LOCAL ELICITATION OF MONOTONICITY IN SOLVING GENERALIZED EQUATIONS AND PROBLEMS OF OPTIMIZATION

Terry Rockafellar
University of Washington, Seattle
University of Florida, Gainesville

Jonathan M. Borwein Commemorative Converence
Newcastle, Australia
25–29 September 2017

Context: Convex Functions and Variational Analysis

$\longrightarrow$ another domain in which Jon Borwein participated heavily!

Optimization: starting with a research explosion in the 1950s, it was realized that convex versus nonconvex was the real watershed in optimization, instead of the traditional linear versus nonlinear.

Convex functions: studies of the role of convexity and inequality constraints in optimization led to very new prespectives where:

- real-valued functions could instead be extended-real-valued,
- their graphs needed to be replaced by their epigraphs,
- set-valued subgradient mappings can generalize differentiation.

Variational analysis: extensions beyond convexity led to:

- variational geometry with unilateral tangents and normals,
- variational convergence/epiconvergence replacing pointwise,
- variational ineqs./generalized equations, variational principles,
- monotonicity properties that generalize "positive definiteness."

Overview of This Talk

- Problems of finding a zero of a maximal monotone mapping have many applications in optimization, equilibrium, and elsewhere
- The proximal point algorithm generates, from any starting point, a sequence that converges to some particular solution
- Spingarn (1983) invoked that for a partial inverse of the mapping to get a procedure conducive to problem decomposition
- Rock. and Wets (1991) followed that lead in developing the progressive hedging algorithm in convex stochastic programmming
- Penannen (2002) showed how localized max monotonicity in the proximal point algorithm guarantees local convergence

These ideas can be articulated even if local max monotonicity is just elicitable. In optimization that is how augmented Lagrangians are able to support multiplier methods through second-order theory

Problem Format for Discussion and Elaboration

Ingredients:

- a Hilbert space H, taken here to be finite-dimensional
- some subspace $L \supset H$ with orthogonal complement $L^\perp$
- some set-valued mapping $T : H \rightrightarrows H$

$$\mathrm{gph}\, T = \left\{ (z, w) \,\middle|\, w \in T(z) \right\} \subset H \times H$$

Basic Problem

determine $\bar{z} \in L$ and $\bar{w} \in L^\perp$ such that $\bar{w} \in T(\bar{z})$

The "monotone" case of this: T being maximal monotone

monotonicity: $\langle w' - w, z' - z \rangle \geq 0$ when $w \in T(z)$, $w' \in T(z')$

maximality: $\nexists$ monotone $T' \neq T$ with $\mathrm{gph}\, T' \supset \mathrm{gph}\, T$

Subspace interpretation: L stands for a "linkage constraint"

Connection with Variational Inequalities

Variational inequality: $-F(\bar{z}) \in N_C(\bar{z})$ for

$F : H \to H$ some continuous mapping

$C \subset H$ some nonempty closed convex set

$N_C =$ the normal cone mapping associated with C,

$v \in N_C(z) \iff z \in C,\ \langle v, z' - z\rangle \le 0,\ \forall z' \in C$

Variational inequality with linkage: $C = L \cap B$

where $L \subset H$ is a subspace, $B \subset H$ is a closed convex set

Normal cone formula: assuming $L \cap \operatorname{ri} B \neq \emptyset$, say

$$N_{L\cap B}(z) = N_L(z) + N_B(z) \text{ with } N_L(z) = \begin{cases} L^\perp & \text{if } z \in L \\ \emptyset & \text{if } z \notin L \end{cases}$$

Reduction to the basic problem: taking $T = F + N_B$

$$-F(\bar{z}) \in N_{L\cap B}(\bar{z}) \iff \bar{z} \in L \text{ and } \exists \bar{w} \in L^\perp \cap (F + N_B)(\bar{z})$$

Monotone case: T is max mono. if F is monotone rel. to B,

$$\langle F(z') - F(z), z' - z\rangle \ge 0 \ \ \forall z', z \in B$$

Partial Inverse Approach of Spingarn, 1983

- Represent H as the product space $L \times L^{\perp}$
- Write z and w as (x, u) and (v, y) with $x, v \in L$, $u, y \in L^{\perp}$

