CRITICAL MULTIPLIERS IN VARIATIONAL SYSTEMS VIA SECOND-ORDER GENERALIZED DIFFERENTIATION

BORIS MORDUKHOVICH

Wayne State University

talk given at the **Workshop on Variational and Stochastic Analysis**

based on joint work with **Ebrahim Sarabi**(Miami U.)

Santiago, Chile, March 2017

Supported by NSF grant DMS-1512846 and
by the US Air Force grant 15RT0462

VARIATIONAL SYSTEMS

Given C^2 mappings $\Phi\colon I\!R^n \to I\!R^m$, $\Psi\colon I\!R^n \times I\!R^m \to I\!R^l$ and a convex extended-real-valued function $\theta\colon I\!R^m \to \overline{I\!R} := (-\infty, \infty]$, consider the variational system (VS)

$$\Psi(x, v) = 0,\ v \in \partial\theta\Big(\Phi(x)\Big)$$

It can be treated as a generalized KKT system being reduced to the classical KKT for nonlinear programs (NLPs) with equality and inequality constraints when $\theta = \delta_\Omega$ with $\Omega := I\!R^s \times I\!R_-^{m-s}$

We consider a significantly more general class of θ for which (VS) cover various smooth and nonsmooth optimization problems and may not be even related to optimization

CONVEX PIECEWISE LINEAR FUNCTIONS

Recall that $\theta\colon I\!R^m \to \overline{I\!R}$ is convex piecewise linear, $\theta \in CPWL$, if it admits the following equivalent descriptions

- The epigraph $\operatorname{epi}\theta$ is a convex polyhedron in $I\!R^{m+1}$

- There are $\alpha_i \in I\!R$, $l \in I\!N$, and $a_i \in I\!R^m$ for $i \in T_1 : = \{1, \ldots, l\}$ such that θ is represented by

$$\theta(z) = \max\Big\{\langle a_1, z\rangle - \alpha_1, \ldots, \langle a_l, z\rangle - \alpha_l\Big\}, \qquad z \in \operatorname{dom}\theta$$

and $\theta(z) = \infty$ otherwise, where the domain $\operatorname{dom}\theta$ is a convex polyhedron given by

$$\operatorname{dom}\theta = \Big\{z \in I\!R^m \Big|\ \langle d_i, z\rangle \le \beta_i, \quad i \in T_2 := \{1, \ldots, p\}\Big\}$$

with some $d_i \in I\!R^m$, $\beta_i \in I\!R$, $p \in I\!N$

THE SETTING

Impose the following connection between Φ and Ψ in (VS)

$$\Psi(x,v) = f(x) + \nabla\Phi(x)^* v, \quad (x,v) \in I\!R^n \times I\!R^m$$

where $f\colon I\!R^n \to I\!R^n$ is a smooth mapping and where A^* stands for the matrix transposition. Consider $\bar{x} \in I\!R^n$ satisfying the stationarity condition

$$0 \in f(\bar{x}) + \partial(\theta \circ \Phi)(\bar{x})$$

and define the set of Lagrange multipliers associated with $\bar{x}$ by

$$\Lambda(\bar{x}) = \left\{ v \in I\!R^m \middle| \, \Psi(\bar{x},v) = 0, \; v \in \partial\theta(\bar{z}) \right\}$$

with $\bar{z} = \Phi(\bar{x})$ in what follows

GRAPHICAL DERIVATIVE

The graphical derivative of a set-valued mapping $F\colon I\!R^n \rightrightarrows I\!R^p$ at $(\bar{x},\bar{y}) \in \operatorname{gph} F$ is

$$DF(\bar{x},\bar{y})(u) = \Big\{v \in I\!R^p \Big|\ (u,v) \in T\Big((\bar{x},\bar{y}); \operatorname{gph} F\Big)\Big\}, \quad u \in I\!R^n$$

where $T(z;\Omega)$ is a contingent cone to Ω at z defined by

$$T(z;\Omega) := \Big\{w \in I\!R^m \Big|\ \exists\, z_k \stackrel{\Omega}{\to} z,\ \alpha_k \ge 0,\ \alpha_k(z_k - z) \to w\Big\}$$

We use the 2nd-order construction $D\partial\theta$ noting that for $\theta \in C^2$

$$\Big(D\partial\theta\Big)\Big(\bar{z},\theta(\bar{z})\Big)(u) = \Big\{\nabla^2\theta(\bar{z})u\Big\}, \quad u \in I\!R^m$$

For $\theta \in CPWL$ the construction $D\partial\theta$ is explicitly calculated entirely via the given data of θ

