CRITICALITY OF LAGRANGE MULTIPLIERS IN CONSTRAINED OPTIMIZATION WITH APPLICATIONS TO SQP

BORIS MORDUKHOVICH

Wayne State University, USA

Lecture 2 at the **Forum on Developments and Origins of Operations Research**

Shenzhen, China, November 2021

ORGANIZERS: **Operations Research Society of China & Southern University of Science and Technology**

VARIATIONAL SYSTEMS

Consider the variational system (VS)

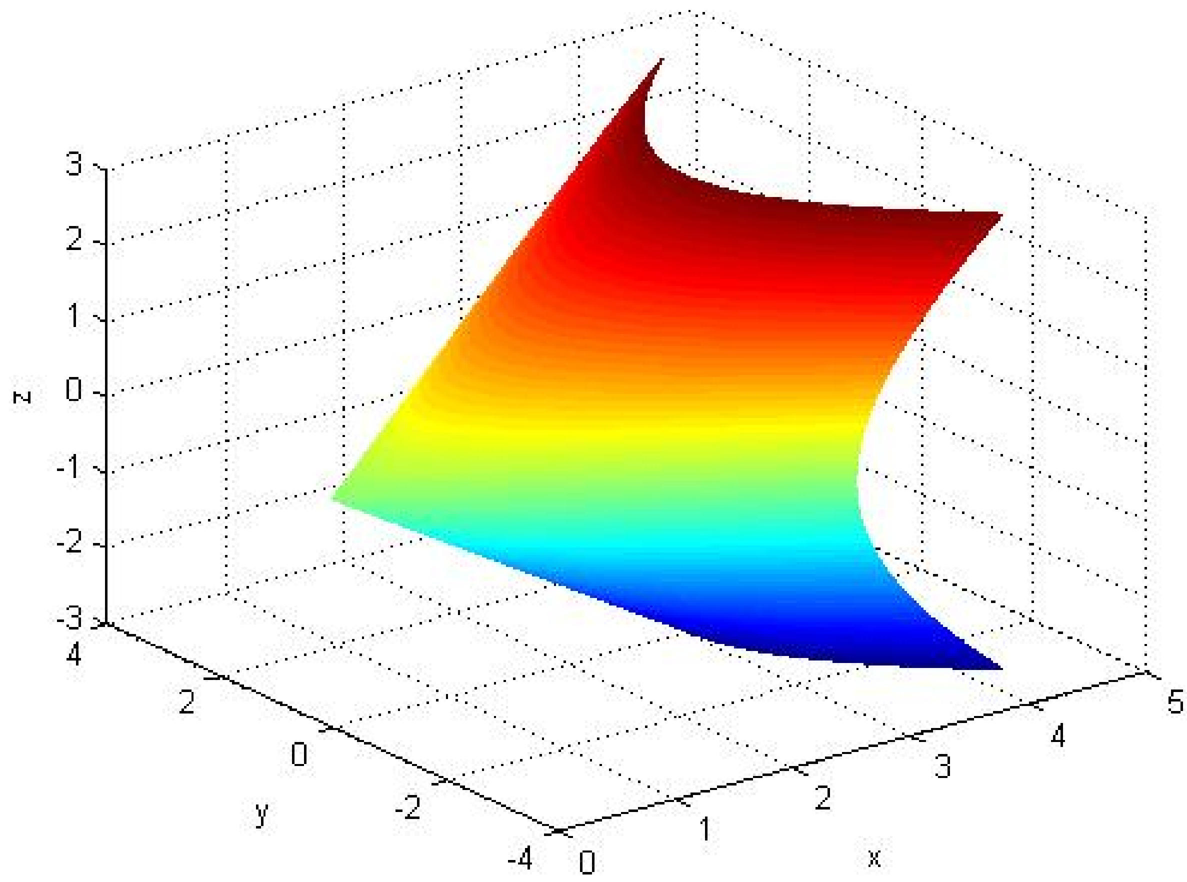
$$\Psi(x,\lambda) := f(x) + \nabla\Phi(x)^*\lambda = 0, \qquad \lambda \in N_\Theta\big(\Phi(x)\big)$$

where $f\colon I\!R^n \to I\!R^n$ is a $\mathcal{C}^1$ while $\Phi\colon I\!R^n \to I\!R^m$ is a $\mathcal{C}^2$ mappings, $\Theta \subset I\!R^m$ is a closed set with N_Θ standing for its (limiting) normal cone. A major source for such systems comes from KKT in constrained optimization (CO)

$$\text{minimize } \varphi_0(x) \text{ subject to } \Phi(x) \in \Theta$$

with $f = \nabla\varphi_0$. When Θ is a convex cone, the latter problems are known as conic programs

