A. F. Izmailov
Moscow State University, Faculty of Computational Mathematics and Cybernetics, Moscow, Russia

M. V. Solodov
Instituto de Matemática Pura e Aplicada, Rio de Janeiro, Brazil

A survey on dual behavior of Newton-type methods for constrained optimization[*]

Abstract

For constrained optimization problems with nonunique Lagrange multiplier associated to a solution, we consider a certain thin subclass of multipliers called critical (and in particular, violating the second-order sufficient condition for optimality) that exhibits some very special properties. Specifically, convergence to a critical multiplier appears to be a typical scenario of dual behaviour of primal-dual Newton-type methods when critical multipliers do exist. Moreover, along with the possible absence of dual convergence, attraction to critical multipliers is precisely the reason for slow primal convergence usually observed on degenerate problems. On the other hand, critical multipliers turn out to have some special analytical stability properties: noncritical multipliers should not be expected to be stable subject to parametric perturbations of optimality systems.

Key words: constrained optimization, degenerate constraints, second-order sufficiency, Newton method, SQP.

1. Introduction

In this survey, we discuss possible scenarios of behaviour of the dual part of sequences generated by primal-dual Newton-type methods when applied to optimization problems with nonunique multipliers associated to a solution. We consider the subclass of multipliers called critical (this subclass violates the

[*]Research of the first author is supported by the Russian Foundation for Basic Research Grants 06-01-00530, 07-01-00270, 07-01-00416 and 07-01-90102-Mong, and by RF President's Grant NS-693.2008.1 for the support of leading scientific schools. The second author is supported in part by CNPq Grants 301508/2005-4 and 471267/2007-4, by PRONEX–Optimization, and by FAPERJ.

second-order sufficient condition for optimality), possessing some very special properties. Specifically, convergence to a critical multiplier appears to be a typical scenario of dual behaviour of primal-dual Newton-type methods when critical multipliers do exist, and when dual sequence converges. Moreover, along with the possible absence of dual convergence, attraction to critical multipliers is precisely the reason for slow primal convergence that is usual for problems with degenerate constraints. It is interesting to note that this negative effect of attraction to critical multipliers was first discovered experimentally. Now we can claim that this is indeed a persistent effect, and this claim is supported both by some kind of theoretical explanations, as well as by serious numerical evidence. On the other hand, critical multipliers turn out to have some special analytical stability properties: noncritical multipliers should not be expected to be stable subject to generic parametric perturbations of optimality systems. All this is quite remarkable because the set of critical multipliers is normally "thin" within the set of all multipliers.

Before proceeding, let us stress that many questions concerned with critical multipliers remain open. What we currently have at hand is some characterization of what *would have happened* in the case of convergence of Newton-type methods to a noncritical multiplier, plus some explanation of why this behavior is unlikely (at least for some particular cases), plus quite serious numerical evidence showing that this behavior is indeed unlikely. This is actually enough for all practical purposes: there is absolutely no doubt that the effect of attraction exists, and everyone who deals with Newton-type methods for optimization problems with degenerate constraints must take care of it. However, a completely satisfactory theoretical result should apparently be of a positive nature, demonstrating that *the set of critical multipliers is indeed an attractor* in some sense, and we hope that a result of this type might eventually be proved. But for now, it does not exist.

The paper is organized as follows. In Section 2,we recall the notion of critical multipliers for equality-constrained problems, and discuss the effect of attraction to such multipliers for the Newton-Lagrange method (NLM), which is the particular case of sequential quadratic programming (SQP) methods for this specific class of problems. In particular, we present the result demonstrating that in the simple case of full degeneracy, and when there are no terms of order higher than 2, convergence to noncritical multipliers is highly unlikely, at least when there exist multipliers satisfying second-order sufficient conditions for minimizers or maximizers. In Section 3, we proceed with special stability properties of critical multipliers for the Lagrange optimality system subject to parametric perturbations. Finally, in Section 4, we discuss extensions of the notion of criticality and the related theory to problems with mixed equality and inequality constraints.

2. Critical multipliers

Let us now recall the notion of a *critical multiplier*, suggested in [18], which plays a central role in understanding dual behaviour of Newton-type schemes for degenerate problems. It is convenient to do this for the case of equality-constrained problem first.

To this end, consider the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & F(x) = 0, \end{array} \tag{1.1}$$

with smooth objective function $f : \mathbf{R}^n \to \mathbf{R}$, and smooth constraint mapping $F : \mathbf{R}^n \to \mathbf{R}^l$.

Stationary points of problem (1.1) and the associated Lagrange multipliers are characterized by the Lagrange optimality system

$$\frac{\partial L}{\partial x}(x, \lambda) = 0, \quad F(x) = 0,$$

where

$$L : \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}, \quad L(x, \lambda) = f(x) + \langle \lambda, F(x) \rangle,$$

is the Lagrangian function. Let $\bar{x} \in \mathbf{R}^n$ be a stationary point, and let $\mathcal{M}(\bar{x})$ be the (nonempty) set of Lagrange multipliers associated with $\bar{x}$. As is well known, if $\bar{x}$ is a local solution, then $\mathcal{M}(\bar{x})$ is nonempty if some constraint qualification (CQ) holds at $\bar{x}$, and for equality-constrained problems, most of the relevant CQs reduce to the following classical regularity condition:

$$\operatorname{rank} F'(\bar{x}) = l,$$

which also guarantees that the set $\mathcal{M}(\bar{x})$ of Lagrange multipliers is a singleton.

