Proceeding Series of the Brazilian Society of Applied and Computational Mathematics, v. 6, n. 1, 2018.

Trabalho apresentado no XXXVII CNMAC, S.J. dos Campos - SP, 2017.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

A note on the convergence of an augmented Lagrangian algorithm to second-order stationary points

Roberto Andreani[1]
Institute of Mathematics, Statistics and Scientific Computing, Unicamp, Campinas, SP, Brazil
Leonardo D. Secchin[2]
Department of Applied Mathematics, UFES, São Mateus, ES, Brazil

Abstract. Many algorithms that ensure second-order necessary optimality conditions were developed in the literature. To the best of our knownledge, none of them guarantee Strong Second-Order Necessary Condition (SSONC). Gould and Toint [5] showed that we do not expect SSONC in the barrier method. In this paper, we argue by an example that the same is true for the second-order augmented Lagrangian method introduced in [1]. This reinforces the Weak Second-Order Necessary Condition as the appropriate condition for the convergence analysis of second-order optimization algorithms.

Keywords. second-order optimality conditions, augmented Lagrangian methods, constrained optimization

1 Introduction

In this section, we consider the optimization problem

$$\min f(x) \quad \text{s.t.} \quad g(x) \leq 0 \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are twice continously differentiable functions. Let $\ell(x, \mu)$ be the Lagrangian function

$$\ell(x, \mu) = f(x) + \sum_{i=1}^{m} \mu_i g(x)$$

where $(x, \mu) \in \mathbb{R}^n \times \mathbb{R}^m_+$. It is well known that under a *constraint qualification*, a local minimizer x^* of (1) satisfies the first order Karush-Kuhn-Tucker (KKT) necessary conditions [4]. That is, there are multipliers μ^* such that

$$\begin{aligned} \nabla_x \ell(x^*, \mu^*) = \nabla f(x^*) + \sum_{i=1}^{m} \mu_i^* \nabla g(x^*) &= 0, \\ g(x^*) \leq 0, \ \ \mu^* &\geq 0 \\ \text{and} \quad \mu_i^* g_i(x^*) = 0, \ \ i &= 1, \ldots, m. \end{aligned}$$

[1]andreani@ime.unicamp.br
[2]leonardo.secchin@ufes.br

In this case, and under a suitable constraint qualification, x^* also satisfies the Strong Second-Order Necessary Condition (SSONC) [4]

$$d^T \nabla_{xx} \ell(x^*, \mu^*) d \geq 0 \quad \text{for all} \quad d \in C(x^*, \mu^*)$$

where

$$C(x^*, \mu^*) = \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{l} d^T \nabla g_i(x^*) = 0 \text{ for all } i \in I(x^*) \text{ such that } \mu_i^* > 0 \\ d^T \nabla g_i(x^*) \leq 0 \text{ for all } i \in I(x^*) \text{ such that } \mu_i^* = 0 \end{array} \right\}$$

and

$$I(x^*) = \{ j \mid g_j(x^*) = 0 \}$$

is the set of the indices of the active constraints at x^*. Reciprocally, if a KKT point x^* with associated multipliers μ^* conforms to the Second-Order Sufficient Condition (SOSC), i.e., if

$$d^T \nabla_{xx} \ell(x^*, \mu^*) d > 0 \quad \text{for all} \quad d \in C(x^*, \mu^*) \backslash \{0\},$$

then x^* is a strict local minimizer of (1) [4].

Many algorithms in the literature were developed in order to guarantee convergence to points that satisfy second-order necessary conditions (see for example [2] and references there in). Unfortunately, checking SSONC is an NP-hard [6] problem. Thus, it is common to deal with a less stringent condition, namely, the Weak Second-Order Necessary Condition (WSONC). A KKT point x^* with associated multipliers μ^* satisfies WSONC if $\nabla_{xx} \ell(x^*, \mu^*)$ is positive semidefinite over

$$C^W(x^*) = \left\{ d \in \mathbb{R}^n \mid d^T \nabla g_i(x^*) = 0 \text{ for all } i \in I(x^*) \right\}.$$

The subspace $C^W(x^*)$ is independent of multipliers and $C(x^*, \mu^*) \subset C^W(x^*)$ (this inclusion is strict; for example, in $\min_{x \leq 0} x^3$ at the origin). Checking WSONC consists of solving a quadratic programming problem, a computationaly feasible task. We note that when we only have equality constraints, the analogues of C and C^W coincide, and no further analysis is necessary. This is the motivation to consider only inequality constraints in (1).

