linkage variational inequality in coupled form* and call (1.1) as specialized to (1.2) the associated *linkage variational inequality in decoupled form*, when viewed as a condition on $\bar{x}$ for some $\bar{y} \in [F + N_C](\bar{x})$. This variational inequality version of the present topic was taken up in the conference paper [24] with examples of what decoupling could offer in practical advantages. We'll come back to such examples later in a wider context in Sections 4 and 5.

When $T = \partial\varphi$ for a lower semicontinuous (lsc) function $\varphi : H \to (-\infty, \infty]$, not necessarily convex,[3] we speak of the *optimization case* of (1.1). It goes hand in hand with the problem

$$\text{minimize } \ \varphi(x) \ \text{ subject to } \ x \in S, \ \text{ or equivalently, minimize } \ \varphi + \delta_S \ \text{ over } H, \tag{1.5}$$

where δ_S denotes the indicator of S (the function having 0 as its value on S but ∞ outside of S). The fundamental first-order necessary condition for local optimality of $\bar{x}$ in (1.5) is $0 \in \partial(\varphi + \delta_S)(\bar{x})$, but $\partial(\varphi + \delta_S)(\bar{x}) \subset \partial\varphi(\bar{x}) + \partial\delta_S(\bar{x})$ under various constraint qualifications [29, 10.9], and solving $0 \in \partial\varphi(\bar{x}) + \partial\delta_S(\bar{x})$ is the same as (1.1) for $T = \partial\varphi$, inasmuch as $\partial\delta_S(\bar{x}) = N_S(\bar{x})$. Thus, (1.1) for $T = \partial\varphi$ constitutes a first-order necessary condition for local optimality in (1.5) in widespread circumstances. Moreover it stands for global optimality when φ is convex. Other instances of (1.1) connected with optimization come up in primal-dual first-order optimality conditions involving Lagrange multipliers and in characterizations of game-like equilibrium in multi-agent optimization.

The optimization case of the linkage problem covers vastly more territory than might at first be apparent. Because every element of H can be expressed uniquely as $x + u$ for some $x \in S$ and $u \in S^\perp$, (1.5) can be posed equivalently as

$$\text{minimize } \ f(x,u) = \varphi(x+u) \ \text{ in } \ (x,u) \in S \times S^\perp \ \text{ subject to } \ u = 0. \tag{1.6}$$

This connects it with the general theory of Lagrangians and duality in [19, 21, 29] in which a problem in a variable x is embedded in a family of problems parameterized by a "perturbation" variable u. In fact, (1.6) is fully able to support that format, inasmuch as any problem of minimizing $f(x,u)$ on a product Hilbert space $H = H_1 \times H_2$ subject to $u = 0$ can be written this way with $S = H_1 \times \{0\}$ and $S^\perp = \{0\} \times H_2$. The rule for subgradients in (1.6) is

$$(v,y) \in \partial f(x,u) \quad \iff \quad v + y \in \partial\varphi(x+u), \quad \text{where } \ ((x,u),(v,y)) \in S \times S^\perp, \tag{1.7}$$

so the linkage problem (1.1) for $T = \partial\varphi$ corresponds in (1.6) to solving the first-order optimality condition $(0, \bar{y}) \in \partial f(\bar{x}, 0)$. Karush-Kuhn-Tucker relations and very much more can be cast in this pattern through the particular choice of f or φ.

A linkage subspace S within a Hilbert space H can arise in many ways, as the preceding discussion ought to make clear. However, particular applications may dictate particular choices. For our approach here, an especially motivating example of S is the nonanticipativity subspace in multistage stochastic programming and stochastic variational inequalities, which dictates that decisions can only depend on information already known, not on information yet to be revealed in the future [28, 30]. Kirchoff's laws for flows in a network might also be a source of linkage, as could be the relationships between the

[3] The finite-dimensional theory of subgradients of nonconvex functions is furnished in [29], and an infinite-dimensional version is in [15].

quantities in different time periods in a problem with deterministic dynamics. In other situations S may be introduced for technical reasons. That will be seen in the applications to splitting which we'll develop in Section 4 along with block-separability-based problem decomposition, thereby extending similar but more limited applications of such kind which were made in [24] and [32].

For the workability of the general linkage problem (1.1) as a model for which solution methods can be designed, we suppose that the projection mappings

$$P = P_S : H \to S \;\text{ and }\; P^\perp = P_{S^\perp} : H \to S^\perp$$

are relatively convenient to execute. Convenience for one entails convenience for the other, since $P + P^\perp = I$. The representation of each element of H as the sum of its projections on S and $S^\perp$ was central to Spingarn's "partial inverse" developments in [31], where the importance and versatility of problem (1.1) were first emphasized. But he required maximal monotonicity of T in the sense pioneered by Minty [14], whereas a major aim here is to get away from having to insist on that.

Some background on monotonicity will anyway help in what is coming. Recall that a mapping $T : H \rightrightarrows H$ is *monotone* (globally) if

$$\langle y' - y, x' - x \rangle \geq 0 \;\text{ when }\; y \in T(x),\; y' \in T(x'). \tag{1.8}$$

It is *maximal monotone* if furthermore the graph of T, namely

$$\operatorname{gph} T = \{\, (x, y) \in H \times H \,|\, y \in T(x)\},$$

can't be extended by adjoining another pair (x, y) without violating (1.8), or in other words if there is no monotone mapping T' with $\operatorname{gph} T' \supset \operatorname{gph} T$ and $\operatorname{gph} T' \neq \operatorname{gph} T$. There is also *strong* monotonicity, which replaces the 0 in (1.8) by $\sigma||x' - x||^2$ for some $\sigma > 0$, and similarly the maximal version of that. Strong monotonicity at level σ is equivalent to the monotonicity of $T - \sigma I$. For localized versions of these concepts we can take a cue from Pennanen [16] and speak of T being monotone or maximal monotone *in some subset* of $H \times H$ if these properties hold relative to that subset.

The subgradient mappings $\partial\varphi$ associated with lower semicontinuous proper *convex* functions φ on H provide prime examples of maximal monotone T [20], as do the continuous linear mappings $A : H \to H$ whose *symmetric* part is positive semidefinite. It's crucial to appreciate, though, that localized maximal monotonicity of a subgradient mapping $\partial\varphi$ doesn't refer to just to localization in the effective domain of φ (although it could), for the reason that the localization is articulated in $H \times H$, not simply H. More about this will come shortly.

In the case of $T = F + N_C$ for a closed convex set $C \subset$ and a continuous mapping $F : C \to H$, maximal monotonicity follows from F being monotone relative to C, i.e.,

$$\langle F(x') - F(x), x' - x \rangle \geq 0 \;\text{ for }\; x, x' \in C. \tag{1.9}$$

Likewise for strong monotonicity. When (1.9) holds on $U \cap C$ for some open convex subset U of H, then $F + N_C$ is maximal monotone in $U \times H$. The optimization case of $T = F + N_C$ corresponds to having $F = \nabla f$ for a continuously differentiable function f on C. Then $T = \partial\varphi$ for $\varphi = f + \delta_C$.

With these facts at our disposal, we can proceed towards the statement of the algorithm featured in this paper. An observation about the role of the subspace $S^\perp$ in the linkage problem (1.1) will guide us in passing beyond surface manifestations of monotonicity:

$$\begin{array}{l}\text{the pairs } (\bar{x}, \bar{y}) \text{ that solve (1.1) remain the same} \\ \text{when } T \text{ is replaced by } T + eP^\perp \text{ for some } e \geq 0,\end{array} \tag{1.10}$$

which holds because having $\bar{x} \in S$ corresponds to $P^\perp(\bar{x}) = 0$. We have the freedom to select the value of the elicitation parameter e in order to enhance properties of T for algorithmic purposes.

Definition 1: elicitation, global and local. *Maximal monotonicity will be said to be elicitable globally at a level $e \geq 0$ if the mapping $T + eP^{\perp} : H \rightrightarrows H$ is maximal monotone. It will be said to be elicitable locally at a level $e \geq 0$ with respect to a pair $(\bar{x}, \bar{y})$ solving (1.1) if $T + eP^{\perp}$, which likewise has $(\bar{x}, \bar{y})$ in its graph, is maximal monotone in a neighborhood of $(\bar{x}, \bar{y})$. The elicitability of strong maximal monotonicity, whether global or local, is defined in parallel to this.*

Elicitation at level $e = 0$ is important in this definition as a means of covering *monotone* cases of the linkage problem within the results that we'll obtain. Although we looked at the globally monotone case in [24] for T having the special form in (1.2), we didn't treat the locally monotone case there at all. Both cases are now going to be encompassed for general T.

Of course if monotonicity is elicitable at a level e it is elicitable at all higher levels e as well.[4] This way of identifying "hidden monotonicity" is entirely different from the tactic of Pennanen in [16], which amounts instead to adding some multiple of I to T^{-1}. Maybe his device can somehow be combined with ours to gain further traction, but that won't be studied in this paper.

In the optimization case with $T = \partial\varphi$, elicitability of monotonicity aspects of T translates to elicitability of convexity aspects of φ through the observation that

$$\begin{aligned} &T = \partial\varphi \;\longleftrightarrow\; T + eP^{\perp} = \partial[\varphi + \tfrac{e}{2} d_S^2] \;\text{ where} \\ &d_S(x) = [\text{ distance of } x \text{ from } S\ ] = ||P^{\perp}(x)||. \end{aligned} \tag{1.11}$$

Global elicitability of monotonicity corresponds then to global elicitation of convexity in the sense of $\varphi + \frac{e}{2} d_S^2$ being convex.[5] Local convexity of $\varphi + \frac{e}{2} d_S^2$ on a neighborhood of a solution $\bar{x}$ similarly elicits monotonicity locally with respect to $(\bar{x}, \bar{y})$ for $\bar{y} \in \partial\varphi(\bar{x})$, but that criterion is only sufficient, not necessary. A necessary and sufficient criterion for local elicitation must take into account the localization around $\bar{y}$ as well as the localization around $\bar{x}$. We'll come back to this in Section 3 with the insight that such local elicitation amounts to a kind of *sufficient condition for local optimality* in the associated problem of minimizing φ over the subspace S. It relates to a property called *variational convexity* which we investigated recently in [25]. However, a simple example may help by providing a glimpse of that forthcoming picture of local optimality.

Example: elicitation in smooth optimization. *Let H be finite-dimensional and $T = \nabla\varphi$ for a $\mathcal{C}^2$ function φ, so the linkage problem comes down to finding $\bar{x} \in S$ such that $\nabla\varphi(\bar{x}) \perp S$ (with $\nabla\varphi(\bar{x})$ thus being the $\bar{y} \in S^{\perp}$). This means $\bar{x}$ satisfies the first-order necessary condition for minimizing φ over S. The associated second-order sufficient condition for a local minimum at $\bar{x}$ is the positive definiteness of the Hessian $\nabla^2\varphi(\bar{x})$ relative to S. That's actually equivalent to the existence of $e \geq 0$ such that the Hessian of $\varphi + \frac{e}{2} d_S^2$ is positive definite on H, which corresponds to $\varphi + \frac{e}{2} d_S^2$ being strongly convex around $\bar{x}$ and thus to the local elicitation of strong monotonicity of $T = \nabla\varphi$ around $(\bar{x}, \bar{y})$.*

This situation will be handled in a much broader way in Section 3 without demanding smoothness of φ; see the corollary to Theorem 6 and beyond. For now, we move on to describing our solution method that's able to make use of elicitation of monotonicity or convexity when needed.

Progressive Decoupling Algorithm. *With respect to an elicitation level $e \geq 0$ (global or local) and a proximal parameter $r > e$, proceed as follows. In iteration ν, having $x^{\nu} \in S$ and $y^{\nu} \in S^{\perp}$, determine $\widehat{x}^{\nu} \in H$ as a solution to the generalized equation*

$$0 \in T^{\nu}(\widehat{x}^{\nu}), \;\; \textit{where} \;\; T^{\nu}(x) = T(x) - y^{\nu} + r[x - x^{\nu}]. \tag{1.12}$$

[4] The sum of a maximal monotone mapping and a continuous monotone mapping is again maximal monotone.

[5] A subgradient mapping is maximal monotone if and only if the function is convex lower semicontinuous and proper (i.e., $\not\equiv \infty$ and never $= -\infty$). See [17] for the finite-dimensional case and [2, Theorem 5.6] for the infinite-dimensional case; also [29, Theorem 12.17].

From $\widehat{x}^\nu$ *then get* $x^{\nu+1} \in S$ *and* $y^{\nu+1} \in S^\perp$ *by*

$$x^{\nu+1} = P(\widehat{x}^\nu), \qquad y^{\nu+1} = y^\nu - (r-e)P^\perp(\widehat{x}^\nu) = y^\nu - (r-e)[\widehat{x}^\nu - x^{\nu+1}]. \tag{1.13}$$

The virtue of this procedure is that it addresses problem (1.1) through iterations solving only the generalized equations in (1.12), *which omit the linkage subspace* S.

Theorem 1: convergence.