In this work, we are interested in dual behaviour of primal-dual Newton methods, and thus, the case of interest here is when $\mathcal{M}(\bar{x})$ is not a singleton. At issue, therefore, are *degenerate* problems, where $\bar{x}$ does not satisfy the regularity condition.

The case of violation of classical CQs has been a subject of considerable interest in the past fifteen years, both in the general case (e.g., [16, 38, 9, 13, 14, 1, 4, 24, 8, 37, 25, 36, 21, 35, 23, 20, 19]) and in the special case of equilibrium or complementarity constraints [27, 32, 34, 33, 15, 17, 2, 7, 3]).

In addition to CQs (or lack of them), second-order conditions serve as an important ingredient for convergence and rate of convergence analysis of Newton-type methods. Recall that the second-order sufficient condition for optimality (SOSC) holds at $\bar{x}$ with a multiplier $\bar{\lambda} \in \mathcal{M}(\bar{x})$, if

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in \ker F'(\bar{x}) \setminus \{0\},$$

If SOSC holds (with some multipliers) then the point $\bar{x}$ is a strict local minimizer of problem (1.1), and primal convergence to $\bar{x}$ can be expected for (good implementations of) good algorithms, even in degenerate cases. But the speed of convergence is often slow. It has been observed (see, e.g., [36, Sec. 6], and [23]) that, when primal convergence is slow, the reason for this is not so much degeneracy as such, but some undesirable behaviour of the dual sequence. Among various scenarios of this behaviour, one of the prominent ones appears to be dual convergence to multipliers violating SOSC. Understanding this phenomenon better is important, in particular, for assessing the chances of applicability of various local stabilization/regularization methods (like sSQP) that have been proposed recently to tackle degeneracy [16, 13, 26, 8, 21, 35] (see also [38, 9, 37, 36]). Some of those methods do achieve superlinear or quadratic convergence despite degeneracy, if their primal-dual starting point is close to a point satisfying SOSC. To get such a starting point, the issue of dual behaviour of an "outer" (global) phase of the algorithm is again highly important.

For a current iterate $(x^k,\, \lambda^k)$, the step of the basic SQP method consists of taking the next iterate $(x^{k+1},\, \lambda^{k+1})$ as a stationary-point-multiplier pair of the quadratic programming subproblem

$$\begin{array}{ll} \text{minimize} & \langle f'(x^k),\, x - x^k\rangle + \frac{1}{2}\langle \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)(x - x^k),\, x - x^k\rangle \\ \text{subject to} & F(x^k) + F'(x^k)(x - x^k) = 0. \end{array}$$

In other words, this is NLM, i.e., the standard Newton method for the Lagrange system of equations, given by the iteration system

$$\frac{\partial L}{\partial x}(x^k,\, \lambda^k) + \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)(x^{k+1} - x^k) + (F'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) = 0,$$

$$F(x^k) + F'(x^k)(x^{k+1} - x^k) = 0.$$

Definition 1. **[18]** *A multiplier* $\bar{\lambda} \in \mathcal{M}(\bar{x})$ *associated with a stationary point* $\bar{x}$ *of problem* (1.1) *is referred to as* critical *if*

$$\exists \xi \in \ker F'(\bar{x}) \setminus \{0\} \text{ such that } \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi \in \operatorname{im}(F'(\bar{x}))^{\mathrm{T}},$$

and noncritical *otherwise.*

It can be immediately seen that SOSC is violated for any critical multiplier $\bar{\lambda}$. If $F'(\bar{x}) = 0$ (the case of full degeneracy), criticality of $\bar{\lambda}$ reduces to saying that the matrix $\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})$ is singular. In the general case, criticality can be interpreted as singularity of the so-called reduced Hessian of the Lagrangian.

Evidently, critical multipliers form a special subclass within the multipliers violating SOSC, and what is important about them is that they serve as attractors for the dual sequences of Newton-type methods: convergence to such multipliers is something that should be expected to happen when the dual sequence converges and critical multipliers exist. This is quite remarkable, considering that the set of critical multipliers is normally "thin" in $\mathcal{M}(\bar{x})$. Moreover, the reason for slow primal convergence is precisely dual convergence to critical multipliers, or the lack of convergence. If dual sequence were to converge to a noncritical multiplier, primal convergence rate would have been superlinear.

Let us now justify these claims for the "most pure" case of full degeneracy, and when there are no terms of order higher than 2. Specifically, let the objective function f and the constraint mapping F be quadratic:

$$f(x) = \frac{1}{2}\langle Ax, x\rangle, \quad F(x) = \frac{1}{2}B[x, x],$$

where $A(= f''(0))$ is a symmetric $n \times n$-matrix, and $B(= F''(0)) : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^l$ is a symmetric bilinear mapping. The stationary point of interest is $\bar{x} = 0$, and $\mathcal{M}(\bar{x}) = \mathbf{R}^l$.

