DISCUSSION

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

Andreas Fischer

The paper by Izmailov and Solodov provides an excellent overview of recent research on a phenomenon of slow convergence of optimization algorithms. The phenomenon was unknown about 10 years ago. For me, one of the most striking outcomes of this research is that the phenomenon may cause serious problems to software if applied to certain optimization problems. Although the occurrence of the phenomenon also depends on the point where an algorithm starts from, the sole existence of a critical multiplier frequently forces an algorithm to show this slow convergence. By now, this behavior is not fully understood and remedies are sought. The invited paper is well suited for researchers to become interested in the topic.

In the remainder of the discussion, I would like to first make a short excursion to a research problem that started to attract people in the second half of the nineties. Then, I will raise some questions related to both the excursion and the phenomenon of slow convergence.

A root of the invited paper's topic goes back to about 20 years. At that time, researchers thought about algorithmically dealing with optimization problems whose set of Lagrange multipliers associated to a stationary point contains more the one element (degenerate case). More in detail, let us consider the optimization problem

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad g(x) \leq 0, \; h(x) = 0, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, and $\mathbb{R}^n \to \mathbb{R}^l$ are sufficiently smooth functions. For the sake of simplicity, Izmailov and Solodov deal with equality constraints only. Here, inequality constraints are involved as well. A point $\bar{x} \in \mathbb{R}^n$ is called stationary

---

A. Fischer (✉)
Department of Mathematics, Technische Universität Dresden, 01062 Dresden, Germany
e-mail: Andreas.Fischer@tu-dresden.de

for problem (1), if $\bar{x}$ together with some multipliers $\mu \in \mathbb{R}^m$ and $\lambda \in \mathbb{R}^l$ satisfy the Karush–Kuhn–Tucker (KKT) system

$$\frac{\partial}{\partial x}\left(f(x) + \langle \mu, g(x) \rangle + \langle \lambda, h(x) \rangle \right) = 0, \quad h(x) = 0,$$
$$\mu \geq 0, \quad g(x) \leq 0, \quad \langle \mu, g(x) \rangle = 0. \tag{2}$$

The set of Lagrange multipliers associated to $\bar{x}$ is denoted by $\mathcal{M}(\bar{x})$. Our notation is close to that of Izmailov and Solodov. Moreover, references from their paper will appear as [IS #].

The construction of locally superlinearly convergent methods for problems where the set $\mathcal{M}(\bar{x})$ of Lagrange multipliers has infinitely many (nonisolated) elements was not an easy task in general. Note that classical conditions for superlinear convergence of methods for problem (1), like SQP-type methods in Robinson (1974) and Wilson (1963), or algorithms based on the reformulation of the KKT system (2) as a nonsmooth system of equations (Facchinei et al. 1998a; Qi and Jiang 1997), imply the local uniqueness of a KKT point. A review of some early works on the topic of dealing with nonunique multipliers can be found in Fischer (1999). For further references we just refer the reader to the invited paper by Izmailov and Solodov. Here, we would like to mention only two approaches that, in a wider sense, might be useful to deal with inequality constraints in the degenerate case. The first one (Ralph and Wright 1997, 2000) is based on an interior point technique for solving monotone variational inequalities and allows local superlinear convergence under conditions that do not imply the uniqueness of the multiplier. A related result for not necessarily convex optimization problems is given in Vicente and Wright (2002). The second approach is about identifying active inequality constraints at a stationary point $\bar{x}$ if problem (1) is degenerate. Such techniques may help to locally replace inequality constraints by equations. They were developed, applied, and extended in several contexts, for example see Dan et al. (2002), De Leone and Lazzari (2010), Facchinei et al. (1998b), Izmailov and Solodov (2008), Oberlin and Wright (2006) and Wright (2003). A basic ingredient of many identification techniques is a computable local error bound that holds in some neighborhood of the set of KKT points or, similarly, in the neighborhood of a certain particular KKT point. In the latter case, such an error bound is a function $\delta : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^l \to [0, \infty)$ which, for some KKT point $(\bar{x}, \bar{\mu}, \bar{\lambda})$ and some $\gamma \in (0, 1]$, $C > 0$, satisfies

