On an Approximation of the Hessian of the Lagrangian

Florian Jarre
Institut für Mathematik, Universität Düsseldorf
Universitätsstraße 1, D–40225 Düsseldorf, Germany
e-mail: jarre@opt.uni-duesseldorf.de

Abstract. In the context of SQP methods or, more recently, of sequential semidefinite programming methods, it is common practice to construct a positive semidefinite approximation of the Hessian of the Lagrangian. The Hessian of the augmented Lagrangian is a suitable approximation as it maintains local superlinear convergence under appropriate assumptions. In this note we give a simple example that the orthogonal projection of the Hessian of the Lagrangian onto the cone of semidefinite matrices may lead to arbitrarily slow local convergence, and is thus not a suitable approximation.

Key words. SQP method, Lagrangian, approximate Hessian.

1 Introduction

The subproblems arising in SQP methods or in sequential semidefinite programming methods (SSP methods), use a quadratic objective function whose Hessian ideally coincides with the Hessian of the Lagrangian. However, when the Hessian of the Lagrangian is not positive definite, the solution of the SSP subproblems is computationally unattractive. An effective remedy used with SQP methods is to replace the Hessian of the Lagrangian by the Hessian of the augmented Lagrangian. Under suitable conditions, this matrix is positive definite and this substitution maintains local superlinear convergence of the SQP method, [1, 11]. The augmented Lagrangian depends on certain penalty parameters, a good choice of which is not always obvious. Alternatively, the Hessian of the Lagrangian can be replaced by a reduced Hessian that approximates the Hessian on the null space of the active gradients. Again, the reduced Hessian is positive definite under suitable assumptions. Here, the identification of the active constraints is a critical issue.

Another popular and computationally cheap approach for generating a positive semidefinite approximation to the Hessian of the Lagrangian is to apply a damped BFGS update, [8, 9].

For the recent SSP methods [4, 3, 5], the cost for solving the SSP subproblems is rather high—it is the solution of a linear semidefinite program—so that a cheap approximation to the Hessian is less crucial than the goal of finding the best possible approximation.

An approximation that does not depend on some unknown parameters (as the augmented Lagrangian or the reduced Hessian do) and that is closest possible to the Hessian of the Lagrangian, is to use the projection onto the cone of positive semidefinite matrices. Due to the computational cost, the use of this approximation is prohibitive in standard SQP methods. However, for more complicated SSP subproblems, this approximation might be affordable.

The aim of this short note is to show that, regardless of the computational cost, this approximation is not suitable in the context of SQP methods or SSP methods.

2 A positive semidefinite approximation

Consider the following nonlinear optimization problem

$$(P) \qquad \text{minimize} \;\; f(x) \quad \text{subject to} \quad x \in I\!R^n : \; f_i(x) \le 0 \text{ for } 1 \le i \le m$$

with twice continuously differentiable functions $f,\ f_i\ \ (1 \le i \le m)$. By defining an $m \times m$ diagonal matrix with diagonal entries $f_i(x)$, it is evident that problem (P) is a special case of a nonlinear semidefinite program. The SSP subproblem of this nonlinear semidefinite program coincides with the SQP subproblem for (P). In the following we therefore restrict ourselves to the simpler problem (P).

Let the Lagrangian $L: \ I\!R^{n+m} \to I\!R$ be defined by

$$L(x,y) := f(x) + \sum_{i=1}^{m} y_i f_i(x).$$

The Hessian of the Lagrangian is denoted by

$$H(x,y) := \nabla_x^2 L(x,y).$$

For brevity we will write shortly $H := H(x,y)$. For a given vector x, and a given matrix H the SQP subproblem is given by
(QP)

$$\underset{\Delta x \in I\!R^n:}{\text{minimize}} \;\; \nabla f(x)^T \Delta x + \frac{1}{2} \Delta x^T H \Delta x \quad \text{subject to} \quad f_i(x) + \nabla f_i(x)^T \Delta x \le 0 \text{ for } 1 \le i \le m.$$

Replacing H in (QP) by a semidefinite matrix $\hat{H}$ yields a modified convex quadratic subproblem.