For each $\lambda \in \mathbf{R}^l$ (and each $x \in \mathbf{R}^n$), set

$$H(\lambda) = A + \lambda B \left(= \frac{\partial^2 L}{\partial x^2}(x, \lambda) \right).$$

Then the NLM iteration can be written in the form

$$H(\lambda^k)x^{k+1} + (B[x^k])^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) = 0,$$

$$B[x^k, x^{k+1}] = \frac{1}{2}B[x^k, x^k],$$

which implies the estimate

$$\begin{aligned}\langle H(\lambda^k)x^k,\, x^{k+1}\rangle &= -\langle \lambda^{k+1}-\lambda^k,\, B[x^k,\, x^k]\rangle\\ &= -2\langle \lambda^{k+1}-\lambda^k,\, B[x^k,\, x^{k+1}]\rangle\\ &= o(\|x^k\|\|x^{k+1}\|),\end{aligned}$$

provided the dual trajectory $\{\lambda^k\}$ converges. Furthermore, assuming that it converges to a noncritical multiplier $\bar{\lambda}$, we derive

$$\begin{aligned}x^{k+1} &= -(H(\lambda^k))^{-1}(B[x^k])^{\mathrm{T}}(\lambda^{k+1}-\lambda^k)\\ &= o(\|x^k\|),\end{aligned}$$

and

$$B[x^k,\, x^k] = o(\|x^k\|^2).$$

We thus proved

Theorem 1. **[22]** *Let $\{(x^k,\, \lambda^k)\}$ be a trajectory generated by NLM for problem (1.1) with quadratic data, and suppose that this trajectory converges to $(\bar{x},\, \bar{\lambda})$ with $\bar{x}=0$ and some $\bar{\lambda}\in\mathcal{M}(\bar{x})(=\mathbf{R}^l)$.*

Then

$$\left\langle \frac{\partial^2 L}{\partial x^2}(0,\, \bar{\lambda})x^k,\, x^{k+1}\right\rangle = o(\|x^k\|\|x^{k+1}\|),$$

and if $\bar{\lambda}$ is a noncritical multiplier, then

$$x^{k+1} = o(\|x^k\|),$$

$$F''(0)[x^k,\, x^k] = o(\|x^k\|^2).$$

Theorem 1 implies that, if the dual limit $\bar{\lambda}$ is noncritical, then the primal trajectory converges to $\bar{x}=0$ superlinearly. Moreover, two additional asymptotic relations hold, the second of which implies that $\{x^k\}$ converges tangentially to the null set of the quadratic mapping corresponding to $F''(0)$. We next explain why the behaviour characterized by these two relations is highly unlikely to occur.

Suppose that there exists an infinite subset K of $\{0,\, 1,\, \ldots\}$ such that $x^k\neq 0$ and $x^{k+1}\neq 0\ \forall\, k\in K$, the subsequence $\{x^k/\|x^k\|\mid k\in K\}$ converges to some $\xi\in\mathbf{R}^n$, and the subsequence $\{x^{k+1}/\|x^{k+1}\|\mid k\in K\}$ converges either to ξ or to $-\xi$. In particular, this is automatic if the entire $\{x^k\}$ converges to $\bar{x}=0$ tangentially to a direction $\xi\neq 0$, which is quite a reasonable numerical behaviour, typically observed in the experiments. Then the asymptotic relations in Theorem imply the equalities

$$\left\langle \frac{\partial^2 L}{\partial x^2}(0,\, \bar{\lambda})\xi,\, \xi \right\rangle = 0, \quad F''(0)[\xi,\, \xi] = 0,$$

which further imply

$$\begin{aligned} \langle f''(0)\xi,\, \xi\rangle &= \langle f''(0)\xi,\, \xi\rangle + \langle \bar{\lambda},\, F''(0)[\xi,\, \xi]\rangle \\ &= \left\langle \frac{\partial^2 L}{\partial x^2}(0,\, \bar{\lambda})\xi,\, \xi \right\rangle \\ &= 0. \end{aligned}$$

Hence,

$$\begin{aligned} \left\langle \frac{\partial^2 L}{\partial x^2}(0,\, \lambda)\xi,\, \xi \right\rangle &= \langle f''(0)\xi,\, \xi\rangle + \langle \lambda,\, F''(0)[\xi,\, \xi]\rangle \\ &= 0 \quad \forall\, \lambda \in \mathcal{M}(\bar{x}), \end{aligned}$$

which means that SOSC does not hold with any multiplier associated with $\bar{x}$. Moreover, the following weaker SOSC does not hold at $\bar{x}$:

$$\forall\, \xi \in \ker F'(\bar{x}) \setminus \{0\}\ \exists\, \lambda \in \mathcal{M}(\bar{x}) \text{ such that } \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \lambda)\xi,\, \xi \right\rangle > 0.$$

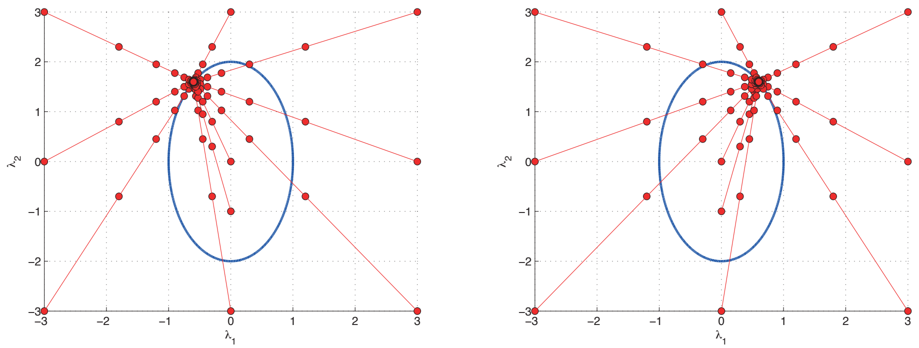
Moreover, the corresponding SOSC for maximizers does not hold as well. Clearly, this is a rather special situation, and beyond this special case, convergence to noncritical multipliers is highly unlikely.

