Normality and Quasinormality in Nonlinear Programming and Optimal Control

JAVIER F ROSENBLUETH
IIMAS, Universidad Nacional Autónma de México
Apartado Postal 20-726, CDMX 01000
MEXICO
jfrl@unam.mx

Abstract: In nonlinear programming, the notion of quasinormality provides a constraint qualification which is in general weaker than that of normality, and it emerges naturally from an extended Lagrange multiplier rule. In this paper, we explain in detail its origin and some of its consequences in optimization theory for finite dimensional problems involving equality and inequality constraints, and provide a possible generalization to optimal control problems.

Key–Words: Lagrange multipliers, equality and inequality constraints, quasinormality, normality, optimal control

1 The finite dimensional case

Let us begin with a brief explanation of the main aspects of the theory related to the notions of normality and quasinormality in a nonlinear programming context.

Consider the problem, which we shall label (N), of minimizing f on the set S, where $f, g_i\colon \mathbb{R}^n \to \mathbb{R}$ ($i \in A \cup B$) are continuously differentiable functions, $A = \{1, \ldots, p\}$, $B = \{p+1, \ldots, m\}$, and

$$S = \{x \in \mathbb{R}^n \mid g_\alpha(x) \le 0 \ (\alpha \in A),$$

$$g_\beta(x) = 0 \ (\beta \in B)\}.$$

It is well-known that, if x_0 is a local solution to the problem (N) and an additional assumption is satisfied for the functions defining the constraints, then there exists a multiplier $\lambda = (\lambda_1, \ldots, \lambda_m)$ such that, with respect to x_0, λ, f and $g_1, \ldots, g_m$, the Karush-Kuhn-Tucker (KKT) conditions (or *first order Lagrange multiplier rule*) hold. Those conditions are given by

i. $\lambda_\alpha \ge 0$ ($\alpha \in A$).
ii. $\lambda_\alpha g_\alpha(x_0) = 0$ ($\alpha \in A$).
iii. If $F(x) := f(x) + \langle \lambda, g(x) \rangle$ then $F'(x_0) = 0$.

Here, g denotes the function mapping $\mathbb{R}^n$ to $\mathbb{R}^m$ whose components are $g_1, \ldots, g_m$, and $\langle \cdot, \cdot \rangle$ refers to the standard inner product in $\mathbb{R}^m$ so that $\langle \lambda, g(x) \rangle = \sum_1^m \lambda_i g_i(x)$. In this system, the function F corresponds to the standard *Lagrangian*, and the real numbers $\lambda_1, \ldots, \lambda_m$ are called the *Kuhn-Tucker* or *Lagrange multipliers*.

Given $x_0 \in S$, let us denote by $\Lambda(x_0)$ the set of all $\lambda \in \mathbb{R}^m$ satisfying the KKT conditions.

In general, if x_0 solves (N) locally, the KKT conditions may not hold unless further assumptions on g, known as *constraint qualifications*, are imposed. In other words, a constraint qualification should be imposed in order to guarantee that $\Lambda(x_0) \neq \emptyset$.

A simple, natural constraint qualification is easily obtained by means of an extended necessary condition involving a cost multiplier. It is the well-known Fritz John necessary optimality condition, which can be stated as follows (see, for example, [2, 6–8, 10] and, for a nonsmooth multiplier rule, see [3, p 221]).

Theorem 1 *If x_0 solves* (N) *locally, then there exist $\lambda_0 \ge 0$ and $\lambda \in \mathbb{R}^m$, not both zero, such that*

i. $\lambda_\alpha \ge 0$ ($\alpha \in A$).
ii. $\lambda_\alpha g_\alpha(x_0) = 0$ ($\alpha \in A$).
iii. *If* $F_0(x) := \lambda_0 f(x) + \langle \lambda, g(x) \rangle$, $F_0'(x_0) = 0$.