One of the most important class of optimization algorithms to solve (1) are the augmented Lagrangian methods. Auslander [3] showed that the classical barrier and the pure quadratic penalty methods converge to points that satisfy WSONC if its iterates fulfills SOSC for the subproblems. Gould and Toint [5] provided an example where the limit point of the classical barrier method does not satisfy SSONC (naturally, only WSONC). That is, we can not expect convergence to SSONC points with the barrier method. Thus, one question is whether the same occurs with quadratic penalty-like methods. We are particularly interested in the second-order augmented Lagrangian method developed in [2], named Algencan-second. In this method, each iterate satisfies approximately a second-order condition for the Lagrangian subproblem. The authors showed that under a suitable constraint qualification (which is weaker than the well known Linear Independence Constraint Qualification – LICQ), the feasible limit points of Algencan-second fulfill WSONC.

This papers is organized as follows. In Section 2 we briefly present ALGENCAN-SECOND algorithm for (1). In Section 3 we give an example where SOSC for the subproblem is valid at every iterate of ALGENCAN-SECOND method but its limit point does not satisfy SSONC. Finally, conclusions are given in Section 4.

Notation. The symbol $\|\cdot\|_\infty$ will denote the sup-norm. If $z \in \mathbb{R}^n$, the components of z_+ are defined by $(z_+)_i = \max\{0, z_i\}$, $i = 1, \dots, n$. Also, $\mathbb{R}^q_+ = \{z \in \mathbb{R}^q \mid z_i \geq 0, i = 1, \dots, q\}$. We denote by $\lambda_0(A)$ the smallest eigenvalue of a symmetric matrix A.

2 Second-order augmented Lagrangian method

The Powell-Hestenes-Rockafellar (PHR) augmented Lagrangian associated to the problem (1) is defined by

$$L_\rho(x, \mu) = f(x) + \frac{\rho}{2} \sum_{i=1}^{m} \left[\left(g_i(x) + \frac{\mu_i}{\rho} \right)_+ \right]^2,$$

where $(x, \mu) \in R^n \times \mathbb{R}^m_+$ and $\rho > 0$. We define, for each $x \in \mathbb{R}^n$ and $\varepsilon > 0$, the *approximate ε-Hessian* of L_ρ (with respect to x) as

$$\nabla^2_\varepsilon L_\rho(x, \mu) = \nabla^2 f(x) + \sum_{i=1}^{m} (\mu_i + \rho g_i(x))_+ \nabla^2 g_i(x) + \rho \sum_{i \in I_\varepsilon(x,\mu,\rho)} \nabla g_i(x) \nabla g_i(x)^T$$

where

$$I_\varepsilon(x, \mu, \rho) = \left\{ j \;\middle|\; \frac{1}{\sqrt{\rho}} (\mu_j + \rho g_j(x)) \geq -\varepsilon \right\}.$$

Observe that $\nabla^2_0 L_\rho(x, \mu)$ is the true Hessian of L_ρ where it exists. At these points the eigenvalues of $\nabla^2_\varepsilon L_\rho$ give upper bounds to the eigenvalues of the true Hessian. Furthermore, if the true Hessian of L_ρ is semidefinite positive then $\nabla^2_\varepsilon L_\rho$ is also semidefinite positive.

We resume ALGENCAN-SECOND method for the problem (1) in Algorithm 1 below.

In order to guarantee the well-definiteness of optimization algorithms, a common assumption is that the sequence of iterates belongs to a compact set, frequently ensured by box-constraints on the original problem. ALGENCAN-SECOND algorithm was designed to handle box-constraints. However, this requires a more sophisticated exposure that is unnecessary for our purposes (for more details, see [2]). Hence, we simply assume the following assumption that guarantees the well-definiteness of the algorithm presented here.

Assumption A1. The Lagrangian subproblem (2) always have a global solution, and the sequence $\{x^k\}$ generated by Algorithm 1 belongs to a compact set.

The iterate x^k in Step 2 of Algorithm 1 may be computed by the GENCAN-SECOND [2] algorithm, a box-constrained solver based on an active-set strategy and in spectral projected gradient steps that is able to deal with directions of negative curvature.

Proceeding Series of the Brazilian Society of Applied and Computational Mathematics, v. 6, n. 1, 2018.

Algorithm 1 ALGENCAN-SECOND for problem (1)

Let $\mu_{\max} > 0$, $\gamma > 1$ and $0 < \tau < 1$. Let $\{\varepsilon_k\}$ be a sequence of positive scalars such that $\lim_{k\to\infty} \varepsilon_k = 0$. Let $\mu_i^1 \in [0, \mu_{\max}]$, $i = 1, \ldots, m$, and $\rho_1 > 0$. Initialize $k \leftarrow 1$.