The symmetric matrix H possesses the eigenvalue decomposition

$$H = UDU^T$$

with a diagonal matrix D and a unitary matrix U. Let D^+ be the matrix obtained from D by replacing the negative entries of D with zeros. The matrix

$$H^+ := \underset{\tilde{H} \succeq 0}{\text{argmin}} \;\; \|\tilde{H} - H\|_F^2$$

is the projection of H onto the cone of positive semidefinite matrices; where the norm $\|.\|_F$ denotes the Frobenius norm generated by the trace inner product on the space of square matrices. Since U is an orthogonal matrix, it follows that H^+ is given by

$$H^+ = UD^+U^T.$$

Thus, the eigenbases of H and H^+ coincide, and, with respect to the Frobenius norm, H^+ is the best possible semidefinite approximation to H.

To illustrate the local convergence properties of the modified SQP-method based on some modified Hessian $\hat{H}$, let us assume, for simplicity, that the constraints f_i are linear and that the active constraints are identified and satisfied with equality at an early stage of the SQP method. Then, local superlinear convergence may be lost if, and only if, $\hat{H}$ provides wrong curvature information on the null space of the Jacobian of the active constraints.

What is needed for the modified SQP subproblem is some positive definite approximation of H that coincides with H on the null space of the Jacobian of the active constraints (because this is the space in which the iterates are still "moving"). Since the eigenvectors to negative eigenvalues of H are not necessarily orthogonal to this null space, the approximation H^+ typically does not coincide with H on this null space, and is thus not suitable for the modified SQP subproblem.

In contrast to the projection H^+ of H, the Hessian of the augmented Lagrangian – under certain conditions – does provide a suitable approximation in the context of SQP methods. For a real number t let $t^+ := \max\{0, t\}$. For a given real parameter $r > 0$, the augmented Lagrangian $\Lambda : I\!R^{n+m} \to I\!R$ (see e.g. [6, 7, 10, 2]) is given by

$$\Lambda(x, y) := f(x) + \frac{r}{2} \sum_{i=1}^{m} \left(\left(\left(f_i(x) + \frac{y_i}{r} \right)^+ \right)^2 - \left(\frac{y_i}{r} \right)^2 \right).$$

Near a strictly complementary solution of (P) the augmented Lagrangian is twice continuously differentiable, and if the second order sufficient conditions for optimality hold and r is sufficiently large,

$$H_a(x, y) := \nabla_x^2 \Lambda(x, y)$$

is positive definite at the local minimizer and can be considered as a positive definite approximation to H. Since H and H_a coincide on the null space of the Jacobian of the active constraints, the Hessian of the augmented Lagrangian is a suitable approximation in the context of SQP methods.

The possibly very poor quality of the modified SQP-subproblem based on the projection H^+ will be illustrated with a simple example and will be compared with the subproblem based on the matrix H_a:

3 An example

For $\mu > 0$ we consider the problem

$$(P_\mu) \qquad \text{minimize} \;\; f_\mu(x) \quad \text{subject to} \quad x \in I\!R^2 : \; x \geq 0,$$

where

$$\begin{aligned} f_\mu(x) &:= \mu^2 x_1^3 + 4\mu x_1 x_2 + 2x_2^2 + (\mu^2 - 4\mu + 1)x_1 - 4x_2 + 2 \\ &= x_1(1 + \mu^2(x_1 - 1)^2) + 2(\mu x_1 + x_2 - 1)^2. \end{aligned}$$

From the second representation of f_μ it becomes obvious that the global minimizer of (P_μ) is the point $\bar{x} = (0,\ 1)^T$.

Let an approximate local minimizer $x_\epsilon := (0,\ 1+\epsilon)^T$ with $|\epsilon| < 1$ be given. The derivatives of f_μ are then given by

$$\nabla f_\mu(x_\epsilon) = \begin{pmatrix} \mu^2 + 1 + 4\mu\epsilon \\ 4\epsilon \end{pmatrix} \quad \text{and} \quad \nabla^2 f_\mu(x_\epsilon) = \begin{pmatrix} 0 & 4\mu \\ 4\mu & 4 \end{pmatrix}.$$