The KKT conditions correspond to the case $\lambda_0 = 1$, and constraint qualifications are conditions for which Theorem 1 holds precisely with $\lambda_0 = 1$. Let us define *normality* as a condition which, based on Theorem 1, implies that $\Lambda(x_0) \neq \emptyset$.

Definition 2 *A point $x_0 \in S$ is said to be normal relative to S if $\lambda = 0$ is the only solution to*

i. $\lambda_\alpha \ge 0$ ($\alpha \in A$).
ii. $\lambda_\alpha g_\alpha(x_0) = 0$ ($\alpha \in A$).
iii. $\sum_1^m \lambda_i g_i'(x_0) = 0$.

Observe that, in Theorem 1, if x_0 is also a normal point relative to S, then $\lambda_0 > 0$ (since, by normality relative to S, if $\lambda_0 = 0$ then $\lambda = 0$, contradicting the nontriviality condition) in which case the set of multipliers obtained by dividing each multiplier by λ_0 is a new set having $\lambda_0 = 1$. This implies nonemptiness of $\Lambda(x_0)$. For the question of uniqueness of the multipliers, we refer to [9, 11] for the nonlinear problem and to [1, 4] for multipliers in an optimal control context.

Now, an extended Lagrange multiplier rule which yields the notion of quasinormality and is more general than the Fritz John necessary optimality condition, can be found [8]. The result is strongly related to a proof of Theorem 1 given in [10], which we shall summarize in three simple steps.

Denote by $I(x_0) = \{\alpha \in A \mid g_\alpha(x_0) = 0\}$ the set of *active indices* at $x_0 \in S$.

Proof of Theorem 1.

Without loss of generality, assume that $x_0 = 0$ and $f(x_0) = 0$. Let

$$B(\epsilon) := \{x \in \mathbb{R}^n : |x| \le \epsilon\}$$

and choose $\delta > 0$ such that

$$g_\alpha(x) < 0 \text{ for all } \alpha \notin I(x_0) \text{ and } x \in B(\delta).$$

Define

$$G(x) = \sum_{\alpha \in I(x_0)} g_\alpha^+(x)^2 + \sum_{\beta \in B} g_\beta(x)^2$$

where $f^+(x) = \max(f(x), 0)$.

Step 1. For all $0 < \epsilon \le \delta$ there exists $\sigma > 0$ such that, for all x with $|x| = \epsilon$,

$$f(x) + |x|^2 + \sigma G(x) > 0.$$

Proof: Suppose not. Let $\epsilon \in (0, \delta]$ be such that, for all $q \in \mathbf{N}$, $\exists x_q$ with $|x_q| = \epsilon$ such that

$$f(x_q) + |x_q|^2 \le -qG(x_q). \tag{1}$$

Since $\{x_q\}$ is bounded, there exist a subsequence (we do not relabel) and x^* such that $x_q \to x^*$. Then

$$|x^*| = \lim |x_q| = \epsilon$$

and $f(x_q) \to f(x^*)$. By dividing both members of (1) by $-q$ and letting $q \to \infty$ we conclude that $G(x^*) = 0$ and so $x^* \in S$. Thus

$$\lim f(x_q) = f(x^*) \ge f(x_0) = 0.$$

But by (1), $f(x_q) \le -\epsilon^2$, a contradiction.

Step 2. For all $0 < \epsilon \le \delta$ there exists $\bar{x}$ with $|\bar{x}| < \epsilon$, and a unit vector $(\lambda_0, \ldots, \lambda_m)$ with $\lambda_0 \ge 0$, $\lambda_\alpha \ge 0$ $(\alpha \in I(x_0))$, and $\lambda_\beta \in \mathbb{R}$ $(\beta \in B)$ such that

$$\lambda_0[f'(\bar{x}) + 2\bar{x}] + \sum_{\alpha \in I(x_0)} \lambda_\alpha g'_\alpha(\bar{x}) + \sum_{\beta \in B} \lambda_\beta g'_\beta(\bar{x}) = 0. \tag{2}$$