ICE-CREAM CONE IN 3D



SUBGRADIENT GRAPHICAL DERIVATIVE

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

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

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

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

We use the **2nd-order subgradient graphical derivative** construction $D\partial\theta$ while noting that for $\theta \in \mathcal{C}^2$ we have

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

For various important classes of extended-real-valued functions θ considered below the construction $D\partial\theta$ is explicitly calculated entirely via the given data of θ

CRITICAL AND NONCRITICAL MULTIPLIERS

Given $\bar{x} \in I\!R^n$ satisfying the stationary condition

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

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

$$\Lambda(\bar{x}) := \Big\{\lambda \in I\!R^m \;\Big|\; \Psi(\bar{x}, \lambda) = 0,\; \lambda \in N_{\Theta}\Big(\Phi(\bar{x})\Big)\Big\}$$

DEFINITION (BM-Sarabi18) The multiplier $\bar{\lambda} \in \Lambda(\bar{x})$ is critical if there is $\xi \neq 0$ satisfying

$$0 \in \nabla_x \Psi(\bar{x}, \bar{\lambda})\xi + \nabla\Phi(\bar{x})^* D N_{\Theta}(\Phi(\bar{x}), \bar{\lambda})\Big(\nabla\Phi(\bar{x})\xi\Big)$$

The Lagrange multiplier $\bar{\lambda} \in \Lambda(\bar{x})$ is noncritical when the above inclusion admits only the trivial solution $\xi = 0$.

ORIGINAL DEFINITION AND DISCUSSION

ORIGINAL DEFINITION (Izmailov05) for $\theta = \delta_{\{0\}^m}$ (i.e. for KKT systems in NLPs with smooth equality constraints): the multiplier $\bar{\lambda}$ is critical if the primal-dual system

$$\nabla_x \Psi(\bar{x}, \bar{\lambda})\xi \in \mathrm{rge}\,\nabla\Phi(\bar{x})^*, \quad \nabla\Phi(\bar{x})\xi = 0$$

admits a nontrivial solution

Izmailov, Solodov, and their collaborators demonstrate that critical multipliers are largely responsible for slow convergence of major primal-dual algorithms of optimization. This is due to the fact that the set of critical multipliers is an attractor for the dual sequence of multipliers $\{\lambda_k\}$ in such algorithms and thus slow down the convergence of the primal sequence $\{x_k\}$ in numerical methods. Therefore, critical multipliers should be ruled out for appropriate classes of stationary/optimal solutions

REDUCIBLE SETS

DEFINITION A closed set $\Theta \subset I\!R^m$ is $\mathcal{C}^2$-cone reducible at $\bar{z} = \Phi(\bar{x}) \in \Theta$ to a closed convex subcone $C \subset I\!R^p$ if there exist a neighborhood $\mathcal{O} \subset I\!R^m$ of $\bar{z}$ and a $\mathcal{C}^2$-smooth mapping $h\colon I\!R^m \to I\!R^p$ such that

$$\Theta \cap \mathcal{O} = \left\{ z \in \mathcal{O} \,\middle|\, h(z) \in C \right\}, \quad h(\bar{z}) = 0,\ \nabla h(\bar{z}) \text{ is surjective}$$

In contrast to the classical definition by Bonnans and Shapiro (2000), we do not assume that C is pointed and Θ is convex

EXAMPLES: second-order/Lorentz/ice-cream cone and their products, SDP cone, copositive cone, and other major constraint systems in conic programming.