Unfortunately, if we perturb the data by higher-order terms, the argument becomes less clean. More precisely, the last two asymptotic relations in Theorem 1 remain valid (in case of convergence to a noncritical multiplier), but the first relation takes the cruder form

$$\left\langle \frac{\partial^2 L}{\partial x^2}(0,\, \bar{\lambda})x^k,\, x^{k+1} \right\rangle = o(\|x^k\|\|x^{k+1}\|) + o(\|x^k\|^2),$$

and generally, this does not lead to the desired conclusions. The same happens for other kinds of perturbations. Furthermore, passing to the general case of degeneracy (when the first derivatives of f and F at $\bar{x}$ may not be equal to zeroes), and reducing it to the fully degenerate case by means of the Liapunov–Schmidt procedure (see, e.g., [12]) results in more-or-less the same effect, and in our numerical experiments for random problems with linear-quadratic data, some examples of convergence to noncritical multipliers were indeed encountered beyond the fully degenerate case. However, the result regarding superlinear primal convergence in the case of dual convergence to a noncritical multiplier remains valid.

We next illustrate the effect of attraction by some examples.



(a) $x^0 = (1,\, 2)$ (b) $x^0 = (2,\, 1)$

Figure 1: Dual trajectories for Example 1.

Example 1. **[4]** The problem

$$\begin{array}{ll} \text{minimize} & -x_1^2 - x_2^2 \\ \text{subject to} & x_1^2 - x_2^2 = 0,\ x_1 x_2 = 0, \end{array}$$

has the unique feasible point (hence, unique solution) $\bar{x} = 0$, with $\mathcal{M}(\bar{x}) = \mathbf{R}^2$. This solution violates both the regularity condition (since $F'(\bar{x}) = 0$) and SOSC for any $\bar{\lambda} \in \mathbf{R}^2$. Critical multipliers are those $\bar{\lambda}$ that satisfy $4\bar{\lambda}_1^2 + \bar{\lambda}_2^2 = 4$.

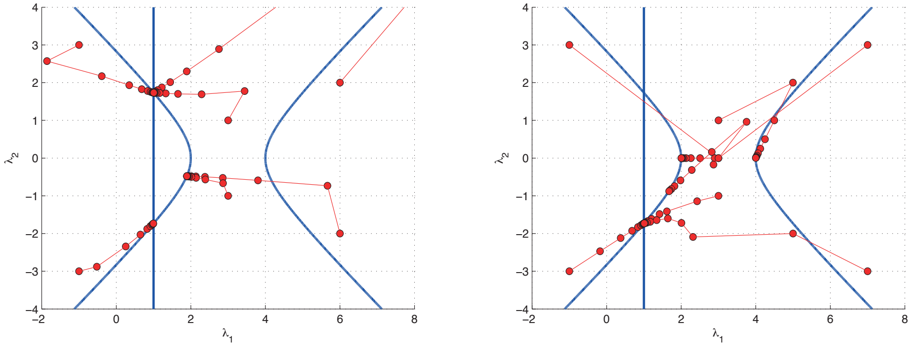
In Figure 1, critical multipliers are represented by the blue line (in the form an oval). Some NLM dual trajectories from primal starting points $x^0 = (1,\, 2)$ and $x^0 = (2,\, 1)$ are represented by red lines, with dots for the dual iterates. It can be shown analytically that for this problem, the dual limit point depends exclusively on x^0 and not on λ^0. Specifically, for any k the full (with $\alpha_k = 1$) primal-dual NLM step is given by

$$x^{k+1} = \frac{1}{2}x^k, \quad \lambda^k = \frac{1}{2}\lambda^k + \left(-\frac{(x_2^0)^2 - (x_1^0)^2}{2((x_1^0)^2 + (x_2^0)^2)}, \frac{2x_1^0 x_2^0}{(x_1^0)^2 + (x_2^0)^2} \right),$$

and the dual trajectory converges to

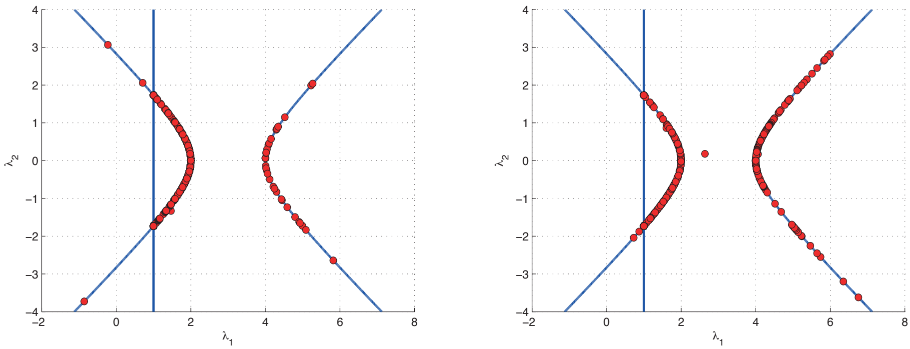
$$\bar{\lambda} = \left(-\frac{(x_2^0)^2 - (x_1^0)^2}{((x_1^0)^2 + (x_2^0)^2)}, \frac{4x_1^0 x_2^0}{(x_1^0)^2 + (x_2^0)^2} \right).$$

Note that the latter is a critical multiplier whatever is taken as x^0, and that convergence is slow (only linear).