Proof: Let $0 < \epsilon \le \delta$, let σ be as in Step 1, and define

$$H(x) = f(x) + |x|^2 + \sigma G(x).$$

Let $\bar{x}$ minimize H on $B(\epsilon)$. Then

$$H(\bar{x}) \le H(0) = 0$$

and so, in view of Step 1, we cannot have $|\bar{x}| = \epsilon$. Thus $\bar{x}$ is interior to $B(\epsilon)$ and $H'(\bar{x}) = 0$. Thus

$$f'(\bar{x}) + 2\bar{x} + \sum_{\alpha \in I(x_0)} 2\sigma g_\alpha^+(\bar{x}) g'_\alpha(\bar{x}) + \sum_{\beta \in B} 2\sigma g_\beta(\bar{x}) g'_\beta(\bar{x}) = 0. \tag{3}$$

Define $L = \{1 + 4\sigma^2 G(\bar{x})\}^{1/2}$,

$$\lambda_0 = 1/L, \quad \lambda_\alpha = 2\sigma g_\alpha^+(\bar{x})/L \ \ (\alpha \in I(x_0)),$$

$$\lambda_\beta = 2\sigma g_\beta(\bar{x})/L \ \ (\beta \in B)$$

and $\lambda_\alpha = 0$ if $\alpha \notin I(x_0)$. Then $(\lambda_0, \ldots, \lambda_m)$ is a unit vector, $\lambda_0 \ge 0$, $\lambda_\alpha \ge 0$ $(\alpha \in I(x_0))$ and, if we divide both members of (3) by L, we obtain (2).

Step 3. For all $q \in \mathbf{N}$ choose $\bar{x}_q$ with $|\bar{x}_q| < \delta/q$ and unit vectors $(\lambda_{0q}, \ldots, \lambda_{mq})$ with nonnegative λ_{0q} and $\lambda_{\alpha q}$ $(\alpha \in I(x_0))$ such that (2) holds. Choose a subsequence for which the unit vectors converge to a limit $(\lambda_0, \ldots, \lambda_m)$.

Since $\bar{x}_q \to x_0$, the equation in (2) holds with this limit-vector and with x_0 in place of $\bar{x}$. □

As mentioned before, an extended Lagrange multiplier rule (due to Hestenes) derived in [8], is strongly based on this particular proof (due to McShane) given in [10] of the Fritz John necessary condition. The extended rule can be stated as follows.

Theorem 3 *If x_0 solves* (N) *locally, then there exist $\lambda_0 \ge 0$ and $\lambda \in \mathbb{R}^m$, not both zero, such that*

i. $\lambda_\alpha \ge 0$ $(\alpha \in A)$.

ii. *For all $\epsilon > 0$ $\exists x$ with $|x - x_0| < \epsilon$ such that $\lambda_\gamma g_\gamma(x) > 0$ whenever $\lambda_\gamma \ne 0$.*

iii. *If $F_0(x) := \lambda_0 f(x) + \langle \lambda, g(x) \rangle$, $F'_0(x_0) = 0$.*

Note that Theorem 1 is a simple consequence of Theorem 3. The notion of quasinormality, introduced in [8], is based on this extended multiplier rule.

Definition 4 *A point $x_0 \in S$ is said to be quasinormal relative to S if $\lambda = 0$ is the only solution to*

i. $\lambda_\alpha \ge 0$ $(\alpha \in A)$.

ii. *For all $\epsilon > 0$ $\exists x$ with $|x - x_0| < \epsilon$ such that $\lambda_\gamma g_\gamma(x) > 0$ whenever $\lambda_\gamma \ne 0$.*

iii. $\sum_1^m \lambda_i g'_i(x_0) = 0$.