CRITICAL MULTIPLIERS FOR REDUCIBLE SETS

THEOREM Let $\bar{x}$ be a stationary point, $\bar{\lambda} \in \Lambda(\bar{x})$, Θ be $\mathcal{C}^2$-cone reducible at $\bar{z} := \Phi(\bar{x})$ to a closed convex cone C, let

$$K_\Theta(\bar{z}, \bar{\lambda}) = T_\Theta(\bar{z}) \cap \{\bar{\lambda}\}^\perp \quad \text{with} \quad \bar{\lambda} \in N_\Theta(\bar{z})$$

be the critical cone to Θ at $\bar{x}$ for $\bar{\lambda}$, and let $\bar{\mu}$ be a unique solution to the system

$$\bar{\lambda} = \nabla h(\bar{z})^* \bar{\mu}, \quad \bar{\mu} \in N_C\big(h(\bar{z})\big)$$

Then $\bar{\lambda}$ is critical multiplier if and only if the system

$$\begin{aligned} &\nabla_x \Psi(\bar{x}, \bar{\lambda})\xi + \nabla\Phi(\bar{x})^* \eta = 0 \\ &\eta - \nabla^2 \langle \bar{\mu}, h \rangle(\bar{z}) \nabla\Phi(\bar{x})\xi \in N_{K_\Theta(\bar{z}, \bar{\lambda})}\big(\nabla\Phi(\bar{x})\xi\big) \end{aligned}$$

admits a solution $(\xi, \eta) \in I\!R^n \times I\!R^m$ such that $\xi \neq 0$

CALMNESS PROPERTIES OF MULTIFUNCTIONS

DEFINITION We say that a set-valued mapping $F\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is calm at $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ if there exist $\ell \geq 0$ and a neighborhood U of $\bar{x}$ such that

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

where $\mathbb{B}$ stands for the closed unit ball. The isolated calmness property of F at $(\bar{x}, \bar{y})$ is defined by

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

It is well known that the calmness and isolated calmness properties are equivalent to metric subregularity and strong metric subregularity of the inverse mapping F^{-1} at $(\bar{y}, \bar{x})$, respectively

CANONICAL PERTURBATION AND SEMI-ISOLATED CALMNESS

Consider the mapping $G\colon I\!R^n \times I\!R^m \rightrightarrows I\!R^n \times I\!R^m$ given by

$$G(x,\lambda) := \begin{bmatrix} \Psi(x,\lambda) \\ -\Phi(x) \end{bmatrix} + \begin{bmatrix} 0 \\ N_{\Theta}^{-1}(\lambda) \end{bmatrix}$$

and the solution map to the canonical perturbation of VS

$$S(v,w) := \Big\{(x,\lambda) \in I\!R^n \times I\!R^m \;\Big|\; (v,w) \in G(x,\lambda)\Big\},\ (v,w) \in I\!R^n \times I\!R^m$$

DEFINITION Given a solution $(\bar{x},\bar{\lambda})$ to (VS), we say the semi-isolated calmness holds for S if the exist $\varepsilon > 0$, $\ell \geq 0$, and neighborhoods $V \times W$ of $(0,0) \in I\!R^n \times I\!R^m$ such that

$$S(v,w) \cap I\!B_\varepsilon(\bar{x},\bar{\lambda}) \subset \Big[\{\bar{x}\} \times \Lambda(\bar{x})\Big] + \ell\Big(\|v\| + \|w\|\Big) I\!B \quad \text{for} \quad (v,w) \in V \times W$$

CHARACTERIZATIONS OF NONCRITICAL MULTIPLIERS

THEOREM Let $(\bar{x},\bar{\lambda})$ be a solution to the variational system (VS). Consider the properties

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

(ii) There are numbers $\varepsilon>0$, $\ell\ge 0$ and neighborhoods (V,W) of $(0,0)\in I\!R^n\times I\!R^m$ such the semi-isolated calmness holds

(iii) There are numbers $\varepsilon>0$ and $\ell\ge 0$ with the error bound

$$\|x-\bar{x}\|+d\Big(\lambda;\Lambda(\bar{x})\Big)\le\ell\Big(\|\Psi(x,\lambda)\|+d\Big(\Phi(x);N_\Theta^{-1}(\lambda)\Big)\Big),\ (x,\lambda)\in I\!B_\varepsilon(\bar{x},\bar{\lambda})$$

Then we have the assertions

(a) Implications (iii)$\Longleftrightarrow$(ii)$\Longrightarrow$(i) always fulfill