(a) $x^0 = (1,\, 2,\, 3)$ (b) $x^0 = (3,\, 2,\, 1)$

Figure 2: Dual trajectories for Example 2.



(a) $x^0 = (1,\, 2,\, 3)$ (b) $x^0 = (3,\, 2,\, 1)$

Figure 3: Distribution of dual iterates at the time of termination for Example 2.

Example 2. The problem

$$\begin{array}{ll} \text{minimize} & x_1^2 - x_2^2 + 2x_3^2 \\ \text{subject to} & -\frac{1}{2}x_1^2 + x_2^2 - \frac{1}{2}x_3^2 = 0,\ x_1x_3 = 0, \end{array}$$

has the unique solution $\bar{x} = 0$, with $\mathcal{M}(\bar{x}) = \mathbf{R}^2$. This solution violates the regularity condition (since $F'(\bar{x}) = 0$) but satisfies SOSC with any $\bar{\lambda}$ such that

$\bar{\lambda}_1 \in (0, 1)$, $(\bar{\lambda}_1 - 3)^2 - \bar{\lambda}_2^2 > 1$. Critical multipliers are those $\bar{\lambda}$ that satisfy $\bar{\lambda}_1 = 1$ or $(\bar{\lambda}_1 - 3)^2 - \bar{\lambda}_2^2 = 1$.

In Figures 2 and 3, critical multipliers are represented by the blue line (they form two branches of a hyperbola and a vertical line). Note that in Figure 3 (b), there is one dual iterate at the time of termination, which is not close to any critical multiplier. This is a result of non-convergence of the dual trajectory. Figure 4 presents the run that produces this point. It is interesting to note that even in this case critical multipliers seem to play an important role for the behaviour of the dual trajectory: the latter moves "along" the set of critical multipliers.

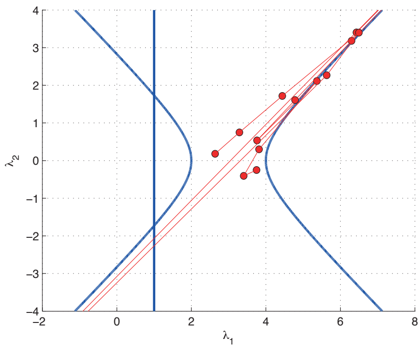


Figure 4: Dual trajectory for Example 2, $x^0 = (3, 2, 1)$, $\lambda^0 = (3.75, -0.25)$.

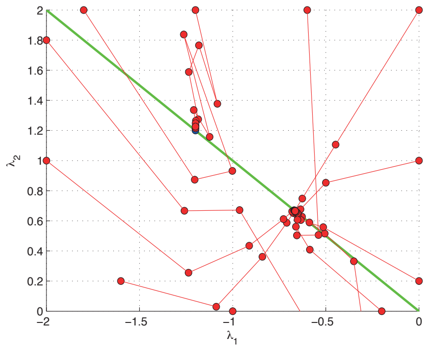


Figure 5: Dual trajectories for Example 3, $x^0 = (1, 2, 3)$.

Example 3. The problem

$$\begin{array}{ll} \text{minimize} & x_1^2 + x_2^2 + x_3^2 \\ \text{subject to} & x_1 + x_2 + x_3 + x_1^2 + x_2^2 + x_3^2 = 0, \; x_1 + x_2 + x_3 + x_1 x_3 = 0, \end{array}$$

has the unique solution $\bar{x} = 0$, with $\mathcal{M}(\bar{x}) = \{\bar{\lambda} \in \mathbf{R}^2 \mid \bar{\lambda}_1 + \bar{\lambda}_2 = 0\}$. This solution violates the regularity condition (since $F'(\bar{x}) = 0$), but satisfies SOSC with any $\bar{\lambda} \in \mathcal{M}(\bar{x})$ such that $\bar{\lambda}_1 > -2/3$. Critical multipliers are $\bar{\lambda} = (-6/5, 6/5)$ and $\bar{\lambda} = (-2/3, 2/3)$.

Figure 5 shows the set of multipliers (green line) and some NLM dual trajectories for $x^0 = (1, 2, 3)$. Critical multipliers $\bar{\lambda} = (-6/5, 6/5)$ and $\bar{\lambda} = (-2/3, 2/3)$ can be barely seen in the figure, because of dual iterates accumulating around those points. As expected, convergence is slow.

The results of numerical experiments with randomly generated problems with linear-quadratic data, and with small specific problems taken from various sources can be found in [23, 20]. These results put in evidence that the effect of attraction exists, and that it destroys the superlinear rate of primal convergence.

A natural and important question is whether the attraction phenomenon shows not only for simple implementations of the basic Newton method but also for its relevant modifications, and for sophisticated professional implementations, which are supplied by various smart heuristics. And if so, whether it still causes lack of superlinear convergence. In [22], an affirmative answers to these questions can be found for two well known and widely used algorithms: the linearly constrained Lagrangian (LCL) method (see [31, 28, 10]), on which the MINOS [29] solver is based, and a class of quasi-Newton SQP methods (see, e.g., [30, Ch. 18] and [5, Ch. 18]), related, in particular, to the SNOPT solver [11].