Clearly, normality implies quasinormality and, if x_0 is a quasinormal point of S in Theorem 3, then $\lambda_0 > 0$ and the multipliers can be chosen so that $\lambda_0 = 1$. In other words, quasinormality is also a constraint qualification.

Now, the importance of these notions relies not only on the fact that they correspond to conditions which guarantee nonemptiness of $\Lambda(x_0)$, but also because they imply *regularity* of the point under consideration, in the sense that the set of tangential constraints at a point x_0, that is,

$$R_S(x_0) = \{h \in \mathbb{R}^n \mid g'_\alpha(x_0; h) \leq 0 \ (\alpha \in I(x_0)), \\ g'_\beta(x_0; h) = 0 \ (\beta \in B)\}$$

coincides with the tangent cone of S at x_0. Thus, if first and second order conditions in terms of the tangent cone are obtained, a criterion in terms of tangential constraints can be derived through regularity and, therefore, through the notions of normality and quasinormality.

2 Optimal control

Let us turn now to a fixed endpoint Lagrange optimal control problem involving equality and inequality constraints in the control functions.

We shall be interested in minimizing the functional

$$I(x, u) := \int_{t_0}^{t_1} L(t, x(t), u(t))dt$$

over all piecewise smooth trajectories x and piecewise continuous controls u, satisfying

$$\begin{cases} \dot{x}(t) = f(t, x(t), u(t)) \ (t \in T), \\ x(t_0) = \xi_0, \ x(t_1) = \xi_1, \\ u(t) \in U \ (t \in T), \end{cases}$$

where the control set is given by

$$U := \{u \in \mathbb{R}^m \mid \varphi_\alpha(u) \leq 0 \ (\alpha \in R), \\ \varphi_\beta(u) = 0 \ (\beta \in Q)\}$$

with $R = \{1, \ldots, r\}$ and $Q = \{r + 1, \ldots, q\}$.

The data for this problem, which we label (P), correspond to an interval $T := [t_0, t_1]$ in $\mathbb{R}$, two points ξ_0, ξ_1 in $\mathbb{R}^n$, and functions L and f mapping $T \times \mathbb{R}^n \times \mathbb{R}^m$ to $\mathbb{R}$ and $\mathbb{R}^n$ respectively, and $\varphi = (\varphi_1, \ldots, \varphi_q)$ mapping $\mathbb{R}^m$ to $\mathbb{R}^q$.

A *control function* u is a piecewise continuous function mapping T to $\mathbb{R}^m$, and a *state trajectory* x corresponding to a control function u is a piecewise $C^1(T, \mathbb{R}^n)$ solution to the differential equation

$$\dot{x}(t) = f(t, x(t), u(t)) \quad (t \in T).$$

A *process* is a pair (x, u) comprising a state trajectory x and an associated control function u and, if u satisfies the control constraints $u(t) \in U$ $(t \in T)$, and x satisfies the endpoint constraints $x(t_0) = \xi_0$ and $x(t_1) = \xi_1$, the process (x, u) is said to be *admissible*. An admissible process *solves* (P) if it achieves the minimum value of I over all admissible processes.

Denote by X the space of piecewise C^1 functions mapping T to $\mathbb{R}^n$, and by $\mathcal{U}_k$ the space of piecewise continuous functions mapping T to $\mathbb{R}^k$ $(k \in \mathbf{N})$. Let $Z := X \times \mathcal{U}_m$, and consider the following two sets:

$$D := \{(x, u) \in Z \mid \dot{x}(t) = f(t, x(t), u(t)) \ (t \in T), \\ x(t_0) = \xi_0, \ x(t_1) = \xi_1\},$$

$$S := \{(x, u) \in D \mid \varphi_\alpha(u(t)) \leq 0, \ \varphi_\beta(u(t)) = 0 \\ (\alpha \in R, \ \beta \in Q, \ t \in T)\}.$$

Our problem (P) is thus that of minimizing I over S.