As the constraints are linear, the Hessian of f_μ and of the Lagrangian

$$L(x,y) = f_\mu(x) + y_1 x_1 + y_2 x_2$$

coincide, $H = \nabla^2 f_\mu(x_\epsilon)$. For simplicity of the presentation we now assume $\mu \gg 1$. For $y_1 > 0$ and sufficiently small $|y_2|$, the Hessian $H_a(x_\epsilon, y)$ of the augmented Lagrangian function with parameter $r > 0$ and the projected Hessian H^+ are then given by

$$H_a(x_\epsilon, y) = \begin{pmatrix} r & 4\mu \\ 4\mu & 4 \end{pmatrix} \quad \text{and} \quad H^+ \approx \begin{pmatrix} 2\mu & 2\mu+1 \\ 2\mu+1 & 2\mu+2 \end{pmatrix}.$$

For sufficiently large r ($\geq 4\mu^2$) the matrix $H_a(x_\epsilon, y)$ is positive semidefinite. The "ideal" SQP subproblem

$$\text{minimize} \quad \begin{pmatrix} \mu^2 + 1 + 4\mu\epsilon \\ 4\epsilon \end{pmatrix}^T \Delta x + \frac{1}{2}\Delta x^T H \Delta x \quad \text{subject to} \quad \Delta x \in I\!R^2 : \; x_\epsilon + \Delta x \geq 0$$

has a nonconvex objective function. When replacing H with $H_a(x_\epsilon, y)$, the solution of the resulting SQP subproblem is given by

$$\Delta x = (0, \; -\epsilon)^T.$$

Hence, $x + \Delta x$ also solves (P_μ). When replacing H with H^+ instead, the resulting search step

$$\Delta x \approx (0, \; -\frac{2\epsilon}{\mu+1})^T$$

is too short by a factor of approximately $\frac{2}{\mu+1}$. For $\mu \to \infty$ the search step is arbitrarily bad, even for tiny $|\epsilon| > 0$. Thus, the SQP method with H replaced by H^+ is linearly convergent with rate $\frac{2}{\mu+1} \ll 1$. Combining several problems of the form (P_μ) with different values of μ to form a single high dimensional, separable problem (P), this simple example can be extended in such a way, that the search step can be a poor direction as well, and a line search will not help much.

4 Acknowledgement

The author is thankful to an anonymous referee who helped to improve the presentation of the paper.

References

[1] Boggs, P.T., Tolle, J.W. (1995): Sequential quadratic programming. Acta Numer. **4**, 1–51

[2] Conn A.R., Gould N.I.M., Toint Ph.L., (1991): A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. SIAM J. Numerical Anal., **28**, 545–572

[3] Correa, R., Ramírez C., H. (2002): A global algorithm for nonlinear semidefinite programming. Research Report 4672, INRIA, Rocquencourt, France

[4] Fares, B., Noll, D., Apkarian, P. (2002): Robust control via sequential semidefinite programming. SIAM J. Control Optim. **40**, 1791–1820

[5] Freund, R.W., Jarre, F. (2003): A sensitivity analysis and a convergence result for a sequential semidefinite programming method. Numerical Analysis Manuscript No. 03–4–09, Bell Laboratories, Murray Hill, New Jersey

[6] Hestenes, M.R. (1969): Multiplyer and gradient methods. J. Opt. Theory Appl. **4**, 303–320.

[7] Powell, M.J.D. (1969): A method for nonlinear constraints in minimization problems. In Fletcher, R. (ed): Optimization, Academic Press, New York, 283–298.

[8] Powell, M.J.D. (1978): A fast algorithm for nonlinearly constrained optimization calculations. Lecture Notes in Mathematics 630, Springer-Verlag, Berlin, 144–157

[9] Powell, M.J.D. (1978): The convergence of variable metric methods for nonlinearly constrained optimzation calculations. In: O.L. Mangasarian, R.R. Meyer, S.M. Robinson eds, Nonlinear Programming, **3**. Academic Press, New York, 27–63.

[10] Rockafellar, R.T. (1973): The multiplyer method of Hestenes and Powell applied to convex Programming. J. Opt. Theory Appl. **12**, 555–562.

[11] Schittkowski, K. (1981): The nonlinear programming method of Wilson, Han, and Powell with an augmented Lagrangian type line search function, parts 1 and 2, Numer. Math. **38**, 83–127