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

Daniel P. Robinson

1 Introduction

The invited discussion paper, *Critical Lagrange Multipliers: what we currently know about them, how they spoil our lives, and what we can do about it*, written by Izmailov and Solodov is a clear and concise overview of dual stabilization methods for solving problems of the form

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) \ \text{ subject to } \ h(x) = 0, \tag{1}$$

where the objective function f and constraint function h are assumed to be twice-continuously differentiable. They focus on the equality constrained problem (1) so that key concepts can be clearly explained, but most of the discussion and many of the results have been extended to problems with inequality constraints. As the title suggests, emphasis is placed on critical Lagrange multipliers and the challenges that they introduce for designing efficient and robust algorithms for solving (1).

The authors are well qualified for preparing such a discussion because they have been the driving force behind understanding critical Lagrange multipliers and for proving related local convergence results. More recently, the authors have developed a globally convergent stabilized sequential quadratic programming (sSQP) method (Izmailov et al. 2014b) (an example of a dual stabilization method) that is intended to benefit from the strong local convergence properties of sSQP that they have established previously. It is in this context, i.e., the development of globally convergent methods

---

This comment refers to the invited paper available at doi:10.1007/s11750-015-0372-1.

D. P. Robinson (✉)
Applied Mathematics and Statistics, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD, USA
e-mail: daniel.p.robinson@jhu.edu

based on sSQP, that I focus most of my commentary. Before proceeding, however, I quickly summarize some key points from their discussion.

Izmailov and Solodov first present the idea of critical Lagrange multipliers. Briefly, they are Lagrange multipliers for which the reduced Hessian matrix associated with the Lagrangian function is singular, i.e., the Hessian of the Lagrangian is singular when restricted to the null space of the constraint Jacobian. They explain that critical Lagrange multipliers are "thin", i.e., the *noncritical* multipliers are relatively open and dense relative to the complete set of Lagrange multipliers. This fact is important since they have proved superlinear convergence results for dual stabilization methods under assumptions that rely on the dual estimates being close enough to a noncritical Lagrange multiplier. Interestingly, they show that conventional Newton-like methods (e.g., sequential quadratic programming methods) often converge to critical Lagrange multipliers empirically, even though the multipliers are "thin".

The authors next discuss dual stabilization methods, one of which is the sSQP method. The sSQP algorithm repeatedly solves the subproblem

$$
\begin{aligned}
&\underset{x,\lambda}{\text{minimize}} \quad \langle f'\left(x^k\right), x - x^k \rangle + \tfrac{1}{2}\langle H\left(x^k, \lambda^k\right)\left(x - x^k\right), x - x^k \rangle + \tfrac{1}{2}\sigma^k \|\lambda\|^2 \\
&\text{subject to} \quad h\left(x^k\right) + h'\left(x^k\right)\left(x - x^k\right) - \sigma^k\left(\lambda - \lambda^k\right) = 0,
\end{aligned}
\tag{2}
$$

and uses an appropriate update to the dual regularization parameter $\sigma_k > 0$. Much like Newton's method for zero-finding, the sSQP algorithm is local, i.e., it is only guaranteed to converge if the starting point is close enough to an appropriate primal–dual solution to (1). The authors give an overview of the local convergence results for such methods, and provide a short discussion on how the methods have been globalized. Importantly, unlike conventional Newton-like methods, the globally convergent sSQP methods appear to be significantly more successful at avoiding critical Lagrange multipliers. As mentioned previously, my commentary will provide additional perspective concerning globalization.

2 The globalization of sSQP

Izmailov and Solodov's discussion of the globalization of sSQP methods is brief because very few methods exist, with most of them being developed within the last 2 or 3 years (Gill et al. 2013, 2014; Izmailov et al. 2014a, b; Fernández et al. 2013; Wright 2003). Moreover, it is probably safe to say that the best way to globalize sSQP is not yet clear. In this section, I discuss various aspects related to this topic that are motivated by the experience that my collaborators and I have gained over the last few years. As I believe that algorithms should be practically and theoretically sound, I will focus on the practical aspects of the globalization process.