(a) *Suppose the linkage problem (1.1) is solvable, and that* $e \geq 0$ *gives a level at which maximal monotonicity is elicited globally. Then the iterations (1.12)–(1.13) for any* $r > e$*, starting from any* $x^0 \in S$ *and* $y^0 \in S^\perp$*, will generate a sequence of pairs* (x^ν, y^ν) *that converges (in the weak topology of* H*) to a pair* $(\bar{x}, \bar{y})$ *solving (1.1). Moreover, this will happen with the intermediate vectors* $\widehat{x}^\nu$ *being uniquely determined as solutions to the generalized equations (1.12) and with*

$$\begin{array}{c} ||(x^{\nu+1}, y^{\nu+1}) - (\bar{x}, \bar{y})||_{r,e} \;\leq\; ||(x^\nu, y^\nu) - (\bar{x}, \bar{y})||_{r,e} \;\; \text{for all } \nu, \\ \text{where } \; ||(x,y)||^2_{r,e} = ||x||^2 + \frac{1}{r(r-e)}||y||^2 \;\; \text{for } \; x \in S, \; y \in S^\perp. \end{array} \tag{1.14}$$

(b) *If maximal monotonicity is just elicited locally at level* $e \geq 0$ *relative to a solution pair* $(\tilde{x}, \tilde{y})$ *in (1.1), the same convergence will hold if the iterations start with* $x^0 + y^0$ *close enough to* $\tilde{x} + \tilde{y}$*,*[6] *and localization is attended to in getting the intermediate vectors* $\widehat{x}^\nu$*. Specifically, there will be a neighborhood* W *of* $\tilde{x} + \tilde{y}$ *such that, when* $x^\nu + y^\nu \in W$*, a unique solution* $\widehat{x}^\nu$ *to (1.12) will lie in* W*, and the* $x^{\nu+1}$ *and* $y^{\nu+1}$ *coming from it by the update rule in (1.13) will then have* $x^{\nu+1} + y^{\nu+1} \in W$*.*

(c) *If the locally elicited monotonicity is strong with modulus* $\sigma > 0$*, then* $\bar{x} = \tilde{x}$*, because locally there is only one possible* x*-component to a solution to the linkage problem. Furthermore, the convergence will ultimately follow the pattern that*

$$||x^{\nu+1} - \bar{x}|| \leq ||(x^{\nu+1}, y^{\nu+1}) - (\bar{x}, y^\nu)||_{r,e} \leq \frac{r}{r+\sigma}||(x^\nu, y^\nu) - (\bar{x}, \bar{y})||_{r,e}. \tag{1.15}$$

The validity of these convergence claims will be confirmed in Section 2 as a consequence of the way we derive the progressive decoupling algorithm from the proximal point algorithm. It will thereby be apparent in the case of global elicitation in (a) that if the problem doesn't have a solution the sequence generated by the iterations must diverge (since the proximal point algorithm has this feature). Most of our attention however will be on local convergence under assumptions about a local solution.

Drawing on results about convergence of the proximal point algorithm when iterations aren't carried out exactly, it would be possible to extend convergence of the progressive decoupling algorithm to a version in which the subproblems in (1.12) aren't solved exactly. However we forgo the exercise of filling in such details here because they would tend to clutter the simple picture of the procedure. Progressive decoupling may work better in some situations than in others, depending on problem structure, and exploring that is a higher priority at this stage, in our view, than pursuing convergence refinements.

For the same reason we hold back also from trying to include numerical results in this paper. There are too many different things to experiment with. Anyway, the progressive decoupling algorithm extends the progressive hedging algorithm [28] in stochastic optimization, which already has a long history of numerical experience; cf. [33] and its references, and more recently [26, 27]. Furthermore, in the case where monotonicity doesn't need to be elicited, i.e., where $e = 0$ works, the procedure is kin to many of analogous character in convex optimization. For more on that, see [24].

[6] the limit $(\bar{x}, \bar{y})$ may be different from $(\tilde{x}, \tilde{y})$.

An issue with progressive hedging in the past, which will carry over also to implementations of progressive decoupling, is selecting the right size for the proximal parameter r. Performance may deteriorate if it is either too high or too low. Little guidance in theory is available for coping with that, although heuristics have been explored numerically; again see [33] and [26], for instance.[7] A new feature here, not present in progressive hedging but which now could be added to that, is extra flexibility coming from the elicitation parameter e. Even when no elicitation of monotonicity is needed (or of convexity in connection with optimization), so that $e = 0$ would do, there might be an advantage in taking $e > 0$. Clues about that can be detected in (1.14), which indicates how the choice of r and e may affect the convergence of the sequences of primal elements x^ν and dual elements y^ν relative to each other. The choices of r and e allow the coefficient $1/r(r-e)$ to range all the way between 0 and ∞. Clearly a low level of this coeffient would favor convergence of x^ν to $\bar{x}$ over convergence of y^ν to $\bar{y}$, while a high level would do the opposite. When the parameters are coupled with respect to the choice of $e > 0$ by

$$r = \sqrt{\left[\frac{e}{2}\right]^2 + 1} + \frac{e}{2}, \;\; \text{so that} \;\; r - e = \sqrt{\left[\frac{e}{2}\right]^2 + 1} - \frac{e}{2}, \tag{1.16}$$

then $r(r-e) = 1$ and the convergence properties in the theorem hold with respect to the norm $||(x,y)||_{r,e}$ in (1.14) simply being $||x+y||$ for $x \in S$ and $y \in S^\perp$. Then neither sequence is favored over the other.

Although the $\bar{x}$ part of a solution $(\bar{x}, \bar{y})$ is unique in the strong elicitation part of the Theorem 1, the $\bar{y}$ part need not be. The implications of the property in (1.15) that's associated with this aren't fully clear, but the indication seems to be that once y^ν is very near or maybe even equal to $\bar{y}$, a hint of linear convergence emerges in x^ν.

In the optimization case $T = \partial\varphi$, the generalized equations (1.12) to be solved in the algorithm's iterations specialize to

$$0 \in \partial\varphi^\nu(\widehat{x}^\nu), \;\; \text{where} \;\; \varphi^\nu(x) = \varphi(x) - \langle y^\nu, x\rangle + \frac{r}{2}||x - x^\nu||^2. \tag{1.17}$$

The direct interpretation of the step in (1.12) is that $\widehat{x}^\nu$ is to be determined as a sort of critical point of the modified function φ^ν. But can that be accomplished by an optimization procedure? More importantly, does $\widehat{x}^\nu$ in fact give a local minimum of φ^ν? That's obviously true when φ^ν is convex, at least locally around $\widehat{x}^\nu$, as would follow from φ being convex locally, but the general circumstances of elicitation may be more subtle than that. Nevertheless minimization can indeed be the rule quite broadly, as will be explained in Theorem 10 in Section 3 in connection with the new sufficient condition for local optimality that's tied to local elicitation of monotonicity in a subgradient mapping.

2 Justification of the algorithm and its convergence

The progressive decoupling algorithm will be derived from the proximal point algorithm [22] in a manner similar to Spingarn's derivation of his method of partial inverses [31], but localization ideas of Pennanen in [16] will enter along with a delicate kind of rescaling. The proximal point algorithm aims at finding $\bar{x}$ such that $0 \in M(\bar{x})$ for a maximal monotone mapping $M : H \rightrightarrows H$ through iterations taking the form:

$$\text{get } x^{\nu+1} \text{ from } x^\nu \text{ by solving } 0 \in M^\nu(x^{\nu+1}), \text{ where } M^\nu(x) = M(x) + r[x - x^\nu] \tag{2.1}$$

[7] Similar issues have also been confronted in other related methodology, as for instance in [4].

with $r > 0$ as parameter. A reminder of its properties will help to set the stage for what comes next.

The attractiveness of the iterations (2.1) when M is maximal monotone is that $x^{\nu+1}$ always exists and is uniquely determined. Although (2.1) just asks for $x^{\nu+1}$ to be an element of $(I + r^{-1}M)^{-1}(x^\nu)$, maximal monotonicity (globally) ensures that the potentially multivalued mapping $(I + r^{-1}M)^{-1}$ is single-valued and even nonexpansive, i.e. globally Lipschitz continuous with Lipschitz constant 1:

$$||(I + r^{-1}M)^{-1}(x') - (I + r^{-1}M)^{-1}(x)|| \leq ||x' - x|| \text{ for all } x, x'. \tag{2.2}$$

The set of fixed points of $(I+r^{-1}M)^{-1}$ is the solution set $M^{-1}(0)$, so a consequence of (2.2), specialized to $x' = x^\nu$, is that

$$||x^{\nu+1} - x|| \leq ||x^\nu - x|| \text{ for all } x \in M^{-1}(0). \tag{2.3}$$

As shown in [22], the generated sequence of vectors x^ν is bounded if and only if $M^{-1}(0) \neq \emptyset$, and it then must converge (in the weak topology of H) to a particular $\bar{x} \in M^{-1}(0)$. Furthermore, if the mapping M^{-1} itself happens to be single-valued and Lipschitz continuous on a neighborhood of 0, the convergence is sure to be at a linear rate; this linear convergence criterion was refined by Luque [12].

Also explained in [22] are conditions for convergence when the subproblems in (2.1) are solved "inexactly." That topic has advanced further through the work of Eckstein and Bertsekas [5] and others; see Pennanen [16] for comprehensive references, also to conditions under which convergence can be guaranteed locally even for nonmonotone M in some circumstances. However, those advances in [5, 16] seem incompatible with the special implementation we'll need to make of the proximal point algorithm to achieve "decoupling."

A result of Pennanen in [16] on local convergence in presence of locally maximal monotonicity will, on the other hand, be essential to us in connection with our elicitation scheme. Here we provide an improved statement of it with a simpler proof.

Theorem 2: localization. *Suppose there exists $\tilde{x} \in M^{-1}(0)$ such that the mapping $M : H \rightrightarrows H$ is maximal monotone in a neighborhood of $(\tilde{x}, 0)$. Then there is a neighborhood W of $\tilde{x}$ such that in the iterations (2.1) of the proximal point algorithm, as long as $x^\nu \in W$, there exists a unique $x^{\nu+1} \in W$ satisfying $0 \in M^\nu(x^{\nu+1})$. The sequence generated this way in W converges to some $\bar{x} \in M^{-1}(0)$ just as if M were maximal monotone globally.*[8]

Proof. Let M_0 be a localization of M obtained by restricting the graph of M to a neighborhood of $(\tilde{x}, 0)$, and let $\overline{M}_0$ be a maximal monotone extension of M_0 (which is always known to exist; see [29, 12.6]). The proximal point algorithm can be executed for $\overline{M}_0$ and, from any starting point, will converge to some point of $\overline{M}_0^{-1}(0)$ with its iterations satisfying

$$||x^{\nu+1} - x|| \leq ||x^\nu - x|| \text{ for all } x \in \overline{M}_0^{-1}(0).$$

Therefore, if the procedure is initiated in some closed ball of radius δ around $\tilde{x}$ it must forever stay in that ball and reach a limit $\bar{x}$ also in that ball. According to (2.1) for $\overline{M}_0$, the iterates involve pairs $(x^{\nu+1}, r(x^\nu - x^{\nu+1})) \in \operatorname{gph} \overline{M}_0$ which converge to $(\bar{x}, 0)$. By choosing the δ-ball around $\tilde{x}$ small enough, we can thus ensure that all these pairs lie in the neighborhood of $(\tilde{x}, 0)$ to which we localized in obtaining M_0. In that neighborhood, however, there's no distinction betwen $\overline{M}_0$ and M_0, or between M_0 and the original M. All that matters is that we stick to the unique iterates that would have come from $\overline{M}_0$, and that can be done by not straying out of a neighborhood of $\tilde{x}$ that's small enough to start with. □

[8] The reason why convergence might be to $\bar{x} \neq \tilde{x}$ is evident from the case where M is maximal monotone globally and the solution set $M^{-1}(0)$ is more than a singleton.

A shortcoming of this localization result, of course, is that it doesn't provide any criterion for knowing whether a small enough neighborhood of a particular, but unknown, solution has been reached and how to be sure of staying within it. This will need to be studied more, but we won't do that here. Conditions for executing the localized version of the proximal point algorithm without losing convergence have been worked out in [10].

Next we make note of a rescaling device which adds flexibility to the proximal point algorithm and will be essential in our decoupling scheme.

Theorem 3: rescaling. *Let J be a (continuous) linear mapping from H onto H. Replace the rule for proximal point iterations in (2.1) by*

$$0 \in M_J^\nu(x^{\nu+1}), \quad \text{where } M_J^\nu(x) = M(x) + rJ^*J[x - x^\nu]. \tag{2.4}$$

Then all the same convergence properties hold except that the norm $||x||$ is supplanted in the distance estimates by the norm $||x||_J = ||Jx||$.

Proof. Because J maps H one-to-one onto H, its inverse J^{-1} exists as a continuous linear mapping from H onto H. With respect to the change of variables $z = Jx$ the rule in (2.4) can be written as $0 \in M(J^{-1}z^{\nu+1}) + rJ^*J[J^{-1}z^{\nu+1} - J^{-1}z^\nu]$, or in other words $0 \in J^{*-1}M(J^{-1}z^{\nu+1}) + r[z^{\nu+1} - z^\nu]$. These iterations correspond to executing the proximal point algorithm on the composed mapping $M' = J^{*-1}TJ^{-1}$, which inherits maximal monotonicity from M (globally or locally). They get $z^{\nu+1}$ from z^ν by applying $(I + r^{-1}M')^{-1}$, and the convergence properties are thereby tied to distances $||z' - z||$, which translate back under the change of variables to distances $||Jx' - Jx|| = ||x' - x||_J$. This means that convergence is governed quantitatively by the norm $||\cdot||_J$ as claimed. □

We are ready now to demonstrate that the progressive decoupling algorithm can be obtained by applying the proximal point algorithm, with a change of variables, to the partial inverse with respect to S (in Spingarn's sense [31]) of the mapping $T + eP^\perp$ for an elicitation level $e \geq 0$. Crucial to this is the fact that every element of H can be written in one and only one way as the sum of an element of S and an element of $S^\perp$, namely its projections on those two subspaces:

$$x = P(x) + P^\perp(x) \text{ with } \langle P(x), P^\perp(x)\rangle = 0.$$

Theorem 4: algorithm derivation. *Let $M : H \rightrightarrows H$ be the mapping obtained from $T + eP^\perp$ by*

$$v + u \in M(x + y) \iff v + y \in (T + eP^\perp)(x + u) \text{ when } x, v \in S,\ u, y \in S^\perp, \tag{2.5}$$

so that solving the linkage problem (1.1) is equivalent to finding an element of $M^{-1}(0)$, since

$$\text{for } \bar{x} \in S,\ \bar{y} \in S^\perp: \qquad 0 \in M(\bar{x} + \bar{y}) \iff \bar{y} \in T(\bar{x}). \tag{2.6}$$

Let J for $r > e$ be the continuous linear mapping from H onto H such that

$$\text{for } x \in S,\ y \in S^\perp: \quad J \text{ takes } x + y \text{ to } x + [r(r - e)]^{-1/2}y. \tag{2.7}$$

The iterations (1.12)–(1.13) of the progressive decoupling algorithm correspond then to applying the proximal point algorithm to M as modified by J in the manner of (2.4).