Step 1. Find an approximate minimizer x^k of the problem

$$\min_x L_{\rho_k}(x, \mu^k). \tag{2}$$

The conditions for x^k are

$$\|\nabla_x L_{\rho_k}(x^k, \mu^k)\|_\infty \le \varepsilon_k \tag{3}$$

and

$$\lambda_0 \left(\nabla^2_{\varepsilon_k} L_{\rho_k}(x^k, \mu^k)\right) \ge -\varepsilon_k. \tag{4}$$

Step 2. Define

$$V_i^k = \max\left\{g_i(x^k), -\mu_i^k/\rho_k\right\}, \quad i = 1, \ldots, m.$$

If $k > 1$ and $\max\{\|g(x^k)\|_\infty, \|V_i^k\|_\infty\} \le \tau \max\{\|g(x^{k-1})\|_\infty, \|V_i^{k-1}\|_\infty\}$ define $\rho_{k+1} = \rho_k$. Otherwise, define $\rho_{k+1} = \gamma\rho_k$.

Step 3. Compute $\mu_i^{k+1} \in [0, \mu_{\max}]$, $i = 1, \ldots, m$. Take $k \leftarrow k + 1$ and go to the Step 1.

Remark 2.1. *The tolerances ε_k may be different in (3) and (4), including the precision for the approximate Hessian. For the sake of simplicity, we presented a simplified version with all equal tolerances.*

Remark 2.2. *We can compute μ_i^{k+1} in the Step 3 of Algorithm 1 projecting $\mu_i^k + \rho_k g_i(x^k)$ into $[0, \mu_{\max}]$. This is the way employed in the* ALGENCAN *[1] implementation provided by TANGO project (*`www.ime.usp.br/~egbirgin/tango`*).*

As we already mentioned, feasible limit points of Algorithm 1 fulfill WSONC under a suitable constraint qualification, for example, LICQ.

3 An example

Let us consider the quadratic problem

$$\min \frac{1}{2} x^T G x \quad \text{s.t.} \quad x \in \mathbb{R}^n_+ \tag{5}$$

where G is the symmetric matrix

$$I - \alpha \frac{zz^T}{z^T z},$$

$n \ge 2$, $z = e - ne_1$, e_i is the i-th canonical vector of $\mathbb{R}^n$, e is the vector of all entries one and

$$\alpha > \frac{n}{n-1}.$$

This problem is very similar to that given in [5]. We observe that LICQ is valid in the entire space $\mathbb{R}^n$.

First, we will argue that the origin does not satisfy SSONC. The KKT conditions for (5) are

$$Gx - \mu = 0, \quad x^T\mu = 0, \quad x \geq 0 \quad \text{and} \quad \mu \geq 0.$$

At $x = 0$, the unique corresponding multiplier is $\mu = 0$. We have $e_1^T z = 1 - n$, $z^T z = n(n-1)$ and thus

$$e_1^T \nabla^2 l(0,0) e_1 = e_1^T G e_1 = 1 - \alpha \frac{(e_1^T z)^2}{z^T z} = 1 - \alpha \frac{(1-n)^2}{n(n-1)} = 1 - \alpha \frac{(n-1)}{n} < 0.$$

As $e_1 \in C(0,0) = \{d \mid -e_i^T d \leq 0, i = 1, \ldots, n\}$, the origin does not satisfy SSONC.

From now on, we will analyze the application of Algorithm 1 on (5). The PHR augmented Lagrangian takes the form

$$L_\rho(x,\mu) = \frac{1}{2} x^T G x + \frac{\rho}{2} \sum_{i=1}^{n} \left(-x_i + \frac{\mu_i}{\rho} \right)_+^2 .$$

Assumption A2 $\mu_i/\rho - x_i > 0$, $i = 1, \ldots, n$.

If A2 is valid, then $\nabla_x L_\rho(x,\mu) = Gx - \mu + \rho x$, $I_\varepsilon(x,\mu,\rho) = \{1, \ldots, m\}$ and

$$\nabla_\varepsilon^2 L_\rho(x,\mu) = G + \rho \sum_{i=1}^{n} e_i e_i^T = G + \rho I. \tag{6}$$

In this case, the matrix $\nabla_\varepsilon^2 L_\rho(x,\mu)$ will be positive definite for all $\rho > 0$ sufficiently large. Furthermore, for these ρ we will have

$$\nabla_x L_\rho(x,\mu) = Gx - \mu + \rho x = 0 \quad \Leftrightarrow \quad x = (G + \rho I)^{-1} \mu. \tag{7}$$

Without loss of generality, we can assume that $\rho > \alpha - 1$. Thus, by Shermann-Morrison formula we will obtain

$$(G + \rho I)^{-1} = \left[(1+\rho) I - \frac{\alpha z}{z^T z} z^T \right]^{-1} = \frac{1}{1+\rho} I + \frac{\alpha z z^T}{(1+\rho)(1+\rho-\alpha) z^T z}$$

and thus

$$x = \left(\frac{1}{1+\rho} \right) \mu + \left(\frac{\alpha z^T \mu}{(1+\rho)(1+\rho-\alpha) z^T z} \right) z. \tag{8}$$

Assumption A3 $\mu_i = \mu_1$ for all i, that is, $\mu = \mu_1 e$.