Then $L = \{(x, u) \,|\, u = 0\}$, $L^{\perp} = \{(v, y) \,|\, v = 0\}$, so that

$\bar{z} \in L$, $\bar{w} \in L^{\perp}$, $\bar{w} \in T(\bar{z})$ corresponds to $(0, \bar{y}) \in T(\bar{x}, 0)$

Partial inverse: of T with respect to L

$\widetilde{T} : H \rightrightarrows H$ defined by $(v, u) \in \widetilde{T}(x, y) \iff (v, y) \in T(x, u)$

Then $(0, \bar{y}) \in T(\bar{x}, 0)$ corresponds to $(0, 0) \in \widetilde{T}(\bar{x}, \bar{y})$

Spingarn's observation: T is max mono. $\iff$ $\widetilde{T}$ is max mono.

The proximal point algorithm can then be applied to $\widetilde{T}$ to solve $(0, 0) \in \widetilde{T}(\bar{x}, \bar{y})$ and thereby solve $\bar{z} \in L$, $\bar{w} \in L^{\perp}$, $\bar{w} \in T(\bar{z})$

The topic here: extending this beyond just the monotone case

Spingarn's Application of the Proximal Point Algorithm

Recall context: solving $(0,0) \in \widetilde{T}(\bar{x}, \bar{y})$ for the partial inverse
$$\widetilde{T}(x,y) = \{(v,u) \,|\, (v,y) \in T(x,u)\}$$ under max monotonicity

Proximal point iterations: generating (x_k, y_k) for $k = 1, 2, \ldots$
$$(x_{k+1}, y_{k+1}) = [I + r^{-1}\widetilde{T}\,]^{-1}(x_k, y_k), \quad r > 0$$

Elaboration: this works out in terms of T and the notation
$$u_{k+1} = r^{-1}[y_{k+1} - y_k], \qquad y_{k+1} = y_k - ru_{k+1},$$
to mean $(0,0) \in T_k(x_{k+1}, u_{k+1})$ for the max monotone mapping
$$T_k(x,u) = T(x,u) - (0, y_k) + r[(x,u) - (x_k, 0)]$$
Reverting to earlier notation by letting z_k and w_k stand for $(x_k, 0)$ and $(0, y_k)$, and $\hat{z}_{k+1}$ for (x_{k+1}, u_{k+1}), we get iterations as follows

- from $z_k \in L$ and $w_k \in L^{\perp}$ determine $\hat{z}_{k+1}$ by solving
  $0 \in T_k(\hat{z}_{k+1})$ where $T_k(z) = [T + rI\,](z) - [w_k + rz_k]$
- take $z_{k+1} = P_L(\hat{z}_k)$ (projection), $w_{k+1} = w_k - r[z_{k+1} - \hat{z}_{k+1}]$

Motivating Background in Optimization Duality

Framework: Hilbert spaces X and U, finite-dimensional,
some lsc proper function $f : X \times U \to (-\infty, \infty]$

Optimization problem: minimize $f(x, 0)$ with respect to x
perturbed version: minimize $f(x, u)$ in x for some $u \neq 0$

Typical first-order condition: utilizing general subgradients
$\bar{x}$ locally optimal $\longleftrightarrow$ $\exists \bar{y} \in U$ such that $(0, \bar{y}) \in \partial f(\bar{x}, 0)$

Connection to the basic problem: through its portrayal above
$H = X \times U, \;\; L = X \times \{0\}, \;\; L^{\perp} = \{0\}, \;\; T = \partial f$

Convex case of this: $f(x, u)$ is a convex function of (x, u)
then $T = \partial f$ is maximal monotone

Duality: then too, $(0, \bar{y}) \in \partial f(\bar{x}, 0) \iff (\bar{x}, 0) \in \partial f^*(0, \bar{y})$

Associated dual problem: minimize $f^*(0, y)$ with respect to y
peturbed version: minimize $f^*(v, y)$ in y for some $v \neq 0$

Partial Inverse Interpretation in the Optimization Setting

Lagrangian function: $l(x,y) = \inf_u\{f(x,u) - \langle y,u\rangle\}$

$f(x,u)$ convex in u $\Rightarrow f(x,u) = \sup_y\{l(x,y) + \langle y,u\rangle\}$

Common situation in subgradient calculus:

$$(v,y) \in \partial f(x,u) \quad \Longleftrightarrow \quad (v,-u) \in \partial l(x,y)$$

and then the partial inverse of ∂f can be identified with

$$\widetilde{T} : (x,y) \rightrightarrows \{(v,u) \,|\, (v,-u) \in \partial l(x,y)\}$$