With respect to the smoothness of the functions delimiting the problem, we assume that L, f and φ are C^2 and the $q \times (m + r)$-dimensional matrix

$$\left(\frac{\partial \varphi_i}{\partial u^k} \ \ \delta_{i\alpha}\varphi_\alpha \right)$$

$(i = 1, \ldots, q; \ \alpha = 1, \ldots, r; \ k = 1, \ldots, m)$ has rank q on U (here $\delta_{\alpha\alpha} = 1$, $\delta_{\alpha\beta} = 0$ $(\alpha \neq \beta)$). This condition is equivalent to the condition that, at each point u in U, the matrix

$$\left(\frac{\partial \varphi_i}{\partial u^k} \right) \qquad (i = i_1, \ldots, i_p; \ k = 1, \ldots, m)$$

has rank p, where $i_1, \ldots, i_p$ are the indices $i \in \{1, \ldots, q\}$ such that $\varphi_i(u) = 0$ (see [5] for details).

Given $(x, u) \in Z$ we shall use the notation $(\tilde{x}(t))$ to represent $(t, x(t), u(t))$, for a specific $(x_0, u_0) \in Z$, we set $A(t) := f_x(\tilde{x}_0(t))$ and $B(t) := f_u(\tilde{x}_0(t))$, and '$*$' denotes transpose.

Define the Hamiltonian function H mapping $T \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^q \times \mathbb{R}$ to $\mathbb{R}$ as

$$H(t, x, u, p, \mu, \lambda) :=$$

$$\langle p, f(t, x, u)\rangle - \lambda L(t, x, u) - \langle \mu, \varphi(u)\rangle.$$

First order necessary conditions, following the maximum principle, can be stated as follows (see, for example, [3, 7]). This result can be seen as the analogous of Theorem 1, the Fritz John necessary optimality condition.

Theorem 5 *If* (x_0, u_0) *solves* (P)*, then there exist* $\lambda_0 \geq 0$*,* $p \in X$*, and* $\mu \in \mathcal{U}_q$*, not vanishing simultaneously on* T*, such that*

a. $\mu_\alpha(t) \geq 0$ $(\alpha \in R, \ t \in T)$*;*

b. $\mu_\alpha(t)\varphi_\alpha(u_0(t)) = 0$ $(\alpha \in R,\ t \in T)$*;*
c. $\dot{p}(t) = -H_x^*(\tilde{x}_0(t), p(t), \mu(t), \lambda_0)$ *on every interval of continuity of* u_0.
d. $H_u(\tilde{x}_0(t), p(t), \mu(t), \lambda_0) = 0$ $(t \in T)$.

Define the *Lagrange multipliers* for this problem as those couples (p, μ) for which the above conditions hold with $\lambda_0 = 1$.

Definition 6 *For* $(x_0, u_0) \in S$*, denote by* $\Lambda(x_0, u_0)$ *the set of all* $(p, \mu) \in X \times \mathcal{U}_q$ *satisfying*
a. $\mu_\alpha(t) \geq 0$ $(\alpha \in R,\ t \in T)$*;*
b. $\mu_\alpha(t)\varphi_\alpha(u_0(t)) = 0$ $(\alpha \in R,\ t \in T)$*;*
c. $\dot{p}(t) = -A^*(t)p(t) + L_x^*(\tilde{x}_0(t))$ $(t \in T)$*;*
d. $B^*(t)p(t) = L_u^*(\tilde{x}_0(t)) + \varphi'^*(u_0(t))\mu(t)$ $(t \in T)$.

As in the nonlinear programming problem, a constraint qualification should be imposed on a solution (x_0, u_0) to the problem in order to guarantee nonemptiness of $\Lambda(x_0, u_0)$. Let us define normality relative to S by imposing the null solution as the unique solution to the system given in Theorem 5 when the cost multiplier vanishes.