A3 implies $z^T\mu = \mu_1(e^T e - n e_1^T e) = 0$, and from (8) we have

$$x = \left(\frac{1}{1+\rho} \right) \mu. \tag{9}$$

We initialize Algorithm 1 with

$$\mu^0 = \mu_1^0 e > 0, \quad \rho_0 > 0, \quad x^0 = \frac{1}{1+\rho_0}\mu^0,$$

and ρ_0 large enough to ensure $\rho_0 > \alpha - 1$ and the positive definiteness of $G + \rho_0 I$. We have

$$\frac{\mu_i^0}{\rho_0} - x_i^0 = \left(\frac{1}{\rho_0} - \frac{1}{1+\rho_0}\right)\mu_i^0 > 0, \quad i = 1, \dots, n,$$

that is, A2 is satisfied at the initial point. Thus, by (7) we obtain $\nabla_x L_{\rho_0}(x^0, \mu^0) = 0$ and by (6), x^0 satisfies SOSC for the Lagrangian subproblem. In the following iterations,

- $\mu_i^k - \rho_k x_i^k = [1 - \rho_k/(1+\rho_k)]\mu_i^k > 0$ and thus $\mu_i^{k+1} > 0$ (see Remark 2.2). Inductively, the expression (9) is valid at iteration k because assumptions A2 and A3 are valid in the previous iteration and $\rho_k \geq \rho_0 > \alpha - 1$. Consequently, $\nabla_x L_{\rho_k}(x^k, \mu^k) = 0$ by (7);
- as A2 is valid and $\rho_k \geq \rho_0$, $\nabla_\varepsilon^2 L_{\rho_k}(x^k, \mu^k) = G + \rho_k I \succ 0$. That is, SOSC for the Lagrangian subproblem (2) is valid in *all* iterations of Algorithm 1;
- all feasible points satisfy LICQ, and thus we conclude that $\rho_k \to \infty$. Otherwise Algorithm 1 would converge to a limit point $x^* = \eta e$ with $\eta > 0$, which is not a KKT point. But this contradicts the convergence theory of augmented Lagrangian methods [1, 2].

By the above discussion, we conclude that the sequence

$$\left\{ x^{k+1} = \frac{1}{1+\rho_k}\mu^k \right\}$$

generated by Algorithm 1 converges to the origin, which does not satisfy SSONC although *all* the iterates fulfill SOSC for the subproblems.

Remark 3.1. $x \geq 0$ *may be treated as box-constraints in* ALGENCAN-SECOND, *and then they will not be penalized. We argue that the example is still valid in this case. In fact, the first order condition (3) will be established in an analogous way to deal with box-constraints (see [1]). Also, the second-order condition can only differs from (4) by changing some rows/columns of* $\nabla_\varepsilon^2 L_{\rho_k}(x^k, \mu^k)$ *to respective rows/columns of identity matrix (see [2]). Therefore, no new negative eigenvalues will appear.*

4 Conclusions

We considered a PHR augmented Lagrangian method that, under mild assumptions, converges to points satisfying the Weak Second-Order Necessary optimality Condition (WSONC). We showed by an example that even this method can fail to obtain limits that conforms to Strong Second-Order Necessary optimality Condition (SSONC). This is an analogous result to the one obtained previously for the classical barrier method [5], and reinforce us to believe that there is no method that ensures SSONC. On the other hand, this also reinforces that WSONC is the appropriate condition for the convergence analysis of second-order optimization algorithms.

References

[1] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. On augmented lagrangian methods with general lower-level constraints, *SIAM Journal on Optimization*, 18:1286–1309, 2007. DOI:10.1137/060654797.

[2] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. Second-order negative-curvature methods for box-constrained and general constrained optimization, *Computational Optimization and Applications*, 45:209–236, 1999. DOI:10.1007/s10589-009-9240-y.

[3] A. Auslender. Penalty methods for computing points that satisfy second order necessary conditions, *Mathematical Programming*, 17:229–238, 1979. DOI:10.1007/BF01588245.

[4] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming: Theory and Algorithms*. 3rd ed. John Wiley & Sons, New Jersey, 2006.

[5] N. I. M. Gould and P. L. Toint. A note on the convergence of barrier algorithms to second-order necessary points, *Mathematical Programming*, 85:433–438, 1999. DOI:10.1007/s10107980030a.

[6] K. G. Murty and S. N. Kabadi. Some NP-complete problems in quadratic and nonlinear programming, *Mathematical Programming*, 39:117–129, 1987. DOI:10.1007/BF02592948.

