TOP (2015) 23:48–52
DOI 10.1007/s11750-015-0373-0

DISCUSSION

Rejoinder on: Critical Lagrange multipliers: what we currently know about them, how they spoil our lives, and what we can do about it

A. F. Izmailov · M. V. Solodov

Published online: 27 February 2015

To begin with, we would like to thank the TOP Editor-in-Chief Prof. Miguel Goberna for giving us an opportunity to discuss some issues related to critical Lagrange multipliers, which we consider as one of the most intriguing and challenging subjects of our research over the last decade or so.

We are also especially grateful to the colleagues who accepted the invitation of the Editor-in-Chief to participate in this discussion. All the discussants are leading experts in theoretical and numerical issues of optimization and variational analysis. Their insightful comments shed additional light on various aspects of the phenomenon in question, from a good number of different perspectives covering all the spectrum from "very theoretical" to "very practical". We believe that joint efforts by everybody involved led to a truly fruitful and stimulating discussion. We also hope that due to the diversity, it will be of interest for a wide readership. In addition, all the commentators point out promising directions for future research related to critical multipliers. Below, we summarize their proposals and give some comments on them.

This rejoinder refers to the comments available at doi:10.1007/s11750-015-0368-x, doi:10.1007/s11750-015-0369-9, doi:10.1007/s11750-015-0370-3, doi:10.1007/s11750-015-0371-2.

Research of the first author is supported by the Russian Foundation for Basic Research Grant 14-01-00113 and by CNPq Grant PVE 401119/2014-9 (Brazil). The second author is supported in part by CNPq Grant 302637/2011-7, by PRONEX–Optimization, and by FAPERJ.

A. F. Izmailov (✉)
Operations Research Department, Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University (MSU), Uchebniy Korpus 2, Leninskiye Gory 119991, Moscow, Russia
e-mail: izmaf@ccas.ru

M. V. Solodov
IMPA – Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil
e-mail: solodov@impa.br

1 Critical multipliers and (lack of) Lipschitzian error bound

Both José Mario Martínez and Andreas Fischer emphasize the following key feature of critical multipliers: they are exactly those Lagrange multipliers near which the local Lipschitzian error bound for the solution set of the optimality system fails to hold. This means, in particular, that when the dual part of a primal–dual iterative sequence approaches a critical multiplier, the residual of the optimality system is not a reliable measure for estimating the progress of an algorithm: the residual can become small relatively far from the solution set.

In fact, the lack of local Lipschitzian error bound can be regarded as an alternative definition of a critical multiplier. As remarked by Andreas Fischer, one of the potential advantages of this view is that the definition can be extended to variational problems more general than (or different from) the primal–dual optimality systems. Specifically, when the problem possesses multiple/nonisolated solutions, one can define special "critical solutions" as those near which the local Lipschitzian error bound estimating the distance to the solution set in terms of some residual of the problem does not hold. Figuring out special features of such "critical solutions", and in particular, their impact on the behavior of computational methods, can indeed be of crucial importance for some problem classes. One potential example are generalized Nash equilibrium problems (Facchinei and Kanzow 2010), which naturally possess nonisolated solutions.

Without attempting to draw any far-reaching conclusions at this early stage, we complete this section with two examples demonstrating the potential effect of "critical solutions" of nonlinear equations on the the basic Newton method.

Example 1 Consider the system of equations

$$\Phi(x) = 0 \tag{1}$$

with $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$,

$$\Phi(x) = (x_1\varphi(x),\ x_2\varphi(x)), \tag{2}$$

where

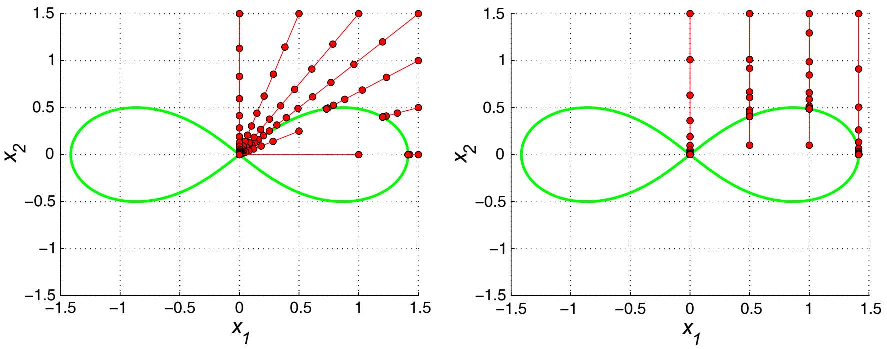
$$\varphi(x) = (x_1^2 + x_2^2)^2 - 2(x_1^2 - x_2^2).$$