Specialization to the convex case:

- this holds with $l(x,y)$ convex in x, concave in y
- $(v,-u) \in \partial l(x,y) \quad \Longleftrightarrow \quad v \in \partial_x l(x,y),\ u \in \partial_y[-l](x,y)$
- $(0,\bar{y}) \in \partial f(\bar{x},0) \quad \Longleftrightarrow \quad (\bar{x},\bar{y})$ is a Lagrangian saddle point
- Spingarn's application of the proximal point algorithm to $\widetilde{T}$ leads to subproblems involving the corresponding

Augmented Lagrangian function:

$$l_r(x,y) = \inf_u\{f(x,u) - \langle y,u\rangle + \tfrac{r}{2}||u||^2\},\ r > 0$$

Augmented Lagrangian Details in Nonlinear Programming

Problem: minimize $f(x,0)$ with respect to x where

$$f(x,u) = \delta_C(x) + g(x) + \delta_K(G(x)+u)$$

in the case of $g : X \to R$ and $G : X \to U$ both smooth, $C \subset X$ closed convex, $K \subset U$ closed convex cone

Corresponding Lagrangians — in terms of $Y = K^* =$ polar cone

$$l(x,y) = \delta_C(x) + g(x) + \langle y, G(x)\rangle - \delta_Y(y)$$

$$l_r(x,y) = \delta_C(x) + g(x) + \langle y, G(x)\rangle + \tfrac{r}{2}||G(x)||^2 - \tfrac{r}{2}\mathrm{dist}_Y^2(y + rG(x))$$

Convex case: $g(x) + \langle y, G(x)\rangle$ convex in $x \in C$ when $y \in Y$

Corresponding execution of Spingarn's algorithm

$$x_{k+1} = \operatorname*{argmin}_x \left\{ l_r(x,y_k) + \tfrac{r}{2}||x - x_k||^2 \right\}, \quad y_{k+1} = y_k - r[x_{k+1} - x_k]$$

= the <u>proximal method of multipliers</u> of Rockafellar (1976)!

Localization of the General Procedure and its Convergence

Definition: a mapping $T : H \rightrightarrows H$ is max monotone locally at $(\bar{z}, \bar{w}) \in \mathrm{gph}\, T$ if $\exists$ neighborhood N of $(\bar{z}, \bar{w})$ such that

- $\langle w' - w, z' - z \rangle \geq 0$ for all (z, w), (z', w'), in $N \cap \mathrm{gph}\, T$,
- $\mathrm{gph}\, T$ can't be extended in N without violating this condition

Local convergence theorem of Pennanen, 2002

The proximal point algorithm for finding $(\bar{z}, 0) \in \mathrm{gph}\, T$ converges **locally** to a solution if T is max monotone **locally** at a solution $(\bar{z}, 0)$ and the iterations proceed from some z_k close enough to $\bar{z}$

Application here: finding $\bar{z} \in L$, $\bar{w} \in L^{\perp}$, with $\bar{w} \in T(\bar{z})$

Pennanen's convergence result can be invoked for $\widetilde{T}$ instead of T

Local convergence of Spingarn's partial inverse procedure

The algorithm for finding $(\bar{z}, \bar{w}) \in (L \times L^{\perp}) \cap \mathrm{gph}\, T$ converges **locally** to a solution if T is max monotone **locally** at a solution $(\bar{z}, \bar{w})$ and the iterations go from (z_k, w_k) close enough to $(\bar{z}, \bar{w})$

Extension to Problems With "Elicitable" Monotonicity

Problem to be solved: find $\bar{z} \in L$, $\bar{w} \in L^{\perp}$, with $\bar{w} \in T(\bar{z})$

Projection device: Let $P_{L^{\perp}}$ be the projection on $L^{\perp}$, so that

$$P_{L^{\perp}} = I - P_L \quad \text{and} \quad z \in L \Leftrightarrow P_{L^{\perp}}(z) = 0$$

Observation: solutions are unaffected if T is replaced by

$$T_s = T + sP_{L^{\perp}} \text{ for some } s > 0$$

Definition of elicitability

Local maximal monotonicity is **elicitable** at a solution $(\bar{z}, \bar{w})$ at level $\bar{s} > 0$ if T_s is maximal monotone locally at $(\bar{z}, \bar{w})$ for $s \geq \bar{s}$