(b) If Θ is $\mathcal{C}^2$-cone reducible at $\bar{z}=\Phi(\bar{x})$ to a closed convex cone C, if the set

$$K_\Theta(\bar{z},\bar{\lambda})^*-\Big[K_\Theta(\bar{z},\bar{\lambda})^*\cap\ker\nabla\Phi(\bar{x})^*\Big]$$

is closed, and if the Lagrange multiplier mapping

$$M_{\bar{x}}(v,w) := \Big\{\lambda \in I\!R^m \;\Big|\; (v,w) \in G(\bar{x},\lambda)\Big\} \text{ for all } (v,w) \in I\!R^n \times I\!R^m$$

is calm at $((0,0),\bar{\lambda})$, then the converse (i)$\Longrightarrow$(ii) also holds

Note that the imposed calmness and closedness assumptions hold automatically if Θ is a convex polyhedron (by Hoffman's lemma). Otherwise, they are essential for noncriticality

SPECIAL CASE

THEOREM Let $(\bar{x}, \bar{\lambda})$ be a solution to the variational system (VS). Then the following are equivalent
(i) The Lagrange multiplier $\bar{\lambda} \in \Lambda(\bar{x})$ is noncritical, $\Lambda(\bar{x}) = \{\bar{\lambda}\}$, and the Lagrange multiplier mapping $M_{\bar{x}}$ is calm at $((0,0), \bar{\lambda})$
(ii) The solution mapping S is isolated calm at $((0,0), (\bar{x}, \bar{\lambda}))$

Here closedness assumption holds automatically since we can show that

$$K_{\Theta}(\bar{z}, \bar{\lambda})^* - \underbrace{\Big[K_{\Theta}(\bar{z}, \bar{\lambda})^* \cap \ker \nabla\Phi(\bar{x})^*\Big]}_{=\{0\}} = K_{\Theta}(\bar{z}, \bar{\lambda})^*$$

SECOND-ORDER CONDITIONS

DEFINITION Let $(\bar{x}, \bar{\lambda})$ be a solution to (VS) under the $\mathcal{C}^2$-cone reducibility of Θ. The the second-order condition is

$$\left\langle \nabla_x \Psi(\bar{x}, \bar{\lambda})\xi, \xi \right\rangle + \left\langle \nabla^2 \langle \bar{\mu}, h \rangle (\bar{z}) \nabla \Phi(\bar{x})\xi, \nabla \Phi(\bar{x})\xi \right\rangle > 0$$
$$\text{for all } \; 0 \neq \xi \in I\!R^n \; \text{ with } \; \nabla \Phi(\bar{x})\xi \in K_\Theta(\bar{z}, \bar{\lambda})$$

In the case of constrained optimization (CO) this condition reduces to the second-order sufficient condition

$$\left\langle \nabla^2_{xx} L(\bar{x}, \bar{\lambda})u, u \right\rangle + \left\langle \nabla^2 \langle \bar{\mu}, h \rangle (\bar{z}) \nabla \Phi(\bar{x})u, \nabla \Phi(\bar{x})u \right\rangle > 0$$
$$\text{for all } \; 0 \neq u \in I\!R^n \; \text{ with } \; \nabla \Phi(\bar{x})u \in K_\Theta(\bar{z}, \bar{\lambda})$$

which can be equivalent described via the sigma term. The latter is a bit stronger than the classical second-order sufficient condition for (CO) corresponding to

$$\sup_{\bar{\lambda} \in \Lambda_c(\bar{x})} \left\{ \left\langle \nabla^2_{xx} L(\bar{x}, \bar{\lambda})u, u \right\rangle \ldots \right\}$$

while it ensures the multiplier noncriticality

NONCRITICALITY UNDER SECOND-ORDER CONDITION

THEOREM Let $(\bar{x}, \bar{\lambda})$ be a solution to (VS), let Θ be $\mathcal{C}^2$-cone reducible at $\bar{z} = \Phi(\bar{x})$ to a closed convex cone C, and let the multiplier mapping $M_{\bar{x}}$ be calm at $((0,0), \bar{\lambda})$. If the second-order condition holds, then the solution map S from is semi-isolatedly calm at $((0,0), (\bar{x}, \bar{\lambda}))$, and hence $\bar{\lambda}$ is a noncritical multiplier corresponding to $\bar{x}$

