Augmented Lagrangians and Hidden Convexity in Sufficient Conditions for Local Optimality

Terry Rockafellar
University of Washington, Seattle

One World Optimization Seminar
12 October 2020

The Role of Sufficient Conditions in Optimization

Classical:
- direct verification of local optimality
- desired to be as close to necessary as possible

Modern:
- design and justification of solution methodologies
- only asked to describe "typical" circumstances

Elementary example with underlying convexity

minimize smooth $f(x)$ for $x \in \boldsymbol{R}^n$, local optimality of $\bar{x}$

first-order: $\nabla f(\bar{x}) = 0$

second-order: $\nabla^2 f(\bar{x})$ positive-definite

$\Longrightarrow$ local strong convexity, reduction to convex optimization

Fundamental question: similar reduction in greater generality?

do sufficienct conditions entail it? or should they?

Hidden Convexity in Classical Nonlinear Programming

Problems with equality constraints: local optimality of $\bar{x}$?

minimize $f_0(x)$ subject to $f_i(x) = 0,\ i = 1, \ldots m$

Lagrangian: $L(x,y) = f_0(x) + y_1 f_1(x) + \cdots + y_m f_m(x)$

first-order: $\nabla_x L(\bar{x},\bar{y}) = 0,\ \ \nabla_y L(\bar{x},\bar{y}) = 0$

second-order: $\nabla^2_{xx} L(\bar{x},\bar{y})$ positive-definite relative to the subspace $S := \left\{ \xi \in R^n \,\middle|\, \nabla f_i(\bar{x}) \cdot \xi = 0,\ i = 1, \ldots, m \right\}$

no local reduction to convex optimization is **evident**, but ...

augmented Lagrangian: $L_r(x,y) = L(x,y) + \frac{r}{2} \sum_{i=1}^{m} f_i(x)^2$

Equivalent statement revealing local convex duality

for $r > 0$ high enough, $L_r(x,y)$ has a convex-concave-type saddle point at $(\bar{x},\bar{y})$, moreover with strong convexity in x

= the secret behind classical "augmented Lagrangian methods"

Framework of "Generalized Nonlinear Programming"

Problem in basic form

$$\text{minimize } f_0(x) + g(F(x)) \text{ for } F(x) = (f_1(x), \ldots, f_m(x))$$

f_i smooth on $\mathbf{R}^n$, but g just closed proper convex on $\mathbf{R}^m$

Some special cases: g = the "modeling function"

- $g(u) = \delta_K$, minimize $f_0(x)$ subject to $F(x) \in K$
- $g(u) = ||u||$ for some norm, "regularization"
- $g(u) = \max\{u_1, \ldots, u_m\}$, min of $f_0(x) + \max\{f_1(x), \ldots, f_m(x)\}$

Mixtures: $F(x) = (F^1(x), \ldots, F^s(x))$ for $F^j : \mathbf{R}^n \to \mathbf{R}^{m_j}$

$g(u) = g_1(u^1) + \cdots g_s(u^s)$ for $u^j \in \mathbf{R}^{m_j}$

$$\text{minimize } f_0(x) + g^1(F^1(x)) + \ldots + g^s(F^s(x))$$

could constrain x to X through $F^s(x) = x$ and $g^s = \delta_X$

Generalized Augmented Lagrangians

Problem reformulation with canonical perturbations

$$\text{minimize} \;\; \varphi(x,u) = f_0(x) + g\big(F(x)+u\big) \;\; \text{subject to } u = 0$$

Lagrangian: with multiplier vectors $y \in \boldsymbol{R}^m$

$$l(x,y) = \inf_u \left\{ \varphi(x,u) - y{\cdot}u \right\} = L(x,y) - g^*(y)$$

Augmented Lagrangian: with augmentation parameter $r > 0$

$$l_r(x,y) = \inf_u \left\{ \varphi(x,u) - y{\cdot}u + \tfrac{r}{2}|u|^2 \right\} = \\ L(x,y) - g_r^*(y + rF(x)) \;\text{ or }\; f_0(x) + g^r\big(\tfrac{1}{r}y + F(x)\big) - \tfrac{1}{2r}|y|^2$$

$$g^r(u) = \min_{u'} \left\{ g(u') + \tfrac{r}{2}|u'-u|^2 \right\} \text{ with } g^{r\,*}(y) = g^*(y) + \tfrac{1}{2r}|y|^2$$
$$g_r(u) = g(u) + \tfrac{r}{2}|u|^2 \text{ with } g_r^*(y) = \min_{y'} \left\{ g^*(y') + \tfrac{1}{2r}|y'-y|^2 \right\}$$

$$f_i\text{'s} \in \mathcal{C}^1 \implies l_r \in \mathcal{C}^1, \qquad f_i\text{'s} \in \mathcal{C}^2 \implies l_r \in \mathcal{C}^{1+}$$