Definition 7 *We shall say that* $(x_0, u_0) \in S$ *is normal relative to S if, given* $(p, \mu) \in X \times \mathcal{U}_q$ *satisfying*
i. $\mu_\alpha(t) \geq 0$ $(\alpha \in R,\ t \in T)$*;*
ii. $\mu_\alpha(t)\varphi_\alpha(u_0(t)) = 0$ $(\alpha \in R,\ t \in T)$*;*
iii. $\dot{p}(t) = -A^*(t)p(t)$ $(t \in T)$*;*
iv. $0 = B^*(t)p(t) - \varphi'^*(u_0(t))\mu(t)$ $(t \in T)$,
then $p \equiv 0$*. Note that, in this event, also* $\mu \equiv 0$.

Observe that, in Theorem 5, if an admissible process (x_0, u_0) solves (P) and is also a normal process relative to S, then $\Lambda(x_0, u_0)$ is nonempty.

Now, a generalization of the notion of quasinormality introduced for the finite dimensional case can be made by considering an extended maximum principle.

Definition 8 *We shall say that* $(x_0, u_0) \in S$ *is quasi-normal relative to S if, given* $(p, \mu) \in X \times \mathcal{U}_q$ *satisfying*
i. $\mu_\alpha(t) \geq 0$ $(\alpha \in R,\ t \in T)$*;*
ii. *For all* $\epsilon > 0$ $\exists u \in \mathcal{U}_m$ *with* $|u(t) - u_0(t)| < \epsilon$ *such that* $\mu_\gamma(t)\varphi_\gamma(u(t))) > 0$ *whenever* $\mu_\gamma(t) \neq 0$.
iii. $\dot{p}(t) = -A^*(t)p(t)$ $(t \in T)$*;*
iv. $0 = B^*(t)p(t) - \varphi'^*(u_0(t))\mu(t)$ $(t \in T)$,
then $p \equiv 0$.

This new notion, which is implied by (but is weaker than) that of normality, can be used to obtain second order necessary conditions which are more refined that those found in the literature. In particular, the main result on second order conditions derived in [5] (see Theorem 6.7), can be shown to hold if the assumption of normality is replaced with that of quasinormality.

References:

[1] J. A. Becerril, K. L. Cortez and J. F. Rosenblueth, Uniqueness of multipliers in optimal control: the missing piece, *IMA J. Math. Control Inform.* 2018, doi.org/10.1093/imamci/dny033

[2] A. Ben-Tal, Second-order and related extremality conditions in nonlinear programming, *J. Optim. Theory Appl.* 31, 1980, pp. 143–165.

[3] F. Clarke, *Functional Analysis, Calculus of Variations and Optimal Control*, Springer-Verlag, London 2013

[4] K. L. Cortez and J. F. Rosenblueth, Normality and uniqueness of Lagrange multipliers, *Discrete Contin. Dyn. Syst.* 38, 2018, pp. 3169–3188.

[5] M. R. de Pinho and J. F. Rosenblueth, Mixed constraints in optimal control: an implicit function theorem approach, *IMA J. Math. Control Inform.* 24, 2017, pp. 197–218.

[6] G. Giorgi, A. Guerraggio and J. Thierfelder, *Mathematics of Optimization: Smooth and Nonsmooth Case*, Elsevier, Amsterdam 2004

[7] M. R. Hestenes, *Calculus of Variations and Optimal Control Theory*, John Wiley, New York 1966

[8] M. R. Hestenes, *Optimization Theory, The Finite Dimensional Case*, John Wiley, New York 1975

[9] J. Kyparisis, On uniqueness of Kuhn-Tucker multipliers in nonlinear programming, *Math. Program.* 32, 1985, pp. 242–246.

[10] E. J. McShane, The Lagrange Multiplier Rule, *Amer. Math. Monthly,* 80, 1973, pp. 922–925.

[11] G. Wachsmuth, On LICQ and the uniqueness of Lagrange multipliers, *Oper. Res. Lett.* 41, 2013, pp. 78–80.