The solution set of this system is the lemniscate shown in Fig. 1a as a thick line. Observe that $\varphi'(x) \neq 0$ for any $x \neq 0$ in the solution set, which easily implies that $\bar{x} = 0$ is the only "critical solution".

Figure 1a shows some sequences generated by the Newton method. Clearly, there is a wide region of starting points from which the sequences converge to the (unique) "critical solution". Moreover, in such cases, the rate of convergence is linear, while in the cases of convergence to other solutions it is superlinear.

Example 2 Consider the system of Eq. (1) with Φ defined by the following slight modification of (2):

$$\Phi(x) = (x_1\varphi(x),\ \varphi(x)).$$



(a) Example 1. **(b)** Example 2.

Fig. 1 Trajectories of the Newton method

The solution set is the same as in Example 2, and the only "critical solution" is again $\bar{x} = 0$.

Figure 1b demonstrates that, in this example, the sequences generated by the Newton method have now no tendency to converge to the "critical solution". Moreover, it can be easily seen analytically that the Newtonian directions in this example are always vertical (when well defined). Thus, the "critical solution" is not the attractor anymore. Nevertheless, the rate of convergence to the "critical solution" is only linear, while the rate of convergence to other solutions is superlinear.

Examples 1 and 2 suggest that the issue of "criticality" may also be relevant for problems without the special primal–dual structure of optimality systems. But the situation is not as clear, so early into the game.

2 Dual stabilization and globalization of stabilized algorithms

The comments by Daniel Robinson are mostly concerned with the challenge of developing efficient and practically sound algorithms based on dual stabilization procedures, intended to avoid attraction to critical multipliers, and thus guarantee fast convergence in the degenerate cases. One of his principal concerns is the use of the so-called "localization conditions" in the local convergence analysis of Newton-type methods for constrained optimization and variational problems (for some examples of such conditions, see, e.g., Izmailov and Solodov 2015, Theorem 5.1 on the classical sequential quadratic programming, Izmailov and Solodov 2015, Theorem 5.9 for the stabilized method).

Daniel Robinson believes that when the subproblems of the algorithm may have multiple solutions, the restrictions in the theoretical analysis on the choice of "appropriate" solutions cannot be ignored from the practical perspective. We agree that this position is not without a reason, of course. While not advocating to ignore the issues, we have a different view on this. But some difference of opinion only makes for a more stimulating discussion, we believe.

First, we would like to note that no localization conditions appear in Theorem 2 in our discussion paper. This is, of course, because our exposition is restricted to the simpler equality-constrained case (why this matters will be commented a bit later). In the case when subproblems' solutions need not be unique, the role of localization conditions is to point out to which kind of possible iterative sequences the rate-of-convergence analysis applies. Thus, the role of such conditions is purely theoretical, of course. They are not part of the algorithm's implementation. The message is that, if the corresponding sequences are generated (which are shown to exist as part of the analysis), then the local rate of convergence is as claimed. If solutions that violate the required condition are generated, the rate of convergence is not guaranteed.

In our view, it is crucial to stress that localization conditions on solutions of subproblems are plain unavoidable, except if subproblems are equality-constrained or strongly convex. See the discussion in Izmailov and Solodov (2015), Section 5.1 and Izmailov and Solodov (2015), Examples 5.1, 5.2. Those examples put in evidence that localization conditions cannot be omitted even when all of the following are satisfied: the linear independence constraint qualification, the second-order sufficient optimality condition, and strict complementarity (thus also the strong second-order sufficient optimality condition). This set of assumptions is the strongest possible!