Proof. Here $J^* = J$, so J^*J is simply the mapping which, in line with (2.7), takes $x + y$ to $x + r^{-1}(r - e)^{-1}y$. The iterations in (2.4), when expressed compatibly with (2.5) in terms of the elements of H being represented by their projections on S and $S^\perp$, follow the pattern that

$$0 \in M(x^{\nu+1} + y^{\nu+1}) + r[x^{\nu+1} - x^\nu] + (r - e)^{-1}[y^{\nu+1} - y^\nu].$$

In terms of letting

$$u^{\nu+1} = -(r-e)^{-1}[y^{\nu+1} - y^\nu] \in S^\perp, \;\text{ hence }\; y^{\nu+1} = y^\nu - (r-e)u^{\nu+1} \in S^\perp, \tag{2.8}$$

we can rewrite this as $-r[x^{\nu+1} - x^\nu] + u^{\nu+1} \in M(x^{\nu+1} + y^{\nu+1})$. Translating through (2.5) we get

$$-r[x^{\nu+1} - x^\nu] + y^{\nu+1} \in (T + eP^\perp)(x^{\nu+1} + u^{\nu+1}), \tag{2.9}$$

where $P^\perp(x^{\nu+1} + u^{\nu+1}) = u^{\nu+1}$, or in other words

$$0 \in T(x^{\nu+1} + u^{\nu+1}) + r[x^{\nu+1} - x^\nu] + eu^{\nu+1} - y^{\nu+1}.$$

In appealing to (2.8) to identify $eu^{\nu+1} - y^{\nu+1}$ with $-y^\nu + ru^{\nu+1}$, we can put this in the form

$$0 \in T(x^{\nu+1} + y^{\nu+1}) - y^\nu + r[x^{\nu+1} - x^\nu + u^{\nu+1}]. \tag{2.10}$$

Now let $\widehat{x}^\nu$ denote $x^{\nu+1} + u^{\nu+1}$, so that $P(\widehat{x}^\nu) = x^{\nu+1}$ and $P^\perp(\widehat{x}^\nu) = u^{\nu+1}$. Then through (2.8) we have the update step (1.13), while (2.10) becomes the subproblem step (1.12). Thus, the proximal point iterations in the pattern described reduce to those of the progressive decoupling algorithm, as we needed to confirm. □

This derivation provides the background for verifying the facts about the convergence of the progressive decoupling algorithm that were presented in Section 1.

Proof of Theorem 1 on convergence. The elicited global maximal monotonicity of $T + eP^\perp$ carries over in the partial inverse operation (2.5) to that of the mapping M to which the proximal point algorithm is applied — with the rescaling modification in Theorem 3 for the J in (2.7) — to get the progressive decoupling algorithm. The convergence properties of the proximal point algorithm translate that way, in terms of the norm in Theorem 3, to the convergence properties claimed here. When the elicited maximal monotonicity is only local, the same properties hold locally on the basis of the Theorem 2. The additional facts claimed in part (c) of the theorem must still be confirmed. Let $(\bar{x}, \bar{y})$ and $(\tilde{x}, \tilde{y})$ solve the linkage problem (1.1), which case also $(\bar{x}, \bar{y}) \in \text{gph}[T + eP^\perp]$ and $(\tilde{x}, \tilde{y}) \in \text{gph}[T + eP^\perp]$ for any $e \geq 0$. Suppose that $T + eP^\perp$ is maximal monotone strongly with modulus $\sigma > 0$ in a neighborhood of $(\tilde{x}, \tilde{y})$, and that $(\bar{x}, \bar{y})$ lies in that same neighborhood. The strong monotonicity implies that

$$\langle \bar{y} - \tilde{y}, \bar{x} - \tilde{x} \rangle \geq \sigma|\bar{x} - \tilde{x}||^2,$$

but the left side of this inequality is 0 because $\bar{x} - \tilde{x} \in S$ while $\bar{y} - \tilde{y} \in S^\perp$. This necessitates $\bar{x} - \tilde{x} = 0$.

Next, to get the local convergence properties in (1.15), we appeal to (2.9) along with $\bar{y} \in T(\bar{x})$ and the strong monotonicity in the neighborhood of $(\tilde{x}, \tilde{y}) = (\bar{x}, \bar{y})$ in which the limit $(\bar{x}, \bar{y})$ must lie, to get

$$\langle -r[x^{\nu+1} - x^\nu] + y^{\nu+1} - \bar{y}, x^{\nu+1} + u^{\nu+1} - \bar{x} \rangle \geq \sigma||x^{\nu+1} + u^{\nu+1} - \bar{x}||^2.$$

Because the x elements lie in S while the y elements and $u^{\nu+1}$ lie in $S^\perp$, and because of the expression for $y^{\nu+1}$ in (2.8), this is the same as

$$r\langle (x^\nu - \bar{x}) - (x^{\nu+1} - \bar{x}), x^{\nu+1} - \bar{x} \rangle + \langle y^\nu - \bar{y} - (r-e)u^{\nu+1}, u^{\nu+1} \rangle \geq \sigma||x^{\nu+1} - \bar{x}||^2 + \sigma||u^{\nu+1}||^2,$$

which can be rewritten as

$$r\langle x^\nu - \bar{x}, x^{\nu+1} - \bar{x} \rangle + \langle y^\nu - \bar{y}, u^{\nu+1} \rangle \geq (r+\sigma)||x^{\nu+1} - \bar{x}||^2 + (r - e + \sigma)||u^{\nu+1}||^2.$$

Since $(r-e+\sigma)/(r+\sigma) \geq (r-e)/r$ for $\sigma > 0$, we can process this further in terms of letting $\theta = \sqrt{(r-e)/r}$ (and appealing to the orthogonality between the x elements and the y and u elements) as

$$\begin{aligned}(r+\sigma)||(x^{\nu+1}-\bar{x})+\theta u^{\nu+1}||^2 &\leq r\langle (x^\nu-\bar{x})+r^{-1}\theta^{-1}(y^\nu-\bar{y}), (x^{\nu+1}-\bar{x})+\theta u^{\nu+1}\rangle \\ &\leq r||(x^\nu-\bar{x})+r^{-1}\theta^{-1}(y^\nu-\bar{y})|| \cdot ||x^{\nu+1}-\bar{x})+\theta u^{\nu+1}||,\end{aligned}$$

from which it follows that

$$||(x^{\nu+1}-\bar{x})+\theta u^{\nu+1}|| \leq \frac{r}{r+\sigma}||(x^\nu-\bar{x})+r^{-1}\theta^{-1}(y^\nu-\bar{y})||. \tag{2.11}$$

Here we can use (2.8) to turn $\theta u^{\nu+1}$ into $\theta(r-e)^{-1}(y^\nu - y^{\nu-1})$. Then from the observation that $\theta)(r-e)^{-1} = 1/\sqrt{r(r-e)}$ and on the other hand $r^{-1}\theta^{-1} = 1/\sqrt{r(r-e)}$, we can translate (2.11) into

$$||(x^{\nu+1}-\bar{x}) - \sqrt{r(r-e)}^{-1}(y^{\nu+1}-y^\nu)|| \leq \frac{r}{r+\sigma}||(x^\nu-\bar{x}) + \sqrt{r(r-e)}^{-1}(y^\nu-\bar{y})||.$$

But this is the same as (1.15) according to the definition of the norm $||\cdot||_{r,e}$, with the initial inequality in (1.15) holding because of the orthogonality between the x and y elements. □

Remark on parameters. Theorem 3 has been used here only in a simple way. It could be exploited further to rescale different components of the vectors in different ways, and that might be helpful in particular applications.

3 Elicitability and its meaning in optimization

For the globally monotone case of the linkage problem (1.1), where T is maximal monotone and the elicitation parameter e can be taken to be 0, the results of the preceding section provide support for the progressive decoupling algorithm which goes beyond the basic variational inequality case $T = F + N_C$ in (1.2) which we treated in [24]. On the optimization side, the corresponding advance is that the minimization of a general lsc convex function φ has now been covered instead of just φ of the form $f + \delta_C$ for a differentiable convex function f. Moreover $e > 0$ can be brought in for its effects on convergence in these circumstances even if not really needed to elicit monotonicity or convexity.

The locally monotone case, where T is only asked to be maximal monotone with respect to a neighborhood of a solution pair $(\bar{x}, \bar{y})$, is new here for the progressive decoupling algorithm and is likewise fully covered at this stage of our developments in the present paper. However, what that means on the optimization side, when $T = \partial\varphi$, can't be appreciated without clarifying how local monotonicity of $\partial\varphi$ around a pair $(\bar{x}, \bar{y})$ in its graph relates to convexity properties of φ.

Here, of course, we are keenly interested also in the nonmonotone case of T, for which $e > 0$ is inescapable, and the potential applicability of that to nonconvex φ. It will be important therefore to gain an understanding of circumstances in which monotonicity can be elicited successfully and how that connects with optimization, where elicitation is tied to the pattern in (1.1) of replacing φ by $\varphi + \frac{e}{2}d_S^2$. We begin on a relatively elementary track with the following observation in the variational inequality setting,

Theorem 5: elicitation via linearity. *Let $T = A + N_C$ for a nonempty closed convex set $C \subset H$ and a (continuous) linear mapping $A : H \to H$ for which*[9]

$$\textit{there exists } \alpha > 0 \textit{ such that } \langle x, Ax\rangle \geq \alpha||x||^2 \textit{ for all } x \in S, \tag{3.1}$$

[9]In finite dimensions, just having $\langle x, Ax\rangle > 0$ for all nonzero $x \in S$ will guarantee this, of course.

and let

$$\beta = \tfrac{1}{2}||P[A+A^*]P^\perp||, \qquad \gamma = ||P^\perp A P^\perp||. \tag{3.2}$$

Then maximal strong monotonicity is elicited globally at level e if

$$e \,>\, e_0 = \alpha^{-1}\beta^2 + \gamma. \tag{3.3}$$

Proof. It suffices to demonstrate that the continuous linear mapping $A_e = A + eP^\perp$ is monotone relative to C when e is sufficiently high, since then $A_e + N_C$ is maximal monotone. In fact we will get the monotonicity of A_e on all of H.

Due to its linearity, the monotonicity of A_e corresponds to having $\langle x, A_e x\rangle \geq 0$ for all $x \in H$. With the values in (3.1) and (3.2) at our disposal we calculate that

$$\begin{aligned}\langle x, A_e x\rangle &= \langle x, Ax\rangle + e\langle x, P^\perp(x)\rangle = \langle\, [P(x)+P^\perp(x)], A[P(x)+P^\perp(x)]\,\rangle + e\langle P^\perp(x), P^\perp(x)\rangle \\ &= \langle P(x), AP(x)\rangle + \langle P(x), [A+A^*]P^\perp(x)\rangle + \langle P^\perp(x), AP^\perp(x)\rangle + e||P^\perp(x)||^2 \\ &\geq \alpha||P(x)||^2 - 2\beta||P(x)||||P^\perp(x)|| + (e-\gamma)||P^\perp(x)||^2\end{aligned}$$

through (3.1) and (3.2) when $x \in C$. Will it be true in terms of $\xi = ||P(x)||$ and $\eta = ||P^\perp(x)||$ that $\alpha\xi^2 - 2\beta\xi\eta + (e-\gamma)\eta^2 > 0$ for all $(\xi,\eta) \neq (0,0)$ when e is higher than the indicated e_0? This is the same as asking when the symmetric 2×2 matrix with diagonal entries α and $e-\gamma$ and off-diagonal entries $-\beta$ is positive definite when $e > e_0$. The trace-determinant test for positive semidefiniteness is whether $\alpha + e - \gamma > 0$ and $\alpha(e-\gamma) - \beta^2 > 0$. Since $\alpha > 0$, the second inequality can be written as (3.3). This implies the first inequality, written as $e > \gamma - \alpha$. Thus, the threshold in (2.14) guarantees the claimed maximal monotonicity. □

Corollary. *Suppose in (3.1) and (3.2) that A is symmetric. Then for $f(x) = \frac{1}{2}\langle x, Ax\rangle$ the function $\varphi = f + \delta_C$ has the property that $\varphi + \frac{e}{2}d_S^2$ is strongly convex on H at all levels of e satisfying (3.3).*

Proof. This corresponds of course to taking A in the theorem to be the Hessian of f. □

The corollary refers to the optimization case of the linkage problem (1.1) that corresponds to minimizing $\varphi = f + \delta_C$ over the subspace S, or in other words, minimizing $f(x)$ subject to $x \in C \cap S$.

Theorem 6: elicitation via derivatives. *Suppose that $T = F + N_C$ for a nonempty closed convex subset $C \subset H$ and a continuously (Fréchet) differentiable mapping $F : C \to H$ with Jacobians $\nabla F(x)$ (as linear mappings from H to H).*

(a) (global elicitation) *Assume*[10]

$$\text{there exists } \alpha(x) > 0 \text{ such that } \langle x', \nabla F(x)x'\rangle \geq \alpha(x)||x'||^2 \text{ for all } x' \in S \text{ when } x \in C, \tag{3.4}$$

and let $\beta(x)$ and $\gamma(x)$ be defined as in (3.2) with respect to $\nabla F(x)$ as A. Let

$$e_0 = \sup_{x \in C\cap S}\Big\{\, \alpha(x)^{-1}\beta^2(x) + \gamma(x)\Big\}. \tag{3.5}$$

If $e_0 < \infty$, then maximal strong monotonicity is elicited globally at all levels $e > e_0$.