2.1 One-phase versus two-phase approaches

A simple strategy for globalizing sSQP is to use a two-phase approach. The first phase (the global phase) may be any globally convergent method, whereas the second phase

(the local phase) is the sSQP method. The basic idea is quite simple: use the global phase to obtain an estimate of a primal–dual solution, and then use that estimate to initialize the local phase and hopefully recover the superlinear rate of convergence expected of sSQP.

There are two main challenges associated with this approach. First, it is difficult to develop conditions that reliably and efficiently decide when the global phase should transition to the local phase. Of course, if the transition occurs too soon, then the global phase could be continued, and the entire process repeated. Unfortunately, it is not difficult to imagine that this back-and-forth approach may sometimes be inefficient. Second, Izmailov and Solodov provide empirical evidence that suggests that conventional methods (e.g., sequential quadratic programming) may often converge to critical Lagrange multipliers. They explain that this has the potential of being a serious problem because numerical experience suggests that it substantially increases the likelihood that the local sSQP phase will *not* converge superlinearly. (This is essentially because the radius of convergence associated with a noncritical Lagrange multiplier decreases superlinearly with respect to its distance to a critical Lagrange multiplier.) It therefore seems that commonly used two-phase approaches, although successfully used on many problems, may never reliably produce superlinearly convergent iterates in practice, as predicted by the local convergence theory. This observation leads me to conclude that efficient and reliable globally convergent sSQP methods will either be *single-phase* approaches or two-phase approaches in which the first-phase also uses some form of dual stabilization (Robinson 2015).

2.2 Assumption concerning subproblem solutions

In this section, I discuss an assumption that is commonly used to establish the local superlinear convergence of sSQP methods. The assumption essentially says that once a primal–dual iterate gets close enough to a primal–dual solution, the solution to the sSQP subproblem (2) that is computed must satisfy certain error estimates [for example, see Izmailov and Solodov (2012, Property 2)]. An assumption of this kind is not surprising since subproblem (2) is generally nonconvex and, therefore, may have many local solutions. The calculation of such special solutions, although critical to proving superlinear local convergence, cannot be guaranteed in practice. This fact leads to a difference of opinion among researchers. One group believes that this assumption on the choice of subproblem solutions is minor. Their argument is usually based on the belief that an active-set QP solver, when applied to the sSQP subproblem, will compute the necessary solution. I do not know of any result in this direction and, in fact, I do not believe it to be true (at least not provably so). In practice, however, the evidence is mixed because all of the numerical experiments that I am aware of show that superlinear convergence is not achieved by sSQP on a nontrivial percentage of test problems. Of course, this somewhat disappointing performance may be caused by reasons other than the assumption on the subproblem solutions. The picture is not completely clear at this point. To my knowledge, there has not been a study that attempts to verify that the "correct" subproblem solutions are computed, but this is probably because such a verification is generally not possible, wherein lies the problem. I belong to the second group of researchers who believe that the assumption placed

on the subproblem solutions is unsatisfactory and should be avoided if possible. In the next section, I outline recent research that provides methods that do not require an assumption on which solution of the subproblem is found.

2.3 Some recent work

As mentioned in the previous two sections, in my opinion, the most promising globally convergent sSQP methods are one-phase methods that do not require any assumptions on which particular solution to the sSQP subproblem (2) is computed. Izmailov and Solodov state that such an assumption on subproblem (2) is unavoidable. This statement is true if a conventional active-set method is used and exact solutions of the subproblem are demanded. However, we have recently proposed an algorithm (Gill et al. 2013, 2014) that is globally convergent and locally equivalent to sSQP. The method uses a non-traditional active-set method and allows for inexact solutions of the subproblem. In particular, the method uses an ϵ-active-set bound-constrained quadratic programming (BCQP) solver and relaxes the termination conditions when certain verifiable conditions are satisfied. The details are complicated, but in general terms, it capitalizes on the close relationship between the solution of the sSQP subproblem (2) and the solution of a certain BCQP subproblem whose objective approximates a primal–dual augmented Lagrangian function (Robinson 2007; Gill and Robinson 2012). By exploiting this relationship, procedures for convexifying the Hessian of the Lagrangian function are used to ensure global convergence. Computable conditions that allow for inexact subproblem solutions are used to establish an equivalence (locally) to sSQP. Although the synchronization of these two aspects into a practical single-phase algorithm proved more difficult than anticipated, the global and local convergence results do not require an assumption about which subproblem solutions are computed.