Cone case: $g = \delta_K$, $g^* = \delta_Y$, for polar cones K and Y

$$\implies \quad g^r(u) = \tfrac{r}{2}\mathrm{dist}_K^2(u), \qquad g_r^*(y) = \tfrac{1}{2r}\mathrm{dist}_Y^2(y)$$

First-Order Optimality, Generalized KKT Conditions

Properties of the objective: $\varphi(x,u) = f_0(x) + g(F(x)+u)$

lsc, proper, subdifferentially continuous/regular

the subgradients $(v,y) \in \partial\varphi(x,u)$ are "regular" subgradients:

$$\varphi(x',u') \geq \varphi(x,u) + (v,y)\cdot[(x',u') - (x,u)] + o\big(|(x.u') - (x,u)|\big)$$

First-order condition on $\bar{x}$ and a multiplier vector $\bar{y}$

$(0,\bar{y}) \in \partial\varphi(\bar{x},0)$ (necessary under a constraint qual.)

Equivalent expression with augmented Lagrangians (any $r > 0$)

$$\nabla_x l_r(\bar{x},\bar{y}) = 0, \quad \nabla_y l_r(\bar{x},\bar{y}) = 0 \ \left[\text{i.e., } \bar{y} \in \operatorname{argmax} l_r(\bar{x},\cdot)\right]$$

this is the first-order condition for a **saddle point** at $(\bar{x},\bar{y})$

the second part = the multiplier condition $\bar{y} \in \partial g(F(\bar{x}))$

Convex optimization: the case where $l_r(x,y)$ is convex in x

but what about having just **local convexity** in x around $\bar{x}$?

Variational Convexity of a Function

Let $f : R^n \to (-\infty, \infty]$ be lsc, proper, subdiff.contin./regular

$$v \in \partial f(x) \quad \Longleftrightarrow \quad f(x') \geq f(x) + v{\cdot}(x' - x) + o(|x' - x|)$$

Definition: f is **variationally convex** at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ if
$\exists$ open convex neighborhood $\mathcal{X} \times \mathcal{V}$ of $(\bar{x}, \bar{v})$ for which
there exists a proper lsc convex function $h \leq f$ on $\mathcal{X}$ such that

$$[\mathcal{X} \times \mathcal{V}] \cap \operatorname{gph} \partial h = [\mathcal{X} \times \mathcal{V}] \cap \operatorname{gph} \partial f$$

and, for (x, v) belonging to this common set, also $h(x) = f(x)$

variational **strong** convexity has h **strongly** convex

Theorem — Rock. 2019

this holds $\Longleftrightarrow$ ∂f is max monotone locally around $(\bar{x}, \bar{v})$

Observation: f variationally convex at $\bar{x}$ for $\bar{v} = 0 \in \partial f(\bar{x})$
$\Longrightarrow$ f has a local minimum at $\bar{x}$

variational strong convexity $\Longleftrightarrow$ tilt stability of this min

Illuminating Examples of Variational Convexity

Recall the property: on neighborhood $\mathcal{X} \times \mathcal{V}$ of $(\bar{x}, \bar{v}) \in \operatorname{gph} \partial f$
there exists a proper lsc convex function $h \leq f$ on $\mathcal{X}$ such that
$$[\mathcal{X} \times \mathcal{V}] \cap \operatorname{gph} \partial h = [\mathcal{X} \times \mathcal{V}] \cap \operatorname{gph} \partial f$$
and, for (x, v) belonging to this common set, also $h(x) = f(x)$

Relation to local convexity? this results when $\mathcal{V}$ = whole space

---

Example in one dimension: $f(x) = |x| - x^2$ for $x \in \boldsymbol{R}$

$\operatorname{gph} \partial f$ reduces around $(\bar{x}, \bar{v}) = (0, 0)$ to a *vertical line segment*

this is an instance of local maximal monotonicity

Example in two dimensions: $x = (x_1, x_2) \in \boldsymbol{R}^2$

$f(x_1, x_2) = x_2 - x_2^2$ if $x_2 = x_1^2$, but $= \infty$ elsewhere

- strong variational convexity for $\bar{x} = (0, 0)$, $\bar{v} = (0, 0)$, despite nonconvexity of $\operatorname{dom} f$, through tilt stablity of min at $(0, 0)$
- this can be seen as a case of nonlinear programmming with a solution satisfying the strong second-order sufficient condition

Second-Order Optimality Via Variational Sufficiency

Recall problem: min $\varphi(x,u) = f_0(x) + g(F(x)+u)$ s.t. $u = 0$
"unaffected" if $\varphi(x,u)$ is replaced by $\varphi_r(x,u) = \varphi(x,u) + \frac{r}{2}|u|^2$