(b) (local elicitation) *Let $\tilde{x}$ and $\tilde{y}$ solve the linkage problem (1.1) for $T = F + N_C$ and suppose that the condition in (3.4) is satisfied by $\nabla F(\tilde{x})$ as A. Then maximal strong monotonicity is elicited locally for all e sufficiently high with respect to $\tilde{x}$ and $\tilde{y}$.*

Proof. For (a) this follows directly from Theorem 5. For (b) more argument is needed. The condition assumed at $\tilde{x}$ carries over to a neighborhood of $\tilde{x}$ by the continuous dependence of $\nabla F(x)$ on x. The

[10]In finite dimensions this is implied for bounded C by having $\langle x', \nabla F(x)x'\rangle > 0$ for all nonzero $x' \in S$ when $x \in C$.

argument of Theorem 5 applies not only to $\nabla F(\tilde{x})$ as A but also to all those linear mappings $\nabla F(x)$ for x in that neighborhood. For each such x a threshold sufficient for monotonicity is available as in (3.3) in terms of values $\alpha(x) > 0$, $\beta(x)$ and $\gamma(x)$, which depend continuously on x. The continuity provides an upper bound e_0 to $\alpha(x)^{-1}\beta(x)^2 + \gamma(x)$ in some neighborhood U of $\tilde{x}$. Then, as long as $e > e_0$, all the linear mappings $\nabla F(x) + eP^\perp$ for $x \in U \cap C$ will be monotone.

Without loss of generality it can be supposed that U is convex and closed. Fix $e \geq e_0$ and let $F_e = F + eP^\perp$, so that $\nabla F_e(x) = \nabla F(x) + eP^\perp$ and is monotone for all $u \in U \cap C$. This implies, as follows, that F itself is monotone on $U \cap C$. In taking x_0 and x_1 in $U \cap C$ and letting $x_\tau = (1-\tau)x_0 + \tau x_1$, the claimed monotonicity corresponds in terms of the function $\theta(\tau) = \langle x_1 - x_0, F(x_\tau) - F(x_0)\rangle$ to having $\theta(1) \geq 0$. Here $\theta(0) = 0$, so we can be sure of having $\theta(1) \geq 0$ if $\theta'(\tau) \geq 0$ for $\tau \in (0,1)$. Since $\theta'(\tau) = \langle x_1 - x_0, \nabla F(x_\tau)(x_1 - x_0)\rangle$, this is a consequence of $\nabla F(x_\tau)$ being monotone.

The strong monotonicity of $F_e = F + eP^\perp$ on $U \cap C$ in combination with its continuity ensures that $F_e + N_{U\cap C}$ is maximal monotone strongly from H to H. This mapping is then maximal monotone relative to $V \times H$ with $V = \operatorname{int} U$, where its graph reduces to that of F. □

Corollary. *Let $f : H \to I\!R$ be twice continuously differentiable with gradients $\nabla f(x)$ and Hessians $\nabla^2 f(x)$ (as linear mappings from H to H). Then, in terms of $F = \nabla f$ and $\varphi = f + \delta_C$, part (a) of Theorem 6 furnishes a criterion for $\varphi + \frac{e}{2}d_S^2$ to be strongly convex globally on H, and hence for strong monotonicity of $T = \partial\varphi$ to be elicited globally. Part (b) furnishes one for $\varphi + \frac{e}{2}d_S^2$ to be strongly convex around a solution $\tilde{x}$, in which case the strong monotonicity of $T = \partial\varphi$ is elicited locally.*

The corollary again refers to the optimization case of (1.1) associated with minimizing $\varphi = f + \delta_C$ over S, which is the same as minimizing $f(x)$ subject to $x \in C \cap S$, but this time f isn't just quadratic. The local version of the corollary with C dropped off (i.e., taken to be all of H) covers, in particular, the example of elicitation presented in Section 1 just ahead of the progressive decoupling algorithm.

This is not the end of the story, because we have not yet really come to grips with monotonicity being elicited locally with respect to a pair $(\tilde{x}, \tilde{y})$ in the graph of T, whether of form $F + N_C$ or not, when a neighborhood of $\tilde{y}$ has to come into play alongside of a neighborhood of $\tilde{x}$. For that, our focus is going to be on the optimization case, where even the meaning of local elicitation at level $e = 0$ hasn't yet been clarified. A concept we introduced recently in [25] will have a central role.

Definition 2: variational convexity. *The lsc function $\varphi : H \to (-\infty, \infty]$ is variationally convex at $\bar{x}$ for $\bar{y} \in \partial\varphi(\bar{x})$ if for some open convex neighborhood $U \times V$ of $(\bar{x}, \bar{y})$ there is a convex lsc function $\psi \leq \varphi$ on U such that, for some $\varepsilon > 0$ and the set $U_\varepsilon = \{\, x \in U \,|\, \varphi(x) \leq \varphi(\bar{x}) + \varepsilon\}$,*

$$[U_\varepsilon \times V] \cap \operatorname{gph} \partial\varphi = [U \times V] \cap \operatorname{gph} \partial\psi \text{ and } \varphi(x) = \psi(x) \text{ at the elements } (x,y) \text{ of this set.} \tag{3.6}$$

It is variationally strongly convex at $\bar{x}$ for $\bar{y}$ if ψ is strongly convex.

These properties are best understood when $\bar{y} = 0$ (which we can revert to by passing from $\varphi(x)$ to $\varphi(x) - \langle \bar{y}, x\rangle$), since we are looking then at the first-order necessary condition for a local minimum: $0 \in \partial\varphi(\bar{x})$. Variational convexity in that case has $0 \in \partial\psi(\bar{x})$ for a convex function $\psi \leq \varphi$ with $\psi(\bar{x}) = \varphi(\bar{x})$. This implies that ψ is minimized locally by $\bar{x}$, and the same then for φ. Thus, in adding variational convexity to having $0 \in \partial\varphi(\bar{x})$, we get a *second-order sufficient condition* for a local minimum of φ at $\bar{x}$, even though φ might not itself be convex on a neighborhood of $\bar{x}$, and even $\operatorname{dom}\varphi$ could be nonconvex. But variational convexity says more than that, because it also concerns nearby pairs (x, y) in $\operatorname{gph} \partial\varphi$.

Variational convexity obviously entails prox-regularity and subdifferential continuity (since those hold for lsc convex functions). It has been investigated from several angles in [25], including connections with tilt stability. More on that will be said later.

Before getting any further with this concept, the definition of $\partial\varphi$ for potentially nonconvex φ needs to be reviewed. In the terminology of [29],[11] y is a *regular* subgradient of φ at x if

$$\varphi(x') \geq \varphi(x) + \langle y, x'-x\rangle + o(||x'-x||), \qquad \text{notation:} \quad y \in \widehat{\partial}\varphi(x). \tag{3.7}$$

It is a (*general*) subgradient of φ at x if

$$\exists y^\nu \in \widehat{\partial}\varphi(x^\nu) \ \text{ with } \ y^\nu \to y \ [\text{weakly}]\,,\ x^\nu \to x,\ \varphi(x^\nu) \to \varphi(x), \qquad \text{notation:} \quad y \in \partial\varphi(x). \tag{3.8}$$

When φ is convex, this two-step definition arrives at the usual subgradients of convex analysis which satisfy the inequality in (3.7) without the error term $o(||x'-x||)$. Beyond that, however, the limits in (3.8) can be essential, including the limit in function values, because the convergence of (x^ν, y^ν) to (x, y) in $\operatorname{gph}\widehat{\partial}\varphi$ need not always entail $\varphi(x^\nu) \to \varphi(x)$. This is reflected in the definition of variational convexity by putting U_ε in place of U at the beginning of (3.6).

An illuminating example of subgradients and variational convexity without actual convexity is offered by the function function $\varphi : I\!R \to I\!R$ defined by

$$\begin{array}{l} \varphi(x) = \max\{1-e^x, 1-e^{-x}\} \ \text{ with } \ \varphi(0) = 0 = \min\varphi \ \text{ and} \\ (x,y) \in \operatorname{gph}\partial\varphi = \operatorname{gph}\widehat{\partial}\varphi \quad \Longleftrightarrow \quad \begin{cases} x>0,\ y = e^{-x}, \\ x = 0,\ y \in [-1,1], \\ x<0,\ y = -e^{x}. \end{cases} \end{array} \tag{3.9}$$

This continuous function is strongly concave to the left of the origin and also to the right of the origin but nonetheless is variationally convex at 0 for $0 \in \partial\varphi(0)$, as seen by taking $\psi(x) = \tau|x|$ for any $\tau \in [0,1)$. If we modify it to

$$\begin{array}{l} \varphi(x) = 1-e^{-x} \ \text{ for } \ x \geq 0 \ \text{ but } \ \varphi(x) = 1 \ \text{ for } \ x<0, \ \text{ with} \\ (x,y) \in \operatorname{gph}\partial\varphi = \operatorname{gph}\widehat{\partial}\varphi \quad \Longleftrightarrow \quad \begin{cases} x>0,\ y = e^{-x}, \\ x = 0,\ y \in (-\infty,1], \\ x<0,\ y = 0, \end{cases} \end{array} \tag{3.10}$$

we get φ discontinuous but lsc at $x = 0$. But $x = 0$ still gives the global minimum and exhibits variational convexity with respect to the subgradient $y = 0$ there. We also have a local minimum with subgradient $y = 0$ at every $x < 0$, and variational convexity, even local convexity, of φ prevails there as well. However, this is different from the global minimum at the origin and needs to be kept apart from that. The way to keep it apart is to focus on the pairs $(x, y) \in \operatorname{gph}\partial\varphi$ for which $\varphi(x) \leq \varphi(0) + \varepsilon = \varepsilon$ for small enough ε, here $\varepsilon \in (0,1)$. Then there are no subgradients included for $x < 0$.

The incentive behind variational convexity is the following result which binds it to subgradient monotonicity.

Theorem 7: subgradient monotonicity. *For lsc $\varphi : H \to (-\infty,\infty]$ and $(\bar{x},\bar{y}) \in \operatorname{gph}\partial\varphi$, if φ is variationally convex at $\bar{x}$ for $\bar{y}$, then there is an open convex neighborhood $U \times V$ of $(\bar{x},\bar{y})$ along with some $\varepsilon > 0$ such that the mapping $T : H \rightrightarrows H$ defined by*

$$\operatorname{gph} T = \{\, (x,y) \in \operatorname{gph}\partial\varphi \,|\, \varphi(x) < \varphi(\bar{x}) + \varepsilon\} \tag{3.11}$$

[11] The terminology in [15], where infinite-dimensions are covered, is different: regular subgradients are "Fréchet" subgradients, while general subgradients are "limiting" subgradients.

is maximal monotone in $U \times V$. Moreover this implication is an equivalence when H is finite-dimensional and $\bar{y}$ is a regular subgradient of φ at $\bar{x}$, i.e., $\bar{y} \in \hat{\partial}\varphi(\bar{x})$. The same facts apply also to the relationship between variational strong convexity and local maximal strong monotonicity.

Proof. The implication from variational convexity to local maximal monotonicity is easy, because (3.6) allows $\mathrm{gph}\,\partial\varphi$ to be replaced in (3.11) by $\mathrm{gph}\,\partial\psi$, with $\partial\psi$ being maximal monotone locally by virtue of the convexity of ψ. The much deeper fact concerning equivalence[12] was established by a long and intricate argument in [25]. □

Other properties besides local subgradient monotonicity were likewise revealed in [25] to be closely related to variational convexity, for instance, having an open convex neighborhood $U \times V$ of $(\bar{x}, \bar{y})$ and $\varepsilon > 0$ such that, for the set $U_\varepsilon = \{\, x \in U \mid f(x) \leq f(\bar{x}) + \varepsilon \}$,

$$(x, y) \in [U_\varepsilon \times V] \cap \mathrm{gph}\,\partial\varphi \implies \varphi(x') \geq \varphi(x) + \langle y, x' - x \rangle \text{ for all } x' \in U. \tag{3.12}$$

Again, it's obvious this is implied by variational convexity from its definition. Similarly, variational strong convexity always brings with it the property that, for some $\sigma > 0$,

$$(x, y) \in [U_\varepsilon \times V] \cap \mathrm{gph}\,\partial\varphi \implies \varphi(x') \geq \varphi(x) + \langle y, x' - x \rangle + \frac{\sigma}{2}||x' - x||^2 \text{ for all } x' \in U. \tag{3.13}$$

Once more too, we have equivalences when H is finite-dimensional and $\bar{y} \in \widehat{\partial}\varphi(\bar{x})$.

Note that in a context of φ having a local minimum at $\bar{x}$, not just the condition $0 \in \partial\varphi(\bar{x})$ but even $0 \in \widehat{\partial}\varphi(\bar{x})$ is necessary. To that extent the need for $\bar{y}$ to be a regular subgradient in the equivalence parts of Theorem 7 isn't much of a handicap.

Anyway, our optimization topic is not minimizing φ over H but minimizing it over the subspace $S \subset H$. For that we have the following insight, which will tie into the setting of elicitation that revolves around $\varphi + \frac{e}{2}d_S^2$ and its subgradient mapping $\partial\varphi + eP^\perp$. Recall in this that $P^\perp$ vanishes on S, so having $\bar{y} \in \partial\varphi(\bar{x})$ is the same as having $\bar{y} \in [\partial\varphi + eP^\perp](\bar{x})$ when $\bar{x} \in S$.

Theorem 8: sufficiency in local minimization. *Consider lsc $\varphi : H \to (-\infty, \infty]$ and vectors $\bar{x} \in S$ and $\bar{y} \in S^\perp$ solving the linkage problem (1.1) for $\partial\varphi$ in its role as a first-order optimality condition for the minimization of φ over S. If*

$$\exists\, e \geq 0 \text{ such that } \varphi + \frac{e}{2}d_S^2 \text{ is variationally convex at } \bar{x} \text{ for } \bar{y}, \tag{3.14}$$

then φ has a local minimum over S at $\bar{x}$, not just a critical point. If the variational convexity is strong, then the local minimum is strong in the sense of the existence of $\sigma > 0$ such that $\varphi(x) \geq \varphi(\bar{x}) + \frac{\sigma}{2}||x - \bar{x}||^2$ for all $x \in S$ in some neighborhood of $\bar{x}$.