We also reported numerical results (Gill et al. 2014) for our method from which we made an interesting observation. Our method performed substantially better compared to SNOPT (Gill et al. 2004)—a conventional sequential quadratic programming method—on problems that did not satisfy the linear independence constraint qualification (LICQ), i.e., on problems for which the set of constraint gradients associated with the constraints active at the solution were linearly dependent. Problems of this type are not discussed in detail by Izmailov and Solodov, perhaps because their presentation is focused on critical Lagrange multipliers. On a related note, it is important to mention that dual regularization may be viewed as a way of controlling the size of the dual variables *and* as a way of systematically regularizing the KKT matrix. Specifically, the dual regularized KKT matrix associated with problem (1) has the form

$$K := \begin{pmatrix} H\left(x^k, \lambda^k\right) & h'\left(x^k\right)^T \\ h'\left(x^k\right) & -\sigma^k I \end{pmatrix},$$

which implies that nonsingularity of K does not require $h'(x^k)$ to have full row rank. In our experience, this property provides an important practical numerical advantage.

3 Final comments

Izmailov and Solodov have provided a clear and concise overview of critical Lagrange multipliers and their effect on dual stabilization methods. They have produced a substantial body of theoretical results, mostly with respect to local convergence. The key remaining question is how to best globalize such methods. In particular, we seek methods with the following properties. (i) They are applicable to problems with both equality and inequality constraints. (ii) The methods are superlinearly convergent under weak assumptions. (iii) The methods are globally convergent under standard assumptions. (iv) The methods substantially outperform Newton-based methods on degenerate problems. (v) The methods are comparable to Newton-based methods on nondegenerate problems. I believe that the method proposed in Gill et al. (2013, 2014) comes the closest to satisfying these criteria, but I am sure that better methods are possible. I also believe that any method with the properties (i)–(v) must include strategies for: (a) the convexification of the subproblem; (b) the use of *primal* regularization; (c) the careful adjustment of the dual regularization parameter(s) when near *and* far from a solution; and (d) the use of inexact solutions of the subproblem.

References

Fernández D, Pilotta EA, Torres G (2013) An inexact restoration strategy for the globalization of the ssqp method. Comput Optim Appl 54(3):595–617

Gill PE, Kungurtsev V, Robinson DP (2013) A stabilized SQP method: global convergence. Center for Computational Mathematics Report CCoM 13–04. University of California, San Diego

Gill PE, Kungurtsev V, Robinson DP (2014) A stabilized SQP method: superlinear convergence. Center for Computational Mathematics Report CCoM 14–01. University of California, San Diego

Gill PE, Murray W, Saunders MA (2004) User's guide for SNOPT 7.1: a Fortran package for large-scale nonlinear programming. Numerical Analysis Report 04–1. Department of Mathematics, University of California, San Diego, La Jolla, CA

Gill PE, Robinson DP (2012) A primal–dual augmented Lagrangian. Comput Optim Appl 51:1–25. doi:10.1007/s10589-010-9339-1

Izmailov A, Solodov M, Uskov E (2014a) Combining stabilized SQP with the augmented lagrangian algorithm. IMPA Preprint A 754

Izmailov A, Solodov M, Uskov E (2014b) Globalizing stabilized sqp by smooth primal–dual exact penalty function. Tech. rep, IMPA preprint

Izmailov AF, Solodov MV (2012) Stabilized SQP revisited. Math Program 133:93–120. doi:10.1007/s10107-010-0413-3

Robinson DP (2007) Primal–dual methods for nonlinear optimization. Ph.D. thesis, Department of Mathematics, University of California San Diego, La Jolla, CA

Robinson DP (2015) Primal–dual active-set methods for large-scale optimization. J Optim Theory Appl 1–35. doi:10.1007/s10957-015-0708-x

Wright SJ (2003) Constraint identification and algorithm stabilization for degenerate nonlinear programs. Math Program 95(1, Ser. B):137–160. ISMP 2000, Part 3. Atlanta, GA