CRITICAL AND NONCRITICAL MULTIPLIERS

DEFINITION Let $\bar{x}$ be a stationary point of (VS). Then $\bar{v} \in \Lambda(\bar{x})$ is a critical multiplier if there exists $0 \neq \xi \in I\!R^n$ satisfying

$$0 \in \nabla_x \Psi(\bar{x}, \bar{v})\xi + \nabla\Phi(\bar{x})^* \Big(D\partial\theta\Big)(\bar{z}, \bar{v})\Big(\nabla\Phi(\bar{x})\xi\Big)$$

The multiplier $\bar{v} \in \Lambda(\bar{x})$ is noncritical for (VS) otherwise

These notions extend those introduced and developed by Izmailov and Solodov (2005+) for NLPs and related smooth KKT. It has been recognized that critical multipliers are largely responsible for slow convergence of major primal-dual algorithms of optimization and thus should be ruled out for appropriate classes of stationary/optimal solutions

CANONICAL PERTURBATIONS

Consider canonically perturbed variational system (PVS) with the parameter pair $(p_1, p_2) \in I\!R^n \times I\!R^m$

$$\begin{bmatrix} p_1 \\ p_2 \end{bmatrix} \in \begin{bmatrix} \Psi(x,v) \\ -\Phi(x) \end{bmatrix} + \begin{bmatrix} 0 \\ (\partial\theta)^{-1}(v) \end{bmatrix}$$

Define the set-valued mapping

$$G(x,v) := \begin{bmatrix} \Psi(x,v) \\ -\Phi(x) \end{bmatrix} + \begin{bmatrix} 0 \\ (\partial\theta)^{-1}(v) \end{bmatrix}$$

and the solution map to (PVS) by

$$S(p_1,p_2) := \Big\{(x,v) \in I\!R^n \times I\!R^m \Big| \ (p_1,p_2) \in G(x,v)\Big\}.$$

ERROR BOUND CHARACT. OF NONCRITICAL MULTIPLIERS

THEOREM The following properties are equivalent

(i) The Lagrange multiplier $\bar{v} \in \Lambda(\bar{x})$ is noncritical for (VS)

(ii) There are numbers $\varepsilon > 0$, $\ell \geq 0$ and neighborhoods U of $0 \in I\!R^n$ and W of $0 \in I\!R^m$ such that for any $(p_1, p_2) \in U \times W$ and any $(x_{p_1 p_2}, v_{p_1 p_2}) \in S(p_1, p_2) \cap I\!B_\varepsilon(\bar{x}, \bar{v})$ we have

$$\|x_{p_1 p_2} - \bar{x}\| + \text{dist}\Big(v_{p_1 p_2}; \Lambda(\bar{x})\Big) \leq \ell\Big(\|p_1\| + \|p_2\|\Big)$$

COMPOSITE OPTIMIZATION

The general scheme of (CO)

$$\text{minimize} \;\; \varphi(x) := \varphi_0(x) + \theta\Big(\Phi(x)\Big), \quad x \in \mathbb{R}^n$$

where $\varphi_0 \colon \mathbb{R}^n \to \mathbb{R}$ and $\Phi \colon \mathbb{R}^n \to \mathbb{R}^m$ are C^2 while $\theta \colon \mathbb{R}^m \to \overline{\mathbb{R}}$ with $\theta \in CPWL$. Implicit constraints $\Phi(x) \in \operatorname{dom} \theta$.

Consider the Lagrangian

$$L(x, v) = \varphi_0(x) + \langle \Phi(x), v \rangle, \quad (x, v) \in \mathbb{R}^n \times \mathbb{R}^m$$

and the collection of Lagrange multipliers

$$\Lambda_{\text{com}}(\bar{x}) := \Big\{ v \in \mathbb{R}^m \Big| \; \nabla_x L(\bar{x}, v) = 0, \; v \in \partial\theta(\bar{z}) \Big\}$$

SECOND-ORDER SUFFICIENT CONDITION FOR (CO)

We say the second-order sufficient condition (SOSC) holds for a feasible solution $\bar{x}$ to (CO) with $\bar{v} \in \Lambda_{\text{com}}(\bar{x})$, $\bar{z} = \Phi(\bar{x})$ if

$$\langle \nabla^2_{xx} L(\bar{x}, \bar{v})u, u\rangle > 0 \quad \text{for all} \quad 0 \neq u \in I\!R^n \quad \text{with} \quad \nabla\Phi(\bar{x})u \in \mathcal{K}(\bar{z}, \bar{v})$$

where $\mathcal{K}(\bar{z}, \bar{v})$ is the critical cone for θ at $(\bar{z}, \bar{v}) \in \text{gph}\, \partial\theta$ which is calculated via the given data of $\theta \in CPWL$

THEOREM SOSC ensures that $\bar{x}$ a strict local minimizer for (CO). Furthermore, any $\bar{v} \in \Lambda_{\text{com}}(\bar{x})$ for which SOSC holds is a noncritical multiplier for (CO) associated with $\bar{x}$