$$\delta(x, \mu, \lambda)^\gamma \geq C \operatorname{dist}[(x, \mu, \lambda), \{\bar{x}\} \times \mathcal{M}(\bar{x})],$$

for all (x, μ, λ) in a neighborhood of $(\bar{x}, \bar{\mu}, \bar{\lambda})$. The locally correct identification of active inequalities according to Facchinei et al. (1998b) works for any $\gamma \in (0, 1]$. Interestingly, if an error bound with $\gamma = 1$ is available, it can be used to construct algorithms that converge locally superlinearly to a KKT point of a degenerate optimization problem. The first algorithm of this kind is the stabilized SQP method in [IS 51]. For more discussion and developments see Sect. 4 of the paper by Izmailov and Solodov. It is worth noting that the existence of an error bound around the KKT point $(\bar{x}, \bar{\mu}, \bar{\lambda})$ with $\gamma = 1$ is equivalent to requiring that (μ, λ) is noncritical [IS 10]. Thus,

for starting points sufficiently close to a noncritical multiplier, critical multipliers do not spoil those stabilized algorithms.

It is certainly not by chance that, in a more general sense, the use of error bounds plays a very important role for solving KKT systems or other problems that have nonisolated solutions. Such problems with nonisolated solutions may arise, for example, if we consider optimization problems with nonisolated primal solutions, systems of equations and inequalities, complementarity problems, multicriteria optimization, KKT conditions or Fritz John conditions (Dorsch et al. 2013) for generalized Nash equilibrium problems. In such cases a first task is to design appropriate algorithms that allow a certain stabilization or regularization (for example see Behling and Fischer 2012; Dong and Fischer 2006; Facchinei et al. 2014; Kanzow et al. 2004, [IS 13]), where we would like to explicitly mention the breakthrough in Yamashita and Fukushima (2001) by means of an appropriate regularization within the Levenberg–Marquardt method. In some cases, it is also not obvious how one can construct a computable local error bound (with $\gamma = 1$) or under which sufficient conditions such a bound exists, see Dreves et al. (2014), Facchinei et al. (1998b), Fischer and Shukla (2008), Izmailov and Solodov (2014) and [IS 11] for corresponding results.

In view of these works on problems with nonisolated solutions the question arises whether in those more general problems a criticality notion is useful or not. More importantly, do globally convergent algorithms for those problems exhibit the tendency to converge slowly to certain solutions. In analogy to the case of nonunique multipliers one could think that criticality of a solution of a more general problem means that, for this solution, there is no local error bound with $\gamma = 1$. However, before making a conjecture, it appears to be advisable to further clarify the criticality of multipliers and their influence on the global convergence behavior of algorithms. Also note that by another description of the feasible set the degeneracy may vanish.

I do not expect that one will be able to completely avoid the phenomenon that algorithms "like" to converge to critical multipliers. Therefore, as Izmailov and Solodov say, it might be helpful "to improve the efficiency in the case of convergence to critical multipliers". There are several attempts to accelerate the linear rate of convergence of algorithms applied to somehow singular problems, see Griewank (1980), Izmailov and Solodov (2002), Oberlin and Wright (2009) and references therein. Another direction of thinking is whether it can be useful to somehow avoid dealing with multipliers. Some of the restoration techniques (see Martínez and Pilotta 2000; Fischer and Friedlander 2010) do not use multipliers (with the effect of local slow convergence). Results in Birgin and Martínez (2005) and Izmailov et al. (2014) show that superlinear convergence is possible (by means of multipliers). Maybe a higher (but expensive) accuracy when dealing with feasibility could help.

References

Behling R, Fischer A (2012) A unified local convergence analysis of inexact constrained Levenberg–Marquardt methods. Optim Lett 6:927–940

Birgin EG, Martínez JM (2005) Local convergence of an inexact-restoration method and numerical experiments. J Optim Theory Appl 127:229–247