Proof. If $\varphi + \frac{e}{2}d_S^2$ is variationally convex at $\bar{x} \in S$ for its subgradient $\bar{y}$ at $\bar{x}$, then the function $\overline{\varphi}(x) = \varphi(x) + \frac{e}{2}d_S^2(x) - \langle \bar{y}, x - \bar{x} \rangle$ is variationally convex at $\bar{x}$ relative to its subgradient $0 \in \partial\overline{\varphi}(\bar{x})$. This implies, as we have seen, that $\overline{\varphi}$ has a local minimum at $\bar{x}$ over H. But $\overline{\varphi}(x) = \varphi(x)$ when $x \in S$ and $\bar{y} \in S^\perp$, so this yields a local minimum of φ at $\bar{x}$ relative to S. The assertion for variational strong convexity uses the quadratic bounds in (3.13) that are associated with that property. □

The results in Theorem 8 suggest a new approach to thinking about second-order optimality, one which isn't dependent on an exotic definition of second derivatives of lsc functions.

[12] It's an open question whether the restriction to finite-dimensional H and a regular subgradient $\bar{y}$ is truly essential.

Definition 3: variational second-order optimality conditions. *The property in (3.14) will be called the variational second-order sufficient condition for φ to have a local minimum at $\bar{x}$ relative to S when $\bar{y} \in \partial\varphi(\bar{x})$. The version of (3.14) with variational strong convexity will accordingly be called the variational strong second-order sufficient condition.*

To appreciate the really broad scope of this idea, let's look at it in the framework of the reformulated minimization problem in (1.6) with subgradients characterized in (1.7). In that notation, the first-order optimality condition is

$$(0, \bar{y}) \in \partial f(\bar{x}, 0) \tag{3.15}$$

with $\bar{y}$ being a Lagrange multiplier for the constraint $u = 0$; recall that u is interpreted as a perturbation. The function $\varphi_e = \varphi + \frac{e}{2} d_S^2$ translates then to

$$f_e(x, u) = f(x, u) + \frac{e}{2}||u||^2, \text{ likewise having } (0, \bar{y}) \in \partial f_e(\bar{x}, 0). \tag{3.16}$$

The corresponding statement of the variational second-order sufficient condition in Definiton 3 is

$$\exists\, \varepsilon \geq 0 \text{ such that } f_e \text{ is variationally convex at } (\bar{x}, 0) \text{ for } (0, \bar{y}). \tag{3.17}$$

Theorem 8 tells us that this guarantees having a local minimum in the very general problem of minimizing an expression $f(x, u)$ subject to $u = 0$.

Another telling example concerns the case where H is finite-dimensional and φ is *prox-regular* at $\bar{x}$ for $\bar{y}$, which is a very common property; see [29, Chapter 13F].[13] The variational *strong* second-order sufficient condition is equivalent then to the function $\varphi(x) + \frac{e}{2} d_S^2(x) - \langle \bar{y}, x \rangle$ having a *tilt-stable* local minimum at $\bar{x}$, as seen from the main characterization in [18].

Theorem 8 and Definition 3 bring us to the main conclusion we have been working toward about local elicitation in the optimization case.

Theorem 9: elicitability from second-order sufficiency. *Suppose that $\bar{x} \in S$ and $\bar{y} \in S^\perp$ with $\bar{y} \in \partial\varphi(\bar{x})$ in the optimization case of the linkage problem. If the variational second-order sufficient condition for a local minimum at $\bar{x}$ is satisfied, then monotonicity is elicitable at that level e in the sense of the mapping $T + eP^\perp$ being maximal monotone locally in the manner of T in (3.11). When H is finite-dimensional and $\bar{y} \in \widehat{\partial}\varphi(\bar{x})$, the converse holds as well.*

In the case of the variational strong second-order sufficient condition, the same relationships hold with the local elicitability of monotonicity being strong.

Proof. This is now immediate from the facts that have been laid out. □

Theorem 9 conveys the picture that the progressive decoupling algorithm is naturally suited to applications of the linkage problem (1.1) in the optimization case when one is not merely looking to solve a first-order optimality condition but hoping really to *converge to a local minimum.* It could equally well be translated to the optimization format in (1.6) which we reviewed in connection with second-order sufficiency just before its statement.

There is still one piece of unfinished business: establishing that the procedure in this setting can be executed with local minimization on the forefront as well.

It has already been noted at the end of Section 1 that the main step of the algorithm in (1.12) reduces in the optimization case to (1.17), where the first-order condition for a local minimum of a modified objective function φ^ν is to be solved. This is to be carried out in a local sense as specified in

[13] For instance, this holds when φ is the composition of some lsc convex function with a $\mathcal{C}^2$ mapping for which a mild constraint qualification holds at $\bar{x}$. Many expressions can be made to fit that.

Theorem 1 in terms of a neighborhood W of some solution pair $(\tilde{x}, \tilde{y})$ that is not necessarily the one to which the sequence will converge. With the information we have developed about how subgradients connect to local optimality, we can shed more light on this matter.

Theorem 10: implementation in minimizing mode. *Let $(\tilde{x}, \tilde{y})$ solve the linkage problem in the optimization case and be such that the variational second-order sufficient condition for a local minimum of φ at $\tilde{x}$ is satisfied with respect to $\tilde{y}$. Then if the iterations of the progressive decoupling algorithm with proximal parameter $r > 0$ are started from a pair (x^0, y^0) close enough to $(\tilde{x}, \tilde{y})$, there will be open convex neighborhoods U of $\tilde{x}$ and V of $\tilde{y}$ such that*

$$\textit{locally solving } \; 0 \in \partial\varphi^\nu(\widehat{x}^\nu) \;\; \textit{for} \;\; \varphi^\nu(x) = \varphi(x) - \langle y^\nu, x\rangle + \frac{r}{2}||x - x^\nu||^2 \tag{3.18}$$

can be impliemented by determining $\widehat{x}^\nu$ as the unique minimizer of φ^ν on U.

Proof. The claims would be valid beyond question if φ were convex, but here all we have from our hypothesis is the variational convexity of $\varphi_e = \varphi + \frac{e}{2} d_S^2$ at $\tilde{x}$ for some $e \geq 0$ with respect to $\tilde{y} \in \partial\varphi_e(\tilde{x})$. According to Theorem 9 that furnishes the maximal monotonicity with respect to some open convex neighborhood $U \times V$ of $(\tilde{x}, \tilde{y})$ and some $\varepsilon > 0$ of the mapping T defined by

$$\operatorname{gph} T = \{\, (x, y) \in \operatorname{gph} \partial\varphi \,|\, \varphi(x) < \varphi(\tilde{x}) + \varepsilon\}. \tag{3.19}$$

We know then from Theorem 1 that the iterations of the algorithm, when initiated in a small enough neighborhood W of $(\tilde{x}, \tilde{y})$ within $U \times V$ can be kept to that neighborhood, where they moreover will uniquely generate a sequence of pairs (x^ν, y^ν) converging to a some solution pair $(\bar{x}, \bar{y})$. Since

$$\frac{r}{2}||x - x^\nu||^2 = \frac{r}{2}||P(x) - x^\nu||^2 + \frac{r - e}{2}||P^\perp(x)||^2 + \frac{e}{2}||P^\perp(x)||^2$$

as long as $x^\nu \in S$, those iterations as portrayed in (3.18) can be viewed as locally solving

$$0 \in \partial\varphi_e^\nu(\widehat{x}^\nu) \;\; \text{for} \;\; \varphi_e^\nu(x) = \varphi_e(x) - \langle \bar{y}, x\rangle + \frac{r}{2}||P(x) - x^\nu||^2 + \frac{r - e}{2}||P^\perp(x)||^2. \tag{3.20}$$

From the variational convexity property of φ_e that we invoked to get the maximal monotonicity in $U \times V$ of T in the form of (3.19), we also have the existence of a convex lsc function $\psi \leq \varphi_e$ such that (3.6) holds for φ_e in place of φ. But this tells us that (3.20) is equivalent to locally solving

$$0 \in \partial\psi^\nu(\widehat{x}^\nu) \;\; \text{for} \;\; \psi^\nu(x) = \psi(x) - \langle \bar{y}, x\rangle + \frac{r}{2}||P(x) - x^\nu||^2 + \frac{r - e}{2}||P^\perp(x)||^2, \tag{3.21}$$

inasmuch as the relation in (3.6) holds then for ψ^ν and φ_e^ν in place of ψ and φ_e. Moreover the solution corresponds to the unique minimum of the strongly convex function ψ^ν, and also the unique minimum of ψ_e^ν, on U. □

It shouldn't be forgotten though that the local minimization in this implementation doesn't involve minimizing a function φ^ν that's necessarily convex, even locally. On the other hand, situations in the corollaries to Theorems 5 and 6 do provide local or even global convexity in the minimization.

Along the lines of proximal terms being employed in the minimization of nonconvex functions, it's interesting to note that an elicitation-like technique is present in a bundle method developed in [8], which includes rules for updating the proximal parameter.

Finally, it must be mentioned that our efforts here aren't the first to explore approaches to problem decomposition in nonconvex optimization which, in Lagrangian terms, resemble multiplier methods and rely on second-order optimality conditions. Those features can already be found in a 1979 article of Bertsekas [3] in the classical framework of smooth nonlinear programming. However, his method is very different and doesn't involve augmented Lagrangians.

4 Applications to problem decomposition and splitting

The special aim of the progressive decoupling algorithm is to determine a solution to the linkage problem (1.1) by solving a sequence of advantageously modified problems in which the linkages are relaxed by a dual representation which progressively captures them more accurately. The decoupling of linkages not only leads to simpler computations, but also opens the way to problem decomposition based on block separability.

Linkage problems with decomposable structure. *Suppose that* $H = H_1 \times \cdots \times H_q$ *for Hilbert spaces* H_j *and that*

$$T(x) = T(x_1, \ldots, x_q) = (T_1(x_1), \ldots, T_q(x_q)) \text{ for mappings } T_j : H_j \rightrightarrows H_j. \tag{4.1}$$

Problem (1.1) then seeks to

$$\text{find } (\bar{x}_1, \ldots, \bar{x}_q) \in S \text{ and } (\bar{y}_1, \ldots, \bar{y}_q) \in S^\perp \text{ such that } \bar{y}_j \in T_j(\bar{x}_j) \text{ for } j = 1, \ldots, q. \tag{4.2}$$

The challenge for "problem decomposition" is coming up with a way of solving (4.2) through subproblems in the separate indices j. The progressive decoupling algorithm can do that.

Progressive decoupling in decomposition mode. *With respect to an elicitation level* $e \geq 0$ *(global or local) and a proximal parameter* $r > e$*, proceed as follows. In iteration* ν*, having* $x^\nu = (x_1^\nu, \ldots, x_q^\nu) \in S$ *and* $y^\nu = (y_1^\nu, \ldots, y_q^\nu) \in S^\perp$*, determine* $\widehat{x}_j^\nu \in H_j$ *for* $j = 1, \ldots, q$*, by solving*

$$0 \in T_j^\nu(\widehat{x}_j^\nu), \quad \text{where } T_j^\nu(x_j) = T_j(x_j) - y_j^\nu + r[x_j - x_j^\nu]. \tag{4.3}$$

From $\widehat{x}^\nu = (\widehat{x}_1^\nu, \ldots, \widehat{x}_q^\nu)$ *then get* $x^{\nu+1} \in S$ *and* $y^{\nu+1} \in S^\perp$ *by*

$$x^{\nu+1} = P(\widehat{x}^\nu), \qquad y_j^{\nu+1} = y_j^\nu - (r-e)[\widehat{x}_j^\nu - x_j^{\nu+1}]. \tag{4.4}$$

This specialization of the decoupling procedure in Section 1 comes simply from the fact that the block-separable structure reduces the condition $0 \in T^\nu(\widehat{x}^\nu)$ in (1.12) to the separate conditions in (4.3) on the components of $\widehat{x}_j^\nu$. All the results about convergence and elicitation then apply. The conditions in (4.3) will only be solved locally, when the elicitation of monotonicity is only local.

In [24] we described such a method of problem decomposition like this but only for T globally monotone, which corresponds here to global elicitation of monotonicity at level $e = 0$. Furthermore, only the variational inequality case $T = F + N_C$ in (1.2) was considered in that work. Everything else is new here, and that holds also in the connections with optimization, which we explain next.

Optimization version of problem decomposition. The optimization case of the linkage problem with $T = \partial\varphi$ exhibits the decomposable structure in (4.1) when

$$\varphi(x) = \varphi(x_1, \ldots, x_q) = \varphi_1(x_1) + \cdots + \varphi_q(x_q) \text{ for lsc functions } \varphi_j : H_j \to (-\infty, \infty], \tag{4.5}$$

which leads to $T_j = \partial\varphi_j$. The step in (4.3) then takes the form of solving, for $j = 1, \ldots, q$,

$$\widehat{x}_j^\nu = [\text{local}] \operatorname*{argmin}_{x_j \in H_j} \varphi_j^\nu(x_j), \quad \text{where } \varphi_j^\nu(x_j) = \varphi_j(x_j) - \langle y_j^\nu, x_j \rangle + \frac{r}{2}||x_j - x_j^\nu||^2, \tag{4.6}$$

at least in various circumstances described in Section 3. Beyond actual convexity of the functions φ_j, included here as the case of elicitation at level $e = 0$, this holds for φ as in the corollaries to Theorems 5 and 6 and more generally for local elicitation as covered by Theorem 9.