3. Parametric pertubrations

We next explain what we mean by saying that critical Lagrange multipliers possess not only special numerical but also special analytical stability properties. Consider now the parametric equality-constrained problem

$$\begin{array}{ll} \text{minimize} & f(\sigma,\, x) \\ \text{subject to} & F(\sigma,\, x) = 0, \end{array} \tag{3.2}$$

where the objective function $f : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}$, and the constraint mapping $F : \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^l$ depend smoothly on the parameter σ varying in the space $\mathbf{R}^s$ of parameter values. The associated parametric Lagrange system has the form

$$\frac{\partial L}{\partial x}(\sigma,\, x,\, \lambda) = 0, \quad F(\sigma,\, x) = 0,$$

where

$$L : \mathbf{R}^s \times \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}, \quad L(\sigma,\, x,\, \lambda) = f(\sigma,\, x) + \langle \lambda,\, F(\sigma,\, x)\rangle,$$

is the Lagrangian function of problem (3.2).

Let $\bar{\sigma} \in \mathbf{R}^s$ be a fixed (base) parameter value, and let $\bar{x}$ be a stationary point of the unperturbed problem corresponding to $\sigma = \bar{\sigma}$, and let $\mathcal{M}(\bar{\sigma},\, \bar{x})$ be the set of associated Lagrange multipliers.

Proposition 1. [18, 19] *Let $\bar{\lambda} \in \mathcal{M}(\bar{\sigma},\, \bar{x})$ be a noncritical multiplier.*

If there exist sequences $\{\sigma^k\} \subset \mathbf{R}^s \setminus \{\bar{\sigma}\}$, $\{x^k\} \subset \mathbf{R}^n$ and $\{\lambda^k\} \subset \mathbf{R}^l$ such that $\{\sigma^k\} \to \bar{\sigma}$, $\{x^k\} \to \bar{x}$, $\{\lambda^k\} \to \bar{\lambda}$, and for each k, the point $(x^k,\, \lambda^k)$ is a *solution of the parametric Lagrange system with* $\sigma = \sigma^k$*, then*

$$\|x^k - \bar{x}\| + \operatorname{dist}(\lambda^k, \mathcal{M}(\bar{\sigma}, \bar{x})) = O(\|\sigma^k - \bar{\sigma}\|),$$

and any limit point (d, ξ) *of the sequence* $\{(\sigma^k - \bar{\sigma}, x^k - \bar{x})/\|\sigma^k - \bar{\sigma}\|\}$ *satisfies the equality*

$$\frac{\partial F}{\partial \sigma}(\bar{\sigma}, \bar{x})d + \frac{\partial F}{\partial x}(\bar{\sigma}, \bar{x})\xi = 0.$$

The tatter equality implies the inclusion

$$\frac{\partial F}{\partial \sigma}(\bar{\sigma}, \bar{x})d \in \operatorname{im} \frac{\partial F}{\partial x}(\bar{\sigma}, \bar{x}).$$

Degeneracy of constraints of the unperturbed problem (3.2) at $\bar{x}$ means that the right-hand side is a proper subspace in $\mathbf{R}^l$, and hence, the latter may hold only for very special sequences $\{\sigma^k\}$, unless the parametrization is pathological.

At the same time, there exist reasonable sufficient conditions for stability of a specific multiplier subject to directional perturbations, based on the so-called 2-regularity concept [18, 19]. We skip the details since they would require an extensive discussion. We only stress that this condition may hold for critical multipliers only.

4. Extension to mixed constraints

The notion of criticality and the related theory can be extended to the the mixed-constrained case in several reasonable ways. Consider the mathematical programming problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & F(x) = 0,\ G(x) \le 0, \end{array} \tag{4.3}$$

where $G : \mathbf{R}^n \to \mathbf{R}^m$ is a smooth mappings defining inequality constraints.

Stationary points of problem (4.3) and the associated Lagrange multipliers are characterized by the Karush–Kuhn–Tucker (KKT) optimality system

$$\frac{\partial L}{\partial x}(x, \lambda, \mu) = 0, \quad F(x) = 0,$$

$$G(x) \le 0,\ \mu \ge 0,\ \langle \mu, G(x) \rangle = 0,$$

where

$$L : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}, \quad L(x, \lambda, \mu) = f(x) + \langle \lambda, F(x) \rangle + \langle \mu, G(x) \rangle,$$

is the Lagrangian function of problem (4.3). Let now $\bar{x} \in \mathbf{R}^n$ be a stationary point of problem (4.3), and let $\mathcal{M}(\bar{x})$ be the (nonempty) set of Lagrange multipliers associated with $\bar{x}$, that is, those pairs $(\lambda, \mu) \in \mathbf{R}^l \times \mathbf{R}^m$ satisfying the KKT system with $x = \bar{x}$. As is well known, if $\bar{x}$ is a local solution of problem (4.3), then $\mathcal{M}(\bar{x})$ is nonempty if the Mangasarian–Fromovitz constraint qualification (MFCQ) holds at $\bar{x}$, that is,

$$\operatorname{rank} F'(\bar{x}) = l \text{ and } \exists \bar{\xi} \in \ker F'(\bar{x}) \text{ such that } G'_{I(\bar{x})}(\bar{x})\bar{\xi} < 0,$$

where $I(\bar{x}) = \{i = 1, \ldots, m \mid G_i(\bar{x}) = 0\}$ is the set of indices of inequality constraints active at $\bar{x}$.