Dan H, Yamashita N, Fukushima M (2002) A superlinearly convergent algorithm for the monotone com-
    plementarity problem without uniqueness and nondegeneracy conditions. Math Oper Res 27:743–753
De Leone R, Lazzari C (2010) Error bounds for support vector machines with application to the identification
    of active constraints. Optim Methods Softw 25:185–202
Dong YD, Fischer A (2006) A framework for analyzing local convergence properties with applications to
    proximal-point algorithms. J Optim Theory Appl 131:53–68
Dorsch D, Jongen HT, Shikhman V (2013) On structure and computation of generalized Nash equilibria.
    SIAM J Optim 23:452–474
Dreves A, Facchinei F, Fischer A, Herrich M (2014) A new error bound result for generalized Nash equi-
    librium problems. Comput Optim Appl 59:63–84
Facchinei F, Fischer A, Herrich M (2014) An LP-Newton method: nonsmooth equations, KKT systems,
    and nonisolated solutions. Math Program 146:1–36
Facchinei F, Fischer A, Kanzow C (1998a) Regularity properties of a new equation reformulation of varia-
    tional inequalities. SIAM J Optim 8:850–869
Facchinei F, Fischer A, Kanzow C (1998b) On the accurate identification of active constraints. SIAM J
    Optim 9:14–32
Fernández D, Solodov MV (2010) Stabilized sequential quadratic programming for optimization and a
    stabilized Newton-type method for variational problems. Math Program 125:47–73
Fischer A (1999) Modified Wilson's method for nonlinear programs with nonunique multipliers. Math Oper
    Res 24:699–727
Fischer A, Friedlander A (2010) A new line search inexact restoration approach for nonlinear programming.
    Comput Optim Appl 46:333–346
Fischer A, Shukla PK (2008) A Levenberg–Marquardt algorithm for unconstrained multicriteria optimiza-
    tion. Oper Res Lett 36:643–646
Griewank AO (1980) Starlike domains of convergence for Newton's method at singularities. Numer Math
    35:95–111
Izmailov AF, Kurennoy AS, Solodov MV (2014) Some composite-step constrained optimization methods
    interpreted via the perturbed sequential quadratic programming framework. Optim Meth Softw. doi:10.
    1080/10556788.2014.924515
Izmailov AF, Solodov MV (2002) Superlinearly convergent algorithms for solving singular equations and
    smooth reformulations of complementarity problems. SIM J Optim 13:386–405
Izmailov AF, Solodov MV (2008) An active-set Newton method for mathematical programs with comple-
    mentarity constraints. SIAM J Optim 19:1003–1027
Izmailov AF, Solodov MV (2014) On error bounds and Newton-type methods for generalized Nash equi-
    librium problems. Comput Optim Appl 59:201–218
Kanzow C, Yamashita N, Fukushima M (2004) Levenberg–Marquardt methods with strong local con-
    vergence properties for solving nonlinear equations with convex constraints. J Comput Appl Math
    172:375–397
Martínez JM, Pilotta EA (2000) Inexact-restoration algorithm for constrained optimization. J Optim Theory
    Appl 104:135–163
Oberlin C, Wright SJ (2006) Active set identification in nonlinear programming. SIAM J Optim 17:577–605
Oberlin C, Wright SJ (2009) An accelerated Newton method for equations with semismooth Jacobians and
    nonlinear complementarity problems. Math Program 117:355–386
Qi L, Jiang H (1997) Semismooth Karush–Kuhn–Tucker equations and convergence analysis of Newton
    and quasi-Newton methods for solving these equations. Math Oper Res 22:301–325
Ralph D, Wright SJ (1997) Superlinear convergence of an interior point method for monotone variational
    inequalities. In: Michael CF, Pang JS (eds) Complementarity and variational problems: state of the
    art. SIAM, Philadelphia, pp 345–385
Ralph D, Wright SJ (2000) Superlinear convergence of an interior-point method despite dependent con-
    straints. Math Oper Res 25:179–194
Robinson SM (1974) Perturbed Kuhn–Tucker points and rates of convergence for a class of nonlinear
    programming algorithms. Math Program 7:1–16
Vicente LN, Wright SJ (2002) Local convergence of a primal–dual method for degenerate nonlinear pro-
    gramming. Comput Optim Appl 22:311–328
Wilson RB (1963) A simplicial algorithm for concave programming. Ph.D. thesis, Graduate School of
    Business Administration, Harvard University, Cambridge

Wright SJ (2003) Constraint identification and algorithm stabilization for degenerate nonlinear programs. Math Program 95:137–160

Yamashita N, Fukushima M (2001) On the rate of convergence of the Levenberg–Marquardt method. Comput 15(Suppl):239–249