Variationally sufficient condition for local optimality (Rock. 2019)

Combine first-order condition $(0,\bar{y}) \in \partial\varphi(\bar{x},0) = \partial\varphi_r(\bar{x},0)$ with
$\exists r > 0$ such that φ_r is variationally convex at $(\bar{x},0)$ for $(0,\bar{y})$
strong version: variational strong convexity

Connection of the strong version with "quadratic growth":
$\iff \exists\, s > 0$, convex nbhds $\mathcal{W}$ of $(\bar{x},0)$ and $\mathcal{Z}$ of $(0,\bar{y})$, with

$$\varphi_r(x',u') \geq \varphi_r(x,u) + (v,y)\cdot[(x',u')-(x,u)] + \frac{s}{2}|(x',u')-(x,u)|^2$$

for $(x',u') \in \mathcal{W}$ when $(v',y') \in \partial\varphi_r(x,u) \cap \mathcal{Z}$

Example: case of $\varphi \in \mathcal{C}^2$

strong var. sufficiency $\iff \begin{cases} \nabla\varphi(\bar{x},0) \perp S := \{(x,u) \mid u = 0\} \\ \nabla^2\varphi(\bar{x},0) \text{ pos.-def. relative to } S \end{cases}$

Saddle Characterization of Variational Sufficiency

assume just that the functions $f_i(x)$ in the problem are $\mathcal{C}^1$

Theorem 1

The variationally sufficient condition for local optimality holds for $\bar{x}$ with multiplier $\bar{y}$ and parameter $r > 0$ $\iff$

$\exists$ convex neighborhood $\mathcal{X} \times \mathcal{Y}$ of $(\bar{x}, \bar{y})$ such that

- $l_r(x, y)$ is convex in $x \in \mathcal{X}$ (as well as concave in $y \in \mathcal{Y}$)
- $(\bar{x}, \bar{y})$ is a saddle point of $l_r(x, y)$ relative to $\mathcal{X} \times \mathcal{Y}$

Local duality: locally articulated primal and dual problems

$(P^r_{\mathcal{X}\times\mathcal{Y}})$ $\min_{x\in\mathcal{X}}$ the convex function $x \to \max_{y\in\mathcal{Y}} l_r(x, y)$

$(D^r_{\mathcal{X}\times\mathcal{Y}})$ $\max_{y\in\mathcal{Y}}$ the concave function $y \to \min_{x\in\mathcal{X}} l_r(x, y)$

Significance of the saddle point condition:

$\bar{x}$ solves $(P^r_{\mathcal{X}\times\mathcal{Y}})$, $\bar{y}$ solves $(D^r_{\mathcal{X}\times\mathcal{Y}})$, $\min(P^r_{\mathcal{X}\times\mathcal{Y}}) = \max(D^r_{\mathcal{X}\times\mathcal{Y}})$

Saddle Characterization of Strong Variational Sufficiency

Theorem 2

The extra property contributes to the saddle point condition by the **strong** version of variational sufficiency is that

$l_r(x, y)$ is **strongly** convex with respect to $x \in X$

Equivalently, this corresponds to having "augmented tilt stability"

Augmented tilt stability:

the mapping $(v, y) \mapsto \operatorname{argmin}_{x \in \mathcal{X}} \{ l_r(x, y) - v \cdot x \}$ is single-valued Lipschitz continuous for (v, y) near $(0, \bar{y})$

[modulus of Lipschitz continuity]=1/[modulus of strong convexity]

Implication for numerical optimization:

strong variational sufficiency is key to "methods of multipliers"

Connecting With Previous Second-Order Sufficiency

strong variational sufficiency versus other conditions?

assume now that the functions $f_i(x)$ are all $\mathcal{C}^2$

Revealing example: **classical nonlinear programming**

$$\text{minimize } f_0(x) \text{ subject to } f_i(x) \begin{cases} \leq 0 \text{ for } i = 1, \ldots, s, \\ = 0 \text{ for } i = s+1, \ldots, m \end{cases}$$

(here $g = \delta_K$ for the cone K polar to $Y = \boldsymbol{R}^s_+ \times \boldsymbol{R}^{m-s}$)

classical Lagrangian: $L(x, y) = f_0(x) + y_1 f_1(x) + \cdots + y_m f_m(x)$

Strong variational sufficiency in this case

Equivalent to the **"strong second-order sufficient condition"**:

$\nabla^2_{xx} L(\bar{x}, \bar{y})$ is pos.-definite relative to the subspace
$S(\bar{x}, \bar{y}) = \{\xi \in \boldsymbol{R}^n \mid \nabla f_i(\bar{x}) \cdot \xi = 0 \text{ for } i \in I(\bar{x}, \bar{y})\}$,

where $I(\bar{x}, \bar{y}) = \{i \in [1, s] \text{ with } \bar{y}_i > 0\} \cup \{i \in [s+1, m]\}$

interpretation: strong variational sufficiency is its natural extension

Characterizing With Generalized Second-Derivatives

Starting fact: if the augmented Lagrangian $l_r(x,y)$ is $\mathcal{C}^2$, then
strong variational sufficiency $\iff$ $\nabla^2_{xx} l_r(\bar{x},\bar{y})$ pos.-definite

but in general, f_i's $\in \mathcal{C}^2$ only have $l_r \in \mathcal{C}^{1+}$
nevertheless $l_r \in \mathcal{C}^{1+}$ makes $\nabla^2 l_r(x,y)$ exist for almost all (x,y)

Hessian bundle:

$$\overline{\nabla}^2 l_r(\bar{x},\bar{y}) = \left\{ H = \lim_{k\to\infty} \nabla^2 l_r(x^k,y^k) \text{ for } (x^k,y^k) \to (\bar{x},\bar{y}) \right\}$$

any such H can be partitioned into H_{xx}, H_{xy}, H_{yx}, H_{yy}

Theorem 3

strong variational sufficiency holds $\iff$
every $H \in \overline{\nabla}^2 l_r(\bar{x},\bar{y})$ has H_{xx} pos.-definite

Task: now translate this to the modeling functions g through

$$l_r(x,y) = f_0(x) + g^r(\tfrac{1}{r}y + F(x)) - \tfrac{1}{2r}|y|^2$$

Characterizing With Generalized Quadratic Forms

Quadratic forms: $q(\omega) = \frac{1}{2}\omega{\cdot}Q\omega$ with Q pos.-semidefinite

generalized: $q(\omega) = \frac{1}{2}\omega{\cdot}Q\omega + \delta_S(u)$ for a subspace S

Generalized twice differentiability: of g at u for $y \in \partial g(u)$

the second-order difference quotient function

$$\Delta_t^2 g(u \,|\, y)(\omega) = [g(u+t\omega) - g(u) - t\, y{\cdot}\omega]/\tfrac{1}{2}t^2 \text{ for } t > 0$$

epi-converges as $t \to 0$ to a generalized quadratic form q

this is true for almost all $(u, y) \in \operatorname{gph} \partial g$

Quadratic bundle: of g at $\bar{u}$ for $\bar{y} \in \partial g(\bar{u})$

consider pairs $(u^k, y^k) \in \operatorname{gph} \partial g$ yielding generalized forms q^k

$$\operatorname{quad} g(\bar{u} \,|\, \bar{y}) := \left\{ q = \text{epi-}\lim q^k \text{ as } (u^k, y^k) \to (\bar{u}, \bar{y}) \right\}$$

Theorem 4, building on the multiplier condition $\bar{y} \in \partial g(F(\bar{x}))$

strong variational sufficiency holds $\iff$

$$q \in \operatorname{quad} g(F(\bar{x}) \,|\, \bar{y}) \implies q(\omega) + \tfrac{1}{2}\omega{\cdot}\nabla_{xx}^2 L(\bar{x}, \bar{y})\omega > 0 \text{ if } \omega \neq 0$$

a bit stronger than such-type conditions known in special cases

What More?

Elaborating for particular modeling functions:

instances of g associated with various useful features

in minimizing $f_0(x) + g\big(f_1(x), \ldots, f_m(x)\big)$

Applying to the method of multipliers, ALM:

$$x^{k+1} \approx \mathrm{argmin}_x\, l_{r_k}(x, y^k), \qquad y^{k+1} = y^k + r_k \nabla_y l_{r_k}(x^{k+1}, y^k)$$

- executes the proximal point algorithm on a dual problem
- variational sufficiency yields localization in nonconvex problems

Extending to "progressive decoupling" methodology:

similar explorations in recent schemes of problem decomposition

Speculating about the bigger picture:

- the idea that convexity can elicited through **primal+dual** localization seems to have far-reaching potential
- this deserves investigation in other areas, like optimal control

References

[1] R.T. Rockafellar (2019) "Variational convexity and local monotonicity of subgradient mappings," *Vietnam Journal of Mathematics* (issue in honor of A.D. Ioffe's 80th birthday)
dx.doi.org/10.1007/s10013-019-00339-S

[2] R.T. Rockafellar (2019) "Progressive decoupling of linkages in optimization and variational inequalities with elicitable convexity or monotonicity," *Set-valued and Variational Analysis*
dx.doi.org/10.1007/s1128-018-0496-1

[3] R.T. Rockafellar (2020) "Augmented Lagrangians and hidden convexity in sufficient conditions for local optimality," *Mathematical Programming*, recently submitted

downloads: sites.washington.edu/∼rtr/mypage.html