FULL STABILITY OF LOCAL MINIMIZERS

Consider the canonical perturbation (CP) of (CO)

$$\text{minimize} \;\; \varphi_0(x) + \theta\Big(\Phi(x) + p_2\Big) - \langle p_1, x\rangle, \quad x \in I\!R^n$$

For fixed $\gamma > 0$ and parameters $(p_1, p_2) \in I\!R^n \times I\!R^l$ define

$$m_\gamma(p_1, p_2) = \inf_{\|x-\bar{x}\|\le\gamma} \Big\{\varphi_0(x) + \theta(\Phi(x) + p_2) - \langle p_1, x\rangle\Big\}$$

$$M_\gamma(p_1, p_2) = \text{argmin}\Big\{\varphi_0(x) + \theta(\Phi(x) + p_2) - \langle p_1, x\rangle \Big| \; \|x - \bar{x}\| \le \gamma\Big\}$$

Then $\bar{x}$ is a fully stable locally optimal solution to CP if the mapping $(p_1, p_2) \mapsto M_\gamma(p_1, p_2)$ is locally single-valued and Lipschitzian with $M_\gamma(0, 0) = \{\bar{x}\}$ and the function $(p_1, p_2) \mapsto m_\gamma(p_1, p_2)$ is locally Lipschitzian around $(0, 0)$ for some $\gamma > 0$

EXCLUDING CRITICAL MULTIPLIERS

THEOREM Let $\bar{x}$ be a fully stable locally optimal solution to (CP). Then the Lagrange multiplier set $\Lambda_{\text{com}}(\bar{x})$ in does not include any critical multipliers

By now we have complete second-order characterizations of full stability for various classes of optimization and optimal control problems as well as variational systems. This allows us to efficiently determine settings where critical multipliers do not appear and thus slow convergence is eliminated

Tilt stability ($p_2 = 0$) may not rule out critical multipliers, but it does under certain nondegeneracy conditions as well as in some other case for NLPs

ROBUST ISOLATED CALMNESS

DEFINITION A set-valued mapping $F\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ enjoys the robust isolated calmness property at $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ if there exist $\ell \geq 0$ and neighborhoods U of $\bar{x}$, V of $\bar{y}$ such that

$$F(x) \cap V \subset \{\bar{y}\} + \ell \|x - \bar{x}\| \mathbb{B} \quad \text{for all} \quad x \in U$$

together with the condition

$$F(x) \cap V \neq \emptyset \text{ for all } \quad x \in U$$

CHARACT. OF ROBUST ISOLATED CALMNESS FOR (CO)

THEOREM Let $\bar{x}$ be a feasible solution to (CO) with $\theta \in CPWL$. Then the following are equivalent

(i) The KKT solution map

$$S_{\mathsf{KKT}}(p_1, p_2) = \Big\{ (x, v) \in \mathbb{R}^n \times \mathbb{R}^m \Big| \; p_1 = \nabla_x L(x, v), \; v \in \partial\theta\big(p_2 + \Phi(\bar{x})\big) \Big\}$$

is robustly isolatedly calm at $((0,0), (\bar{x}, \bar{v})) \in \mathbb{R}^{n+m} \times \mathbb{R}^{n+m}$ and $\bar{x}$ is a locally optimal solution to (CO).

(ii) SOSC holds and $\Lambda_{\mathsf{com}}(\bar{x}) = \{\bar{v}\}$

(iii) $\Lambda_{\mathsf{com}}(\bar{x}) = \{\bar{v}\}$, $\bar{x}$ is a locally optimal solution to (CO), and $\bar{v}$ is a noncritical multipliers for with $\Psi = \nabla_x L$

(iv) S_{KKT} is isolatedly calm at $((0,0), (\bar{x}, \bar{v}))$ and $\bar{x}$ is a locally optimal solution to (CO)

REFERENCES

1. A. F. Izmailov and M. V. Solodov, Newton-Type Methods for Optimization and Variational Problems, Springer, 2014

2. A. B. Levy, R. A. Poliquin and R. T. Rockafellar, Stability of locally optimal solutions, SIAM J. Optim. **10** (2000), 580–604

3. B. S. Mordukhovich, T. T. A. Nghia and R. T. Rockafellar, Full stability in finite-dimensional optimization, Math. Oper. Res. **40** (2015), 226–252

4. B. S. Mordukhovich and M. E. Sarabi, Generalized differentiation of piecewise linear functions in second-order variational analysis, Nonlinear Anal. **132** (2016), 240–273

5. B. S. Mordukhovich and M. E. Sarabi, Critical multipliers in variational systems via second-order generalized differentiation, to appear in Math. Program.

6. R. T. Rockafellar and R. J-B. Wets, Variational Analysis, Springer, 1998