So, unless one is prepared to claim that all algorithms should be designed (at least in some kind of local phase) on the basis of equality-constrained subproblems, or strongly convex ones, we submit that localization conditions cannot be avoided in local rate-of-convergence analysis. It is prudent to recall that such conditions appear already in the classical works by Stephen M. Robinson on local superlinear convergence of the usual sequential quadratic programming (Robinson 1974), and on linearly constrained Lagrangian methods (Robinson 1972). The examples in Izmailov and Solodov (2015) already cited show that these conditions cannot be dropped, even under the strongest of assumptions. The alternative can only consist of declaring that "practical" algorithms are only those based on equality-constrained or strongly convex subproblems, but such a claim would appear too extreme, in our opinion.

We also point out that the theory does not require computing a *global* solution of the nonconvex subproblem, but only its stationary point. Something which is certainly computationally reasonable. In addition, in our numerical experience, so far we have never encountered a case when superlinear convergence is lost because apparently "wrong" subproblem solutions are computed.

Another interesting observation related to dual stabilization and our discussion of the augmented Lagrangian methods is given by José Mario Martínez: for methods of this class the approximate KKT conditions hold asymptotically even when there are no Lagrange multipliers associated to a solution. And this is not so for Newton-type methods. Thus, in a sense, augmented Lagrangian methods (Birgin and Martínez 2014) possess the dual stabilization property even in the case when there are no dual solutions at all! But overall, the class of problems with no Lagrange multipliers still remains an open field from the numerical viewpoint, at least to a large extent.

Another important comment by José Mario Martínez is that sometimes, when a Newton-type method is approaching (or is known to be likely to approach, as in our setting) a critical dual solution with all the negative consequences, it may not be a good idea to insist on "staying Newtonian". It might be profitable to simplify, at least at some stage, to some extent. For example, one may stop using second derivatives of the constraints, and thus estimates of Lagrange multipliers that are known to lead to critical ones. Of course, this cannot possibly restore superlinear rate of convergence. But since it is likely to be lost anyway, the savings of discarding computing second derivatives may pay off overall (say, lead to stable convergence, not superlinear but acceptable). This is definitely true, though certainly problem dependent.

3 Tilt stability, full stability, and critical multipliers

Boris Mordukhovich raises very interesting questions about possible relations between the existence (or not) of critical Lagrange multipliers and some important stability concepts intensively investigated in modern optimization literature. Specifically, at issue are tilt stability (Poliquin and Rockafellar 1998) and full stability (Levy et al. 2000). These notions, originally introduced in the context of unconstrained minimization of an extended real-valued function, can be extended to constrained problems using the reformulation based on the indicator function of the feasible set.

The conjecture by Boris Mordukhovich is that for constrained optimization problems, these kinds of stability properties might exclude the existence of critical multipliers. Given the numerical difficulties associated with critical multipliers, existence of some relevant problem classes of this kind (degenerate but "stable" in some sense) would indeed be good news. In particular, when developing numerical methods, perhaps one might try to aim at seeking not arbitrary solutions, but rather solutions with specific "good" stability properties, thus avoiding the negative effect of critical multipliers.

These issues are currently under investigation, and it is too early to attempt any formal claims and firm conclusions. Our preliminary intuition suggests that tilt stability can be too weak for excluding the existence of critical Lagrange multipliers, because this kind of stability does not allow for any perturbations of the constraints. At the same time, full stability with sufficiently rich parametrization of constraints (say, allowing for arbitrary right-hand side perturbations) can indeed be incompatible with the existence of critical Lagrange multipliers, thus supporting the corresponding conjecture by Boris Mordukhovich.

References

Birgin EG, Martínez JM (2014) Practical augmented Lagrangian methods for constrained optimization. SIAM, Philadelphia

Facchinei F, Kanzow C (2010) Generalized Nash equilibrium problems. Ann Oper Res 175:177–211

Izmailov AF, Solodov MV (2015) Newton-type methods: a broader view. J Optim Theory Appl 164:577–620

Levy AB, Poliquin RA, Rockafellar RT (2000) Stability of locally optimal solutions. SIAM J Optim 10:580–604

Poliquin RA, Rockafellar RT (1998) Tilt stability of a local minimum. SIAM J Optim 8:287–299

Robinson SM (1972) A quadratically convergent algorithm for general nonlinear programming problems. Math Program 3:145–156

Robinson SM (1974) Perturbed Kuhn Tucker points and rates of convergence for a class of nonlinear-programming algorithms. Math Program 7:1–16