Multistage stochastic programming offers an example of the separability in (4.5) in which each j marks a different "scenario." The modified problem for j in (4.6) is solved as if that scenario is the one that surely will be followed; it's a calculation with "hindsight." Progressive decoupling in this situation reduces to the progressive hedging in [28]. In the broader formulation of (4.3) as adapted to variational inequalities, it corresponds to the progressive hedging approach in [26] for solving stochastic variational inequalities. However, all that, until now, has been articulated only in the presence of global convexity or monotonicity. Developments coming in Section 5 which suggest Lagrangian formats for explicitly dealing with constraints, in particular, could further aid in this direction.

Another major illustration of decoupling comes from taking the spaces H_j in the block-separable structure to all be the same and making an "artificial" choice of the linkage space S.

Linkage problems with splitting structure. *Suppose that $H = H_0 \times \cdots \times H_0$ for q copies of a Hilbert space H_0 and that*

$$T(x) = T(x_1, \ldots, x_q) = (T_1(x_1), \ldots, T_q(x_q)) \;\text{ for mappings }\; T_j : H_0 \rightrightarrows H_0. \tag{4.7}$$

Choose the subspace S with orthogonal complement $S^\perp$ to be

$$\begin{aligned} S &= \{\, x = (x_1, \ldots, x_q) \,|\, \exists\, w \text{ with } x_1 = \cdots = x_q = w\}, \\ S^\perp &= \{\, y = (y_1, \ldots, y_q) \,|\, y_1 + \cdots + y_q = 0\}. \end{aligned} \tag{4.8}$$

Problem (4.1) then seeks in effect to

$$\text{find }\; \bar{w} \in H_0 \;\text{ such that }\; 0 \in T_1(\bar{w}) + \cdots + T_q(\bar{w}). \tag{4.9}$$

The decomposition iterations in (4.3)–(4.4) reduce then to the following procedure.

Progressive decoupling in splitting mode. *With respect to an elicitation level $e \geq 0$ (global or local) and a proximal parameter $r > e$, proceed as follows. In iteration ν, having $w^\nu \in H_0$ and $y_1^\nu, \ldots, y_q^\nu \in H_0$ with $y_1^\nu + \cdots + y_q^\nu = 0$, determine $\widehat{x}_j^\nu \in H_0$ for $j = 1, \ldots, q$, by solving*

$$0 \in T_j^\nu(\widehat{x}_j^\nu), \;\text{ where }\; T_j^\nu(x_j) = T_j(x_j) - y_j^\nu + r[x_j - w^\nu]. \tag{4.10}$$

Then update by

$$w^{\nu+1} = \frac{1}{q}\sum\nolimits_{j=1}^{q} \widehat{x}_j^\nu, \qquad y_j^{\nu+1} = y_j^\nu - (r - e)[\widehat{x}_j^\nu - w^{\nu+1}]. \tag{4.11}$$

Here w^ν gives the common value of the components of $x^\nu = (x_1^\nu, \ldots, x_q^\nu)$ in iteration ν. The projection of $\widehat{x}^\nu = (\widehat{x}_1^\nu, \ldots, \widehat{x}_q^\nu)$ to get $x^{\nu+1}$ in the subspace S in (4.8) corresponds then to the first formula in (4.11).

A splitting method like this was also considered by Spingarn [31] for his method of partial inverses, but only in the case of each T_j being maximal monotone (hence the same for T). Our splitting algorithm reduces in that setting to his with $e = 0$ when $r = 1$, but it differs in allowing, even in his circumstances, general $r > e \geq 0$. A similar method of splitting, likewise only under the assumption of maximal monotonicity but allowing general r, has been offered by Mahey et al. in [13]. It has a different update rule than (4.11) and only handles two components ($q = 2$), at least in its given formulation.

More about methology in this monotonicity framework and its relationships to the work of Spingarn is explained in the recent survey [11]. It's all closely related also to the Douglas-Rachford splitting method for monotone mappings as put forward originally by Eckstein and Bertsekas in [6]. Other contributions are to be found in [1, 5] and still more in related formats of convex optimization such as [7]. However, here we have the potential of splitting without insisting on convexity.

Optimization version of splitting. When $T_j = \partial\varphi_j$ the problem really being targeted for solution, in (4.9), is to

$$\text{find } \bar{w} \in H_0 \text{ such that } 0 \in \partial\varphi_1(\bar{w}) + \cdots + \partial\varphi_q(\bar{w}), \tag{4.12}$$

which is a surrogate for

$$\text{minimize } \varphi_1(w) + \cdots + \varphi_q(w) \text{ over } w \in H_0. \tag{4.13}$$

This is elaborated in a seemingly roundabout way by (4.7)–(4.8)–(4.9) into

$$\text{find } \bar{y}_j \in \partial\varphi_j(\bar{x}_j) \text{ such that } \bar{x}_1 = \cdots = \bar{x}_q \text{ and } \bar{y}_1 + \cdots + \bar{y}_q = 0, \tag{4.14}$$

which itself is a surrogate for

$$\text{minimize } \varphi_1(x_1) + \cdots + \varphi_q(x_q) \text{ subject to the linkage constraint } x_1 = \cdots = x_q. \tag{4.15}$$

The iterations of the algorithm in minimization mode, when supported by convexity or the conditions in the corollaries to Theorems 5 and 6 or in Theorem 9, have

$$\widehat{x}_j^\nu = [\text{local}]\,\text{argmin}_{x_j}\, \varphi_j^\nu(x_j), \text{ where } \varphi_j^\nu(x_j) = \varphi_j(x_j) - \langle y_j^\nu, x_j\rangle + \frac{r}{2}||x_j - w^\nu||^2, \tag{4.16}$$

in which w^ν is the current candidate for solving (4.12).

There are new aspects to this method even when the functions φ_j are convex, including just local convexity around a solution. But perhaps the most interesting question to answer is how *local elicitation of maximal monotonicity of* $T = \partial\varphi$, which has never before been contemplated in splitting, may truly be helpful.

It's crucial to appreciate that in posing the minimization in (4.13) in the seemingly redundant manner of (4.15) there is a *beneficial change in the associated optimality conditions.* The linkage constraint in (4.15) brings in Lagrange multipliers that didn't exist in (4.13). The second-order aspects of optimality change as well. In (4.13) they only concern the variable $w \in H_0$, but in (4.15) we can, for instance, invoke the variational second-order sufficient condition for a local minimum relative to the associated linkage subspace S in the product space H. That may elicit variational convexity or something stronger.

A good way to get insights into this is to put the problem in the perturbation framework of (1.6), which was taken up again ahead of Theorem 9 on the context of second-order sufficiency. The elements of our space H can uniquely be represented in the form

$$(w + u_1, \ldots, w + u_q) \text{ with perturbation vectors } u_j \text{ satisfying } u_1 + \cdots + u_q = 0. \tag{4.17}$$

The problem of interest is to

$$\text{minimize } \varphi_1(w + u_1) + \cdots + \varphi_q(w + u_q) \text{ subject to } (u_1, \ldots, u_q) = (0, \ldots, 0), \tag{4.18}$$

but other choices $(u_1, \ldots, u_q) \neq (0, \ldots, 0)$ offer perturbed versions of that problem in which the functions φ_j undergo shifts in one direction or another. The vector $(\bar{y}_1, \ldots, \bar{y}_q)$ paired with $\bar{w}$ assesses the effect of such shifts on the minimum value. It's in that context that the variational second-order sufficient condition for optimality enters and enriches the scene.

Example: splitting in smooth optimization. Suppose the functions φ_j are of class $\mathcal{C}^2$, so that the first-order optimality condition for minimizing $\varphi = \varphi_1 + \cdots + \varphi_q$ over H_0 is

$$0 = \nabla\varphi(\bar{w}) = \nabla\varphi_1(\bar{w}) + \cdots + \nabla\varphi_q(\bar{w}). \tag{4.19}$$

Solving that is the task in (4.12) which is re-expressed in (4.15) in terms of vectors $\bar{y}_j = \nabla\varphi_j(\bar{x}_j)$ with $\bar{x}_1 = \cdots = \bar{x}_q$. To apply splitting methodology to (4.19) in its existing forms, which require monotonicity, the functions φ_j would need to be convex, thereby entailing for Hessian matrices that

$$\nabla^2\varphi_j(w) \text{ for } j = 1,\dots,q \text{ would have to be positive semidefinite for all } w.$$

In stark contrast, our new approach is able to operate in this situation simply on the basis of the classical second-order sufficient condition for a local minimum at $\bar{w}$, namely the

$$\text{positive definiteness of } \nabla^2\varphi(\bar{w}) = \nabla^2\varphi_1(\bar{w}) + \cdots + \nabla^2\varphi_q(\bar{w}),$$

because that locally elicits strong monotonicity.

This echoes the example of local elicitation from smoothness that was offered in Section 1 and validated with various extensions in Section 3.

Linkage problems from simple linear coupling. *Let H and T have the decomposable structure considered earlier in (4.1) but just with $q = 2$. For corresponding problem in (4.2), choose the complementary subspaces as follows with respect to a (continuous) linear mapping $A : H_1 \to H_2$ and its adjoint $A^* : H_2 \to H_1$:*

$$S = \{\, x = (x_1, x_2) \,|\, x_2 = Ax_1\}, \qquad S^\perp = \{\, y = (y_1, y_2) \,|\, y_1 = -A^* y_2\}. \tag{4.20}$$

The linkage problem then takes the form

$$\textit{find } x_1 \in H_1 \textit{ and } y_2 \in H_2 \textit{ such that } y_2 \in T_2(Ax_1) \textit{ and } -A^* y_2 \in T_1(x_1), \tag{4.21}$$

which is a roundabout way of saying

$$\textit{find } x_1 \in H_1 \textit{ such that } 0 \in T_1(x_1) + A^* T_2(Ax_1). \tag{4.22}$$

The corresponding specialization of the progressive decoupling algorithm executes (4.3) for $j = 1, 2$, and then requires projection on S in (4.4). To what extent will this projection step be numerically accessible? The answer to that depends on the mapping A. Projecting on S in (4.4) means determining the point $x^{\nu+1} = (x_1^{\nu+1}, Ax_1^\nu)$ of S that's nearest to $\widehat{x}^\nu = (\widehat{x}_1^\nu, \widehat{x}_2^\nu)$, or in other words, calculating $x_1^{\nu+1} = \operatorname{argmin}_{x_1 \in H_1} \{\, ||x_1 - \widehat{x}_1^\nu||^2 + ||Ax_1 - \widehat{x}_2^\nu||^2\}$. That has a closed-form answer:

$$x_1^{\nu+1} = (I + A^*A)^{-1}[\widehat{x}_1^\nu + A^*\widehat{x}_2^\nu]. \tag{4.23}$$

Practicability thus comes down to the ease of working with $(I + A^*A)^{-1}$. An elementary example is the case of splitting with $q = 2$, which corresponds to $Ax_1 = x_1$, $A^*y_2 = -y_2$, hence $A^*A = I$.

Optimization version of simple linear coupling. When $T_j = \varphi_j$ we are looking to

$$\text{find } \bar{x}_1 \text{ such that } 0 \in \partial\varphi_1(\bar{x}_1) + A^*\partial\varphi_2(A\bar{x}_1), \tag{4.24}$$

but the vectors $y_2 \in A^*\partial\varphi_2(Ax_1)$ are the subgradients of the function $x_1 \to \varphi_2(Ax_1)$ under typical constraint qualifications, so (4.24) is a natural surrogate for the problem:

$$\text{minimize } \varphi_1(x_1) + \varphi_2(Ax_1) \text{ with respect to } x_1 \in H_1. \tag{4.25}$$

This kind of optimization problem has a long history in convex analysis as the vehicle for Fenchel duality when φ_1 and φ_2 are convex functions. Algorithms for solving it through separate minimization operations on modifications of φ_1 and φ_2 have been developed in various quarters and extended to the monotone case of the problem in (4.22). The 2017 survey of Lenoir and Mahey [11] provides a valuable guide to that literature.

The procedure obtained here differs in its dependence on the inverse in (4.23), which may interfere with its viability relative to that existing methodology. But the special feature is that we are *not* requiring φ_1 and φ_2 to be convex. That's made possible through our scheme of eliciting convexity and its connection in Section 3 with second-order sufficient conditions for a local minimum.

An example in smooth optimization can once more bring this out. Suppose that φ_1 and φ_2 are $\mathcal{C}^2$. Then (4.24) reduces to finding $\bar{x}_1$ such that $\nabla\varphi(\bar{x}_1) = 0$ for $\varphi(x_1) = \varphi_1(x_1) + \varphi_2(Ax_1)$. For the problem to fit with the technology of the convex case, all the Hessian matrices $\nabla^2\varphi_1(x_1)$ and $\nabla^2\varphi_2(x_2)$ would have to be positive semidefinite. But to justify local convergence to $\bar{x}_1$ for algorithm developed here, all that is needed is the positive definiteness of the Hessian $\nabla^2\varphi(\bar{x}_1)$.

From the perturbation standpoint in understanding this model and its version of local optimality, an element of $(x_1, x_2) \in S$, having $x_2 = Ax_1$, is shifted to (x_1+u_1, x_2+u_2) by an element $(u_1, u_2) \in S^\perp$, having $u_1 = -A^*u_2$. This seems a bit artificial and perhaps indicates a shortcoming in such a treatment of "linear coupling." In the next section we'll employ a different strategy which succeeds better and even can handle "nonlinear coupling."

5 Extended decomposition with augmented Lagrangians

From now on we keep to the optimization case of the linkage problem with block-separability but furnish the objective with more structure. The problem, in its starting formulation, is to

$$\begin{array}{ll} \text{minimize} & \varphi(x_1, \ldots, x_q) \text{ on a subspace } S \subset H_1 \times \cdots \times H_q \\ \text{where} & \varphi(x_1, \ldots, x_q) = \sum_{j=1}^q f_j(x_j) + g(\sum_{j=1}^q F_j(x_j)), \end{array} \tag{5.1}$$

with F_j being a continuously differentiable mapping from H_j to another Hilbert space H', and the functions $f_j : H_j \to (-\infty, \infty]$ and $g : H' \to (-\infty, \infty]$ being lsc with g also *convex*.