NONCRITICALITY UNDER STRICT COMPLEMENTARITY

Strict complementarity holds at $\bar{x}$ on (VS) if there exists there $\lambda \in \Lambda(\bar{x})$ such that $\lambda \in \operatorname{ri} N_{\Theta}(\Phi(\bar{x}))$ (Bonnans-Shapiro, 2000)

THEOREM Let $\bar{x}$ be a stationary point for (VS), let Θ be $\mathcal{C}^2$-cone reducible at $\bar{z} = \Phi(\bar{x})$ to a closed convex cone C, and let the strict complementarity condition hold at $\bar{x}$. Then a Lagrange multiplier $\bar{\lambda} \in \Lambda(\bar{x})$ is noncritical if and only if either semi-isolated calmness or error bound condition is satisfied

Here we do not impose the calmness and closedness assumptions of the general characterization theorem for noncriticality

SEQUENTIAL QUADRATIC PROGRAMMING METHOD

Consider the constrained optimization problem

$$\text{minimize } \varphi(x) \text{ subject to } \Phi(x) \in \Theta$$

ALGORITHM (basic SQP method) Choose any $(x_k, \lambda_k) \in \mathbb{R}^n \times \mathbb{R}^m$ and set $k = 0$

- If (x_k, λ_k) satisfies the KKT system, then stop
- Compute (x_{k+1}, λ_{k+1}) as a solution to the KKT system (with $H_k = \nabla^2 L(x_k, \lambda_k)$)

$$\begin{aligned}&\text{min } \varphi(x_k) + \langle \nabla \varphi(x_k), x - x_k \rangle + \tfrac{1}{2}\langle H(x_k)(x - x_k), x - x_k \rangle\\ &\text{subject to } \Phi(x_k) + \nabla \Phi(x_k)(x - x_k) \in \Theta\end{aligned}$$

- Increase k by 1 and then go back to Step 1

SUPERLINEAR CONVERGENCE OF SQP METHOD

THEOREM Assume that

- $\bar{x}$ is a local minimizer and $\Lambda(\bar{x}) = \{\bar{\lambda}\}$
- $\bar{\lambda}$ is noncritical
- The multiplier mapping $M_{\bar{x}}$ is calm at $((0,0), \bar{\lambda})$

Then for any starting point (x_0, λ_0) sufficiently close to $(\bar{x}, \bar{\lambda})$, the SQP method converges to $(\bar{x}, \bar{\lambda})$ and the rate of convergence is superlinear

FULL STABILITY OF LOCAL MINIMIZERS

Following [LevPolRoc00], consider the two-parameter perturbation (2P) of (CO) defined by

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

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 (2P) 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 [M19] Let $\bar{x}$ be a fully stable local optimal solution to (CO) when either $\theta \in CPWL$ or $\theta = \delta_\Gamma$ where Γ is $\mathcal{C}^2$-reducible Then the Lagrange multiplier set $\Lambda_{\rm com}(\bar{x})$ 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; see, e.g., [MorNghiaRoc16] with the references therein. 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 cases

REFERENCES

• J. F. Bonnans and A. Shapiro (2000), Perturbation Analysis of Optimization Problems, Springer

• A. F. Izmailov (2005), On the analytical and numerical stability of critical Lagrange multipliers, Comput. Math. Math. Phys. **45**, 930–946

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

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

- B. S. Mordukhovich (2018), Variational Analysis and Applications, Springer

- B. S. Mordukhovich (2019), Avoiding critical multipliers and slow convergence of primal-dual methods for fully stable minimizers, J. Convex Nonlin. Anal. **20**, 1475–1496

- B. S. Mordukhovich, T. T. A. Nghia and R. T. Rockafellar (2016), Full stability in finite-dimensional optimization, Math. Oper. Res. **40**, 226–252

- B. S. Mordukhovich and M. E. Sarabi (2018), Critical multipliers in variational systems via second-order generalized differentiation, Math. Program. **169**, 605–648

• B. S. Mordukhovich and M. E. Sarabi (2019), Criticality of Lagrange multipliers in variational systems, SIAM J. Optim.**29**, 1524–1557

• R. T. Rockafellar and R. J-B. Wets (1998), Variational Analysis, Springer