$\Longrightarrow$ the algorithm can be applied to T_s in place of T

Corresponding modification of the algorithm

- there are two parameters: $s > 0$ sufficiently high and $r > s$
- the w update is now $w_{k+1} = w_k - (r - s)[z_{k+1} - \hat{z}_{k+1}]$

Elicitation Via Augmented Lagrangians in Optimization

Recall framework: $T = \partial f$ for lsc $f(x,u)$ on $X \times U$ convex in u

$\widetilde{T} : (x,y) \rightrightarrows \{(v,u) \mid (v,-u) \in \partial l(x,y)\}$, where

$l(x,y) = \inf_u \{f(x,u) - \langle y,u \rangle\}$ Lagrangian

Algorithm derivation: apply the proximal point algorithm to $\widetilde{T}$

elicitation with $s > 0$: apply it instead to the partial inverse $\widetilde{T}_s$

where $T_s = T + sP_{L^\perp}$ for the projection $P_{L^\perp} : (x,u) \to (0,u)$

New perspective: $\widetilde{T}_s : (x,y) \rightrightarrows \{(v,u) \mid (v,-u) \in \partial l_s(x,y)\}$ for

$l_s(x,y) = \inf_u \{f(x,u) - \langle y,u \rangle + \frac{s}{2}||u||^2\}$ augmented Lagrangian

Key insight for elicitation

Local max monotonicity of $\widetilde{T}_s$ at a solution $(\bar{x},\bar{y})$ means that the augmented Lagrangian $l_s(x,y)$ is locally convex-concave at $(\bar{x},\bar{y})$

How realistic is it to rely on this holding when s is high enough?

Elicitation Specialized to Nonlinear Programming

Problem: minimize $g(x)$ subject to $x \in C$, $G(x) \in K$

i.e., min $f(x,0)$ for $f(x,u) = \delta_C(x) + g(x) + \delta_K(G(x)+u)$

C = closed convex set, K = closed convex cone, Y = polar

Augmented Lagrangian: with parameter $r > 0$

$$l_r(x,y) = \delta_C(x) + g(x) + \langle y, G(x)\rangle + \frac{r}{2}||G(x)||^2 - \frac{r}{2}\mathrm{dist}^2_Y(y + rG(x))$$

Standard case of NLP: $C = R^n$, $Y = R^q_+ \times R^{m-q}$

Known fact for the standard case of NLP

The so-called **strong second-order sufficient conditions** for optimality of $\bar{x}$ with multiplier vector $\bar{y}$ induce $l_r(x,y)$ to be convex-concave around $(\bar{x}, \bar{y})$ when $r > 0$ is sufficiently high

Conjecture: this holds beyond the standard case and even for much more general problem formats in optimization

second-order optimality theory needs further work to resolve this

More the Role of Second-Order Optimality

Traditional paradigm: develop second-order sufficient conditions that are as close as possible to second-order necessary conditions

Contemporary reality: problems are solved numerically and second-order conditions are the key to understanding convergence

Duality result in standard NLP — Rock. (1974)

$(\bar{x}, \bar{y})$ is a local saddle point of the augmented Lagrangian $l_r(x, y)$
$\iff$ $\bar{x}$ is locally optimal and the function $p(u) = \inf_x f(x, u)$
has the property that $p(u) \geq p(0) - \langle \bar{y}, u \rangle - \frac{r}{2}||u||^2$ for u near 0

convexity-concavity of $l_r(x, y)$ near $(\bar{x}, \bar{y})$ extends this to a nbhood

Conjecture about the general analog of SSOC beyond NLP

For the theory of augmented Lagrangians much more broadly, this should be the existence of a neighborhood of $(\bar{x}, \bar{y})$ on which, for r high enough, $l_r(x, y)$ is concave in y but strongly convex in x

References

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

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

[3 R.T. Rockafellar (1974) "Augmented Lagrange multiplier functions and duality in nonlinear programming," *SIAM J. Control* 12, 268–285.

[4] R.T. Rockafellar, R.J-B Wets (1998) *Variational Analysis*, No. 317 in the series *Grundlagen der Mathematischen Wissenschaften*, Springer-Verlag.

website: www.math.washington.edu/∼rtr/mypage.html