Recall that in this work, we are interested in those cases when $\mathcal{M}(\bar{x})$ is not a singleton. For mixed-constrained problems, this kind of degeneracy means that $\bar{x}$ does not satisfy the so-called strict MFCQ (that is, MFCQ combined with the requirement that the multiplier be unique). Hence, $\bar{x}$ also does not satisfy the stronger linear independence constraint qualification (LICQ), which can be expressed in the form

$$\operatorname{rank} \begin{pmatrix} F'(\bar{x}) \\ G'_{I(\bar{x})}(\bar{x}) \end{pmatrix} = l + |I(\bar{x})|,$$

where $|I|$ stands for the cardinality of a finite set I.

Furthermore, for mixed-constrained problems, we say that SOSC holds at $\bar{x}$ with a multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$, if

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall \xi \in C(\bar{x}) \setminus \{0\},$$

where

$$C(\bar{x}) = \left\{\xi \in \ker F'(\bar{x}) \,\middle|\, G'_{I(\bar{x})}(\bar{x})\xi \le 0, \ \langle f'(\bar{x}), \xi \rangle \le 0\right\}$$

is the critical cone of problem (4.3) at $\bar{x}$.

Possible scenarios of dual behavior of Newton-type algorithms for mixed-constrained problems are somewhat more diverse than for purely equality-constrained problems. However, the effect of attraction to appropriately defined critical multipliers still exists, at least when the set of indices of active inequality constraints of subproblems stabilizes, and it does result in slow primal convergence. The detailed discussion of these issues can be found in [20].

In the remainder of this survey, we demonstrate one theoretical result concerned with (non)critical multipliers for general mixed-constrained problems. The result will be given in the form of an error-bound, in order to avoid some extra notation. The corresponding upper-Lipschitz stability result subject to parametric perturbations (like the one in Proposition above) can be easily recovered by the reader. The following notion of a critical multiplier appears suitable in this context.

Definition 1. *A multiplier* $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ *associated with a stationary point* $\bar{x}$ *of problem* (4.3) *is referred to as* critical *if the system*

$$\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi + (F'(\bar{x}))^{\mathrm{T}}\eta + (G'(\bar{x}))^{\mathrm{T}}\zeta = 0,$$

$$F'(\bar{x})\xi = 0,$$

$$G'_{A_+}(\bar{x})\xi = 0,$$

$$\zeta_{A_0} \geq 0, \ G'_{A_0}(\bar{x})\xi \leq 0,$$

$$\zeta_i\langle G'_i(\bar{x}), \xi\rangle = 0, \ i \in A_0,$$

$$\zeta_N = 0,$$

has a solution (ξ, η, ζ) *with* $\xi \neq 0$*, and* noncritical *otherwise.*

Note that when there are no inequality constraints, this notion reduces to the notion of a critical multiplier for equality-constrained problems.

It can be easily seen that SOSC is violated for any critical multiplier $(\bar{\lambda}, \bar{\mu})$.

Theorem 2. [19, 6] *Let* $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ *be a noncritical multiplier.*
Then the estimate

$$\|x-\bar{x}\|+\mathrm{dist}((\lambda, \mu), \mathcal{M}(\bar{x})) = O\left(\left\|\frac{\partial L}{\partial x}(x, \lambda, \mu)\right\| + \|F(x)\| + \|\min\{\mu, -G(x)\}\|\right)$$

holds for $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ *close enough to* $(\bar{x}, \bar{\lambda}, \bar{\mu})$.

Similar estimates but under somewhat stronger assumptions follow from the results in [14, 8]. This estimate does not rely on any CQ: in a neighborhood of a noncritical multiplier, it turns out to be possible to estimate the distance to the primal-dual solution set via the residual of the KKT system in upper-Lipschitzian way. Such results are widely used in modern numerical methods for optimization problems with degenerate constraints.

References

[1] *M. Anitescu,* Nonlinear programs with unbounded Lagrange multiplier sets. Preprint ANL/MCS-P796-0200. Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, 2000.

[2] *M. Anitescu,* On using the elastic mode in nonlinear programming approaches to mathematical programs with complementarity constraints. SIAM J. Optim., 15:1203–1236, 2005.

[3] *M. Anitescu, P. Tseng, and S.J. Wright,* Elastic-mode algorithms for mathematical programs with equilibrium constraints: global convergence and stationarity properties. Math. Program., 110:337–371, 2007.

[4] *A.V. Arutyunov,* Optimality Conditions: Abnormal and Degenerate Problems. Kluwer Academic Publishers, Dordrecht, 2000.

[5] *J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal, and C. Sagastizábal,* Numerical Optimization: Theoretical and Practical Aspects. Springer–Verlag, Berlin, Germany, 2006. Second Edition.

[6] *D. Fernández and M. Solodov,* Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems. IMPA Preprint A568, November 2007 (revised September 2008).

[7] *R. Fletcher, S. Leyffer, D. Ralph, and S. Scholtes,* Local convergence of SQP methods for mathematical programs with equilibrium constraints. SIAM J. Optim., 17:259–286, 2006.