For example, g could be the indicator δ_K of some closed convex set in H', and then the g term in the minimization would enforce the constraint $\sum_{j=1}^q F_j(x_j) \in K$. For $K = \{0\}$, this would be the equation $\sum_{j=1}^q F_j(x_j) = 0$. However, g could instead be some norm that provides "regularization."

More generally, H' could itself be a product $H'_1 \times \cdots \times H'_p$ with $g(z_1, \ldots, z_p)$ being a sum $\sum_{k=1}^p g_i(z_i)$ that imposes different conditions on the different components of $F_j(x_j) = (F_{j1}(x_j), \ldots, F_{jp}(x_j))$. On the other hand, the functions f_j could have the form $f_j(x_j) = f_{0j}(x_j) + \delta_{X_j}(x_j)$ with f_{0j} continuously differentiable and X_j closed. That would enforce the constraint $x_j \in X_j$.

The g term in (5.1) introduces *an additional way of coupling* the variables x_j beyond the constraint $(x_1, \ldots, x_q) \in S$. In fact this might be the only source of coupling because we are now allowing that S might be all of $H_1 \times \cdots \times H_q$, with $S^\perp$ reducing then to the origin of $H_1 \times \cdots \times H_q$. An alternative is a *splitting* version in which S has the form in (4.8). This would correspond in problem (5.1) to having the $H_j = H_0$ and $x_j = w \in H_0$ for all j. For applications to stochastic programming, on the other hand, S could be the nonanticipativity subspace that's critical in that subject.

When the mappings F_j are linear (or affine), we could deal with the extended coupling in (5.1) by an enlargement of S that mimics the graphical pattern in (4.20). But here we allow *nonlinearity* of F_j, and therefore a different approach is needed.

The linkage problem (1.1) for the minimization problem (5.1) centers on the corresponding mapping $\partial\varphi$ and requires knowing when $\bar{v} \in \partial\varphi(\bar{x})$. Here the earlier $\bar{y}$ has been switched to $\bar{v}$ in order to reserve $\bar{y}$ for use as a "Lagrange multiplier vector" for the coupling associated with the g term. Rules of subgradient calculus in [29, Chapter 10B], when applied to the formula for φ in (5.1), say that having $\bar{v} \in \partial\varphi(\bar{x})$ implies, under a constraint qualification, that

$$\exists\, \bar{y} \in \partial g\Big(\sum_{j=1}^{q} F_j(\bar{x}_j)\Big) \text{ such that } \bar{v}_j - \nabla F_j(\bar{x}_j)^*\bar{y} \in \partial f_j(\bar{x}_j) \text{ for } j = 1, \ldots, q, \tag{5.2}$$

and in some situations there's an equivalence. However, it will be better to work explicitly instead of implicitly with such $\bar{y}$ and avoid the issue of constraint qualifications. This can be accomplished by taking a perturbational approach to the problem in (5.1), reformulating it as:

$$\begin{array}{ll} \text{minimize} & \varphi_0(x_1, \ldots, x_q, u) \text{ on the subspace } S_0 \subset H_1 \times \cdots \times H_q \times H' \\ \text{where} & \varphi_0(x_1, \ldots, x_q, u) = \sum_{j=1}^{q} f_j(x_j) + g(\sum_{j=1}^{q} F_j(x_j) + u), \\ \text{and} & S_0 = \{ (x_1, \ldots, x_q, u) \,|\, (x_1, \ldots, x_q) \in S,\ u = 0 \}, \\ & S_0^{\perp} = \{ (v_1, \ldots, v_q, y) \,|\, (v_1, \ldots, v_q) \in S^{\perp},\ y \in H' \}. \end{array} \tag{5.3}$$

The linkage problem in subgradient form is then to

$$\text{find } (\bar{x}_1, \ldots, \bar{x}_q) \in S,\ (\bar{v}_1, \ldots, \bar{v}_q) \in S^{\perp},\ \bar{y} \in H', \text{ with } (\bar{v}_1, \ldots, \bar{v}_q, \bar{y}) \in \partial\varphi_0(\bar{x}_1, \ldots, \bar{x}_q, 0), \tag{5.4}$$

which is advantageous because we have now from subgradient calculus [29, Chapter 10B] that

$$(v_1, \ldots, v_q, y) \in \partial\varphi_0(x_1, \ldots, x_q, u) \iff v_j - \nabla F_j(x_j)^* y \in \partial f_j(x_j),\ \ y \in \partial g\Big(\sum_{j=1}^{q} F_j(x_j)\Big). \tag{5.5}$$

Moreover the subgradient conditions on the right can be expressed by

$$(v_1, \ldots, v_q) \in \partial_x L(x_1, \ldots, x_q, y), \qquad u \in \partial_y[-L](x_1, \ldots, x_1, y), \tag{5.6}$$

with respect to the associated Lagrangian function

$$L(x_1, \ldots, x_q, y) = \inf_{u \in H'} \{ \varphi_0(x_1, \ldots, x_q, u) - \langle y, u \rangle \} = \sum_{j=1}^{q} [f_j(x_j) + \langle y, F_j(x_j) \rangle] - g^*(y), \tag{5.7}$$

where[14] g^* is the convex function conjugate to g, having $g^*(y) = \sup_u \{ \langle y, u \rangle - g(u) \}$.

This way of looking at the problem in (5.1) is helpful in clarifying subgradients and placing them in a Lagrangian format, but it doesn't go far enough. If we apply the progressive decoupling algorithm to (5.3) in minimization mode, we find ourselves trying to work with augmented Lagrangian functions

$$L_r(x_1, \ldots, x_q, y) = \inf_{u \in H'} \{ \varphi_0(x_1, \ldots, x_q, u) - \langle y, u \rangle + \tfrac{r}{2}||u||^2 \} \text{ for } r > 0 \tag{5.8}$$

which lack the block separability with respect to $x_1, \ldots, x_q$, that's enjoyed in (5.7) by L. This obliges us to take a different approach to perturbations in which we rely on the following fact.

[14]In (5.7), the value of the L is ∞ if it should arise that $\sum_{j=1}^{q} f_j(x_j) = \infty$ while also $g^*(y_j) = \infty$.

Lemma 1: expansion. *The inequality* $g(\sum_{j=1}^q F_j(x_j)) \leq \alpha$ *holds for* $\alpha \in I\!R$ *if and only if*

$$\exists\, u_j \;\text{ with }\; \sum_{j=1}^q u_j = 0 \;\text{ such that }\; q^{-1}\sum_{j=1}^q g(q[F_j(x_j)+u_j]) \leq \alpha. \tag{5.9}$$

Proof. Because g is a convex function, it has the property that $g(w) \leq q^{-1}\sum_{j=1}^q g(w_j)$ when $q^{-1}\sum_{j=1}^q w_j = w$, moreover with the inequality holding as an equation when $w_j = w$ for all j. In particular then, $g\big(\sum_{j=1}^q F_j(x_j)\big) = \min\{\, q^{-1}\sum_{j=1}^q g(w_j) \,|\, q^{-1}\sum_{j=1}^q w_j = \sum_{j=1}^q F_j(x_j)\}$. In changing variables from w_j to $u_j = q^{-1}w_j - F_j(x_j)$, this turns into

$$g\Big(\sum_{j=1}^q F_j(x_j)\Big) = \min\Big\{\, q^{-1}\sum_{j=1}^q g(q[F_j(x_j)+u_j]\,)\,\Big|\, \sum_{j=1}^q u_j = 0\Big\},$$

which confirms the claim. □

Note in (5.9) that $q^{-1}\sum_{j=1}^q g(q[F_j(x_j)+u_j]) = \sum_{j=1}^q g(F_j(x_j)+u_j)$ when g is positively homogeneous, as when $g = \delta_K$ for a cone K, or when g is a norm.

The lemma allows us to reformulate (5.1) as a problem in auxiliary variables u_j alongside of the variables x_j:

$$\begin{array}{l}\text{minimize }\; \overline{\varphi}(x_1,\ldots,x_q,u_1,\ldots,u_q) = \overline{\varphi}_1(x_1,u_1)+\cdots+\overline{\varphi}_q(x_q,u_q) \;\text{ on a subspace } \overline{S},\\ \text{where }\; \overline{\varphi}_j(x_j,u_j) = f_j(x_j) + q^{-1}g(q[F_j(x_j)+u_j]\,), \quad \overline{S}\subset\overline{H} = H_1\times\cdots\times H_q\times[H']^q,\end{array} \tag{5.10}$$

$$\begin{array}{l}\overline{S} = \{\,(x_1,\ldots,x_q,u_1,\ldots,u_q)\,|\,(x_1,\ldots,x_q)\in S,\; u_1+\cdots+u_q = 0, \;\text{ for }\; u_j\in H'\},\\ \overline{S}^\perp = \{\,(v_1,\ldots,v_q,y_1,\ldots,y_q)\,|\,(v_1,\ldots,v_q)\in S^\perp,\; y_1=\cdots=y_q = \text{some } y\in H'\}.\end{array} \tag{5.11}$$

This puts us back in the picture of decomposability at the start of Section 4, but with more variables. The first-order optimality condition in the subgradient form of the linkage problem now concerns

$$(\bar{v}_1,\ldots,\bar{v}_q,\bar{y}_1,\ldots,\bar{y}_q)\in\partial\overline{\varphi}(\bar{x}_1,\ldots,\bar{x}_q,\bar{u}_1,\ldots,\bar{u}_q), \;\text{ or }\; (\bar{v}_j,\bar{y}_j)\in\partial\overline{\varphi}_j(\bar{x}_j,\bar{u}_j) \;\text{ for each } j, \tag{5.12}$$

where[15]

$$(\bar{v}_j,\bar{y}_j)\in\partial\overline{\varphi}_j(\bar{x}_j,\bar{u}_j) \iff \bar{v}_j - \nabla F_j(\bar{x}_j)^*\bar{y}_j\in\partial f_j(\bar{x}_j), \quad \bar{y}_j\in\partial g(q[F_j(\bar{x}_j)+\bar{u}_j]). \tag{5.13}$$

It's worth noting from this calculation, in connection with the second-order variational sufficient condition and its role in elicitability in Theorem 9, that the subgradient of $\overline{\varphi}$ in (5.5) will be a regular subgradient as long as the subgradients of the functions f_j in (5.6) are regular subgradients.[16] [17]

The corresponding version of the algorithm in optimization mode for this problem has steps (4.6) followed by updates (4.3), but with added variables that require elaboration of details. The projections $\overline{P}$ and $\overline{P}^\perp = I - \overline{P}$ from $\overline{H}$ onto the subspaces $\overline{S}$ and $\overline{S}^\perp$ in (5.11) are now needed, but clearly

$$\overline{P}(x_1,\ldots,x_q,u_1,\ldots,u_q) = \Big(P(x_1,\ldots,x_q), u_1 - \tfrac{1}{q}\textstyle\sum_{j=1}^q u_j,\ldots,u_q - \tfrac{1}{q}\sum_{j=1}^q u_j\Big). \tag{5.14}$$

Another issue is the extended form of the minimization step, which in place of (4.6) now has

$$\begin{array}{l}(\widehat{x}_j^\nu,\widehat{u}_j^\nu) = [\text{local}]\,\text{argmin}\,\{\,\overline{\varphi}_j^\nu(x_j,u_j)\,|\,x_j\in H_j,\; u_j\in H'\}, \;\text{ where}\\ \overline{\varphi}_j^\nu(x_j,u_j) = \overline{\varphi}_j(x_j,u_j) - \langle v_j^\nu,x_j\rangle - \langle y^\nu,u_j\rangle + \frac{r}{2}||x_j - x_j^\nu||^2 + \frac{r}{2}||u_j - u_j^\nu||^2,\end{array} \tag{5.15}$$

with respect to current elements

$$(x_1^\nu,\ldots,x_q^\nu,u_1^\nu,\ldots,u_q^\nu)\in\overline{S}, \qquad (v_1^\nu,\ldots,v_q^\nu,y^\nu,\ldots,y^\nu)\in\overline{S}^\perp. \tag{5.16}$$

Altogether, then, we get the following algorithm.

[15] This applies the chain rule in [29, 10.6].

[16] Again the chain rule in [29, 10.6] provides the support.

[17] Although we're concentrating now on optimization, it's evident that a broader kind of linkage problem with decomposable structure could be obtained here by replacing the subgradient mappings ∂f_i and ∂g by other mappings.

Progressing decoupling in extended decomposition. *For an elicitation level $e \geq 0$ (global or local) and a proximal parameter value $r > e$, proceed in iteration ν from having current elements as in (5.16) to calculating the pairs in (5.15) for $j = 1, \ldots, q$ with respect to*

$$\overline{\varphi}_j(x_j, u_j) = f_j(x_j) + g_q(F_j(x_j) + u_j), \quad \text{where } g_q(u) = q^{-1}g(qu). \tag{5.17}$$

Then, letting $\widehat{u}^\nu = \frac{1}{q}\sum_{j=1}^q \widehat{u}_j^\nu$, update by

$$\begin{array}{ll} (x_1^{\nu+1}, \ldots, x_q^{\nu+1}) = P(\widehat{x}_1^\nu, \ldots, \widehat{x}_q^\nu), & u_j^{\nu+1} = \widehat{u}_j^\nu - \widehat{u}^\nu, \\ v_j^{\nu+1} = v_j^\nu - (r-e)[\widehat{x}_j^\nu - x_j^{\nu+1}], & y^{\nu+1} = y^\nu - (r-e)\widehat{u}^\nu. \end{array} \tag{5.18}$$

What can be expected of this procedure? It will have the convergence properties in Theorem 1 for producing a solution to the problem in form (5.3)–(5.4), which will then yield a solution to (5.1), but those properties depend on the global or local level of elicitation, so that's the key issue.