[8] *A. Fischer,* Local behaviour of an iterative framework for generalized equations with nonisolated solutions. Math. Program., 94:91–124, 2002.

[9] *A. Fischer,* Modified Wilson's method for nonlinear programs with nonunique multipliers. Math. Oper. Res. 24:699–727, 1999.

[10] *M.P. Friedlander and M.A. Saunders,* A globally convergent linearly constrained Lagrangian method for nonlinear optimization. SIAM J. Optim., 15:863–897, 2005.

[11] *P.E. Gill, W. Murray, and M.A. Saunders,* SNOPT: An SQP algorithm for large-scale constrained optimization. SIAM J. Optim., 12:979–1006, 2002.

[12] *M. Golubitsky and D.G. Schaeffer,* Singularities and Groups in Bifurcation Theory. Vol. 1. Springer-Verlag, New York, Berlin, Heidelberg, 1985.

[13] *W.W. Hager,* Stabilized sequential quadratic programming. *Comput. Optim. Appl.*, 12:253–273, 1999.

[14] *W.W. Hager and M.S. Gowda,* Stability in the presence of degeneracy and error estimation. Math. Program., 85:181–192, 1999.

[15] *A. F. Izmailov and M. V. Solodov,* Complementarity constraint qualification via the theory of 2-regularity. SIAM J. Optim., 13: 368–385, 2002.

[16] *A. F. Izmailov,* Lagrange methods for finding degenerate solutions of conditional extremum problems. Comput. Math. Math. Phys., 36:423–429, 1996.

[17] *A. F. Izmailov,* Mathematical programs with complementarity constraints: regularity, optimality conditions, and sensitivity. Comput. Math. Math. Phys., 44:1145–1164, 2004.

[18] *A. F. Izmailov,* On the analytical and numerical stability of critical Lagrange multipliers. Comput. Math. Math. Phys., 45:930–946, 2005.

[19] *A. F. Izmailov,* Solution sensitivity for Karush–Kuhn–Tucker systems with nonunique Lagrange multipliers. Optimization, 2008. DOI 10.1080/02331930802434922.

[20] *A. F. Izmailov, and M.V. Solodov,* Examples of dual behaviour of Newton-type methods on optimization problems with degenerate constraints. Comput. Optim. Appl., 2007. DOI 10.1007/s10589-007-9074-4.

[21] *A. F. Izmailov, and M.V. Solodov,* Newton-type methods for optimization problems without constraint qualifications. SIAM J. Optim., 15:210–228, 2004.

[22] *A. F. Izmailov, and M.V. Solodov,* On attraction of linearly constrained Lagrangian methods and of stabilized and quasi-newton SQP methods to critical multipliers. Manuscript, 2008.

[23] *A. F. Izmailov, and M.V. Solodov,* On attraction of Newton-type iterates to multipliers violating second-order sufficiency conditions. Math. Program., 117:271–304, 2009.

[24] *A. F. Izmailov, and M.V. Solodov,* Optimality conditions for irregular inequality-constrained problems. SIAM J. Control Optim., 40:1280–1295, 2001.

[25] *A. F. Izmailov, and M.V. Solodov,* The theory of 2-regularity for mappings with Lipschitzian derivatives and its applications to optimality conditions. Math. Oper. Res., 27:614–635, 2002.

[26] *D.-H. Li and L. Qi,* A stabilized SQP method via linear equations. Applied mathematics technical report AMR00/5, The University of New South Wales, 2000.

[27] *Z.-Q. Luo, J.-S. Pang, and D. Ralph,* Mathematical Programs with Equilibrium Constraints. Cambridge University Press, Cambridge, UK, 1996.

[28] *B.A. Murtagh and M.A. Saunders,* A projected Lagrangian algorithm and its implementation for sparse nonlinear constraints. Math. Program. Study, 16:84–117, 1982.

[29] *B.A. Murtagh and M.A. Saunders,* MINOS 5.0 user's guide. Tech. Report SOL 83.20. Stanford University, December 1983.

[30] *J. Nocedal and S.J. Wright* Numerical Optimization. Springer–Verlag, New-York, 1999.

[31] *S.M. Robinson,* A quadratically convergent algorithm for general nonlinear programming problems. Math. Program., 3:145–156, 1972.

[32] *S. Scholtes and M. Stöhr,* Exact penalization of mathematical programs with equilibrium constraints. SIAM J. Control Optim., 37:617–652, 1999.

[33] *S. Scholtes and M. Stöhr,* How stringent is the linear independence assumption for mathematical programs with complementarity constraints? Math. Oper. Res., 26:851–863, 2001.

[34] *H. Scheel and S. Scholtes,* Mathematical programs with complementarity constraints: Stationarity, optimality and sensitivity. Math. Oper. Res., 25:1–22, 2000.

[35] *S.J. Wright,* An algorithm for degenerate nonlinear programming with rapid local convergence. SIAM J. Optim., 15:673–696, 2005.

[36] *S.J. Wright,* Constraint identification and algorithm stabilization for degenerate nonlinear programs. Math. Program., 95:137–160, 2003.

[37] *S.J. Wright,* Modifying SQP for degenerate problems. SIAM J. Optim., 13:470–497, 2002.

[38] *S.J. Wright,* Superlinear convergence of a stabilized SQP method to a degenerate solution. Comput. Optim. Appl., 11:253–275, 1998.