Global elicitation at level $e = 0$ refers to the *fully convex* case of the problem as stated in (5.10), namely $\overline{\varphi}$ being a convex function, which comes down to the expressions $\overline{\varphi}_j(x_j, u_j)$ in (5.17) being convex in (x_j, u_j). Since g_j is convex, that obviously holds when f_j is convex and F_j is affine, but it can also hold in other situations, for example when $H' = I\!R^n$, the components of F_j are convex, and g is nondecreasing. In the fully convex case, the algorithm in optimization mode is sure to converge globally from any starting elements (if a solution to the problem exists at all).

For local elicitation at a level $e > 0$ it's important to have in mind the connection with local optimality that was developed in Section 3, culminating in Theorems 9 and 10. The upshot is that *if the algorithm is initiated close enough to elements satisfying, together with the first-order condition on subgradients, the variational second-order sufficient condition*, and e is high enough (as dictated by that condition), then it will converge to a local minimum and can be implemented with local minimization generating unique pairs $(\widehat{x}_j^\nu, \widehat{u}_j^\nu)$ in (5.15). Of course, it's hard to know when those conditions are fulfilled, but this can be compared with the situation faced by other algorithms, such as Newton's method or procedures in nonlinear programming that call for various second-order optimality properties to be present.

There's more to learn in this direction through further analysis that brings augmented Lagrangians into the picture. Problems of the form

$$\text{minimize } f_j(x_j) + g_q(F_j^\nu(x_j)) \text{ in } x_j \in H_j, \text{ with } F_j^\nu = F_j(x_j) + u_j^\nu, \;\; g_q(u) = q^{-1}g(qu), \tag{5.19}$$

will have an interesting role. The Lagrangian function associated with such a problem, in the theory explained in [29, Chapters 11I, 11K], is

$$L_j^\nu(x_j, y_j) = \inf_{u_j \in H'} \Big\{ f_j(x_j) + g_q(F_j^\nu(x_j) + u_j) - \langle y_j, u_j\rangle \Big\} = f_j(x_j) + \langle y_j, F_j^\nu(x_j)\rangle - q^{-1}g^*(y_j), \tag{5.20}$$

where we have used the fact that the convex function conjugate to g_q is $q^{-1}g^*$. The *augmented Lagrangian* with parameter $r > 0$ is

$$L_{j,r}^\nu(x_j, y_j) = \min_{u_j \in H'} \Big\{ f_j(x_j) + g_q(F^\nu(x_j) + u_j) - \langle y_j, u_j\rangle + \frac{r}{2}||u_j||^2 \Big\}, \tag{5.21}$$

which is also known to be obtainable through duality by

$$L_{j,r}^\nu(x_j, y_j) = \max_{z_j \in H'} \Big\{ L_j^\nu(x_j, z_j) - \frac{1}{2r}||z_j - y_j||^2 \Big\}. \tag{5.22}$$

Although bringing these expressions down to something simpler needn't be our concern here, it's possible in many situations. Especially to note as an example is the case when g is just the indicator of the origin, so that $g^* \equiv 0$. Then (5.10) minimizes $f_j(x_j)$ subject to $F_j^\nu(x_j) = 0$, and the augmented Lagrangian is $L_{j,r}^\nu(x_j, y_j) = L_j^\nu(x_j, y_j) + \frac{r}{2}||F_j^\nu(x_j)||^2$, as would be anticipated.

The Lagrangian $L_j^\nu(x_j, y_j)$ is always concave in y_j, and the same is true also of the augmented Lagrangian $L_{j,r}^\nu(x_j, y_j)$. But the augmented Lagrangian has a further property of significance: it is continuously differentiable in y_j as long as $x_j \in \text{dom}\, f_j$, with the formula

$$\nabla_{y_j} L_{j,r}^\nu(x_j.y_j) = \underset{u_j \in H'}{\operatorname{argmin}} \Big\{ f_j(x_j) + g_q(F^\nu(x_j) + u_j) - \langle y_j, u_j \rangle + \frac{r}{2}||u_j||^2 \Big\}. \tag{5.23}$$

The way forward now, with the Lagrangians at our disposal, is to look at what happens in (5.15) when we carry out the minimization in u_j and afterward look at the residual minimization in x_j. The minimization with respect to u_j, in making use of (5.17), benefits from the following relationship.

Lemma 2: augmentation.

$$\min_{u_j \in H'} \Big\{ f_j(x_j) + g_q(F_j(x_j) + u_j) - \langle y^\nu, u_j \rangle + \frac{r}{2}||u_j - u_j^\nu||^2 \Big\} = L_{j,r}^\nu(x_j, y^\nu) - \langle y^\nu, u_j^\nu \rangle. \tag{5.24}$$

Proof. In the change of variables $u_j' = u_j - u_j^\nu$, the expression being minimized turns instead into $f_j(x_j) + g_q(F_j(x_j) + u_j^\nu + u_j') - \langle y^\nu, u_j' + u_j^\nu \rangle + \frac{r}{2}||u_j'||^2$, where $F_j(x_j) + u_j^\nu = F_j^\nu(x_j)$. With the removal of $\langle y^\nu, u_j^\nu \rangle$, the minimization is the same as that on the right side of (5.21). □

This enables us to restate the decomposition algorithm with steps (5.15), (5.17) and (5.18) in a different manner.

Progressive decoupling with augmented Lagrangians. *With respect to an elicitation level $e \geq 0$ (global or local) and a proximal paramter $r > e$, proceed in iteration ν from having current elements as in (5.16) to calculating, in terms of $F_j^\nu(x_j) = F_j(x_j) + u_j^\nu$ for $j = 1, \ldots, q$,*

$$\widehat{x}_j^\nu = [\text{local}]\ \underset{x_j \in H_j}{\operatorname{argmin}} \Big\{ L_{j,r}^\nu(x_j, y^\nu) - \langle v_j^\nu, x_j \rangle + \frac{r}{2}||x_j - x_j^\nu||^2 \Big\}. \tag{5.25}$$

Then, letting $\widehat{u}^\nu = \frac{1}{q}\sum_{j=1}^q \nabla_{y_j} L_{j,r}^\nu(x_j^{\nu+1}, y^\nu)$, update by

$$\begin{array}{ll} (x_1^{\nu+1}, \ldots, x_q^{\nu+1}) = P(\widehat{x}_1^\nu, \ldots, \widehat{x}_q^\nu), & u_j^{\nu+1} = u_j^\nu - \widehat{u}^\nu, \\ v_j^{\nu+1} = v_j^\nu - (r-e)[\widehat{x}_j^\nu - x_j^{\nu+1}], & y^{\nu+1} = y^\nu - (r-e)\widehat{u}^\nu. \end{array} \tag{5.26}$$

Spingarn in [32] had something similar for traditional convex programming. We extended that in [24] to convex programming with general constraint cones — no elicitation. Schemes that attach proximal terms to the primal variables in an augmented Lagrangian, as in (5.25), are known as proximal methods of multipliers and go back to [23].

The extended decomposition procedure featured here will identify a local minimum of the problem in (5.10)–(5.11) according to the success obtainable in elicitation. In the globally convex case, where the functions $\overline{\varphi}_j$ in (5.10) are convex, which corresponds to global elicitation of convexity at level $e = 0$, there will be global convergence from any starting elements. Local elicitation will work for e high enough under the associated version of the second-order variational sufficient condition as indicated in Theorem 9 (with local minimization implemented as explained in Theorem 10). More effort will be needed to fully understand the details of this in Lagrangian terms. Other recent advances on dealing with nonconvexity in decomposition are in the paper of Hong et al. [9], but nonconvex constraints aren't admitted there.

Many other problem formats besides (5.1) can be elaborated with perturbations into a format like (5.10)–(5.11) to which the progressive decoupling algorithm can applied and augmented Lagrangians will assist in its implementation. Here is another illustration. The problem

$$\text{minimize } \sum\nolimits_{j=1}^{q}\Big[f_j(x_j)+g_j(F_j(x_j))\Big] \text{ subject to } (x_1,\ldots,x_q)\in S, \tag{5.27}$$

having structure as above except that now there are Hilbert spaces H_j' with $\mathcal{C}^1$ mappings $F_j : H_j \to H_j'$ and convex functions $g_j : H_j \to (-\infty,\infty]$, can be elaborated to

$$\begin{array}{l}\text{minimize } \widetilde{\varphi}(x_1,\ldots,x_q,u_1,\ldots,u_q)=\sum_{j=1}^{q}\widetilde{\varphi}_j(x_j,u_j) \text{ on}\\ \widetilde{S}=\{\,(x_1,\ldots,x_q,u_1,\ldots,u_q)\,|\,(x_1,\ldots,x_q)\in S,\ u_j=0,\forall j\}\\ \text{where } \overline{\varphi}_j(x_j,u_j)=f_j(x_j)+g_j(F_j(x_j)+u_j).\end{array} \tag{5.28}$$

The variables u_j in this case give perturbations which lead to multipliers y_j and the complement

$$\widetilde{S}^{\perp}=\{\,(v_1,\ldots,v_q,y_1,\ldots,y_q)\,|\,(v_1,\ldots,v_q)\in S^{\perp},\ \text{ free } y_j\in I\!R^{m_j},\forall j\ \}. \tag{5.29}$$

This time the first-order condition, in terms of elements of $\widetilde{S}$ and $\widetilde{S}^{\perp}$, is

$$(\bar{v}_1,\ldots,\bar{v}_q,\bar{y}_1,\ldots,\bar{y}_q)\in\partial\widetilde{\varphi}(\bar{x}_1,\ldots,\bar{x}_q,\bar{u}_1,\ldots,\bar{u}_q), \text{ or } (\bar{v}_j,\bar{y}_j)\in\widetilde{\partial}\varphi_j(\bar{x}_j,\bar{u}_j) \text{ for each } j, \tag{5.30}$$

which calculus rules can clarify by

$$(\bar{v}_j,\bar{y}_j)\in\widetilde{\partial}\varphi_j(\bar{x}_j,\bar{u}_j) \iff \bar{v}_j-\nabla F_j(\bar{x}_j)^*\bar{y}_j\in\partial f_j(\bar{x}_j),\quad \bar{y}_j\in\partial g_j(F_j(\bar{x}_j)+\bar{u}_j) \tag{5.31}$$

and confirm that the subgradient of $\widetilde{\varphi}$ in (5.30) is regular if the subgradients of f_j in (5.31) are regular. On applying the algorithm with steps (4.6) and updates (4.3) in this setting, a decomposition procedure emerges in a manner very similar to the one above, including a consolidation into augmented Lagrangian steps.

References

[1] Attouch, H., Briceño-Arias, L. M., and Combettes, P. L., "A parallel splitting method for coupled monotone inclusions," *SIAM J. on Control and Optimization* **48** (2010), 3246–3270.

[2] Aussel, D., Corvellec, J.-N., and Lassonde, M., "Mean-value property and subdifferential criteria for lower semicontinuous functions," *Transactions of the American Mathematical Society* **347** (1995), 4147–4161.

[3] Bertsekas, D. P., "Convexification procedures and decomposition methods for nonconvex optimization problems," *J. Optimization Theory and Applications* **29** (1979), 169–197.

[4] Daniilidis, A., and Lemaréchal, C., "On a primal-proximal heuristic in discrete optimization," *Mathematical Programming A* **104** (2005), 105–128.

[5] Eckstein, J., "A simplified form of block-iterative operator splitting and an asynchronous algorithm resembling the multiblock ADMM," *J. Optimization Theory and Applications* **173** (2017), 155–182.

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

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

[8] Hare, W., and Sagastizabel, C., "A redistributed proximal bundle method for nonconvex optimization," *SIAM J. Optimization* **20**, 2442–2473.

[9] Hong, M.G., Luo, Z.Q., and Razaviyayn, M., "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems." *SIAM J. Optimization* **26** (2016), 33–364.

[10] Iusem, A. N., Pennanen, T. and Svaiter, B. F, "Inexact variants of the proximal point algorithm without monotonicity." *SIAM J. Optimization* **13** (2003), 1080–1097.

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

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

[13] 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.

[14] Minty, G. J., "Monotone (nonlinear) operators in Hilbert space." *Duke Mathematical J.* **29** (1962), 341–346.

[15] Mordukhovich, B. S., *Variational Analysis and Geneeralized Differentiation I: Basic Theory*, No. 330 in the series *Grundlehren der Mathematischen Wissenschaften*, Springer-Verlag, 2006.

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

[17] Poliquin, R. A., "Subgradient monotonicity and convex functions," *Nonlinear Analysis: Theory, Methods and Applications* **14** (1990), 385–398

[18] Poliquin, R. A., and Rockafellar, R. T., "Tilt stability of a local minimum," *SIAM J. Optimization* **8** (1998), 287–289.

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

[20] Rockafellar, R. T., "On the maximal monotonicity of subdifferential mappings." *Pacific J. Math.* **33** (1970), 209–216.

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

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

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

[24] Rockafellar, R. T., "Progressive decoupling of linkages in monotone variational inequalities and convex optimization," *Proceedings of the Conference on Nonlinear Analysis and Convex Analysis, Chitose, Japan, 2017*, submitted.

[25] Rockafellar, R. T., "Variational convexity and local monotonicity of subgradient mappings," *Vietnam Journal of Mathematics*, submitted.

[26] Rockafellar, R. T., and Sun, J. "Solving monotone stochastic variational inequalities and complementarity problems by progressive hedging," *Mathematical Programming*, to appear in 2018.

[27] Rockafellar, R. T., and Uryasev, S., "Minimizing buffered probability of exceedance by progressive hedging," *Mathematical Programmming B*, submitted.

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

[29] Rockafellar, R. T., and Wets, R. J-B, *Variational Analysis*, No. 317 in the series *Grundlehren der Mathematischen Wissenschaften*, Springer-Verlag, 1997 (third printing, with corrections: 2009).

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

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

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

[33] Watson, J.-P., and Woodruff, D.L., "Progressive hedging innovations for a class of stochastic mixed-integer resource allocation problems," *Computational Management Science* **8** (2010), 355–370.