Summer 6-1-1984

An Example of Irregular Convergence in Some Constrained Optimization Methods that Use the Projected Hessian ; CU-CS-268-84

Richard H. Byrd
University of Colorado Boulder

AN EXAMPLE OF IRREGULAR CONVERGENCE IN SOME CONSTRAINED OPTIMIZATION METHODS THAT USE THE PROJECTED HESSIAN

by

Richard H. Byrd *

CU-CS-268-84 June, 1984

Department of Computer Science, University of Colordo, Boulder, Colorado 80309

* This research was supported by National Science Foundation Grant MCS-8115475.

ANY OPINIONS, FINDINGS, AND CONCLUSIONS OR RECOMMENDATIONS EXPRESSED IN THIS PUBLICATION ARE THOSE OF THE AUTHOR AND DO NOT NECESSARILY REFLECT THE VIEWS OF THE NATIONAL SCIENCE FOUNDATION.

ABSTRACT

In this paper we give examples illustrating the behavior of the Coleman- Conn horizontal vertical method and of successive quadratic programming with a Hessian approximation exact on the tangent space of the constraints. One example shows that these methods in general are not one-step superlinearly convergent.

An Example of Irregular Convergence in Some Constrained Optimization Methods That Use the Projected Hessian

Many methods for solution of optimization problems with nonlinear constraints make use of the Hessian of the Lagrangian to obtain fast local convergence. Several of these methods including those of Coleman and Conn [1980] make use of approximations to the reduced Hessian. In addition Powell [1978] analyzes a method where the Hessian approximation is accurate only on the null space of the constraints derivatives. These authors show that under reasonable conditions these methods are two-step superlinearly convergent; i.e. the sequence consisting of every other iterate is superlinearly convergent.

An obvious question is whether these methods are one-step superlinearly convergent. In this paper we give an example showing that neither the method of Coleman and Conn, nor successive quadratic programming with an accurate Hessian on that subspace is superlinearly convergent.

In order to do this we will now describe these two methods applied to a problem of the form

$$\underset{x \in R^n}{\text{minimize}} \quad f(x) \tag{1}$$

$$\text{subject to} \quad c(x)=0$$

where f is a real-valued function on R^n and c maps R^n to R^m. We assume both functions are twice differentiable.

The method of successive quadratic programming (SQP), at an iterate x_k, has the form:

Given x_k let d_k be the solution to

$$\text{minimize} \nabla f(x_k)^T d + \tfrac{1}{2} d^T B_k d$$

$$\text{subject to} \nabla c(x_k)^T d = -c(x_k)$$

Then let

$$x_{k+1} = x_k + d_k.$$

Here B_k is an nxn matrix approximating the Hessian of the Lagrangian,

$$\nabla^2 L(x_k, \lambda_k) = \nabla^2 f(x_k) + \Sigma \lambda_i \nabla^2 c_i(x_k).$$

It should be noted that if B_k is the exact Hessian and if λ is the vector of Lagrange multipliers to the quadratic program at x_{k-1} then this is just Newton's method on the Kuhn-Tucker conditions for problem (1).

To discuss convergence of this method, let the orthogonal projection matrix onto the null space of the constraint derivatives be denoted by

$$P_k = I - \nabla c(x_k)(\nabla c(x_k)^T \nabla c(x_k))^{-1} \nabla c(x_k)^T,$$

and let (x^*, λ_*) denote the Kuhn-Tucker pair for problem (1). It has been shown by Boggs, Tolle, and Wang [1982] and, without the assumption of linear convergence, by Fontecilla, Steihaug, and Tapia [1983] that the sequence x_k generated by SQP is Q-superlinearly convergent if and only if

$$\frac{\| P_k(B_k - \nabla^2 L(x_*, \lambda_*))(x_{k+1} - x_k) \|}{\| (x_{k+1} - x_k) \|} \tag{3}$$

converges to zero.

In the case when we only know $P_k \nabla^2 L(x_k, \lambda_k) P_k$ accurately, that is we have second derivative information only on the null space of the constraint derivatives, this result is weakened. Powell [1978] shows that, under the assumption of convergence, if the condition

$$\frac{\| P_k(Bk - \nabla^2 L(x_*, \lambda_*)) P_k(x_{k+1} - x_k) \|}{\| (x_{k+1} - x_k) \|} \to 0 \tag{4}$$

holds then the sequence is two-step superlinearly convergent; that is

$$\frac{\|x_{k+1}-x^*\|}{\|x_{k-1}-x^*\|}\to 0.$$

We now consider the horizontal-vertical algorithm of Coleman and Conn. Following the usual notation let Z_k be a matrix of orthogonal columns spanning the null space of $\nabla c(x_k)^T$, and Y_k be a matrix of orthogonal columns spanning the column space of $\nabla c(x_k)$. Of course Y_k and Z_k are not uniquely determined by $\nabla c(x_k)$, but they may be easily computed from a QR factorization of $\nabla c(x_k)$.

A single iteration of the method is as follows.

$$h_k = -Z_k M_k^{-1} Z_k{}^T \nabla f(x_k)$$

$$v_k = -\nabla c(x_k)(\nabla c(x_k)^T \nabla c(x_k))^{-1} c(x_k+h_k)$$

$$x_{k+1} = x_k + h_k + v_k.$$

Here M_k is an approximation to $Z_k{}^T \nabla^2 L(x_k,\lambda_k) Z_k$.

Note that h_k is a solution of the homogeneous equality constrained quadratic program

$$\text{minimize} \quad \nabla f(x_k)^T h + \tfrac{1}{2} h^T Z_k M_k Z_k{}^T h$$

$$\text{subject to} \quad \nabla c(x_k)^T h = 0.$$

Note also that the constraints are evaluated at two points, x_k and x_k+h_k. The method is intended to be used in this form only near the solution. A more detailed description and motivation may be found in Coleman and Conn [1980]. If we compare the step generated by this algorithm with the step produced by SQP we see two differences. One is the extra evaluation of the constraints in the Coleman-Conn algorithm; the other is in the Hessian approximation used. In SQP any positive definite nxn matrix may be used. In the Coleman-Conn method only the reduced Hessian is approximated. Thus if were not for the extra constraint evaluation, Coleman and Conn's algorithm would be a special case of SQP with a Hessian approximation satisfying $Y_k^T B_k Z_k = 0$.

Convergence results are similar to that of Powell for SQP. Coleman and Conn [1980], [1982] show that if $M_k = Z_k{}^T \nabla^2 L(x_k, \lambda_k) Z_k$ or even if

$$\frac{\|(M_k - \nabla^2 L(x_*, \lambda_*)) Z_k{}^T (x_{k+1} - x_k)\|}{\|(x_{k+1} - x_k\|} \to 0$$

the method converges two-step superlinearly.

In spite of these theoretical results, there have been no examples or reports of computational experiments that indicate these methods are not actually one-step superlinearly convergent. It is also not clear to what extent being two-step superlinear as opposed to one-step superlinearly implies that these methods are therefore slower.

We now give examples showing that convergence for these methods is in general no better than two-step superlinear. We first consider the horizontal-vertical method of Coleman and Conn applied to an example.

To avoid confusion among iterate numbers, components, and powers in the following example we will denote the components of a vector by subscripts in parentheses i.e. $x = (x_{(1)}, x_{(2)})^T$. Iterate numbers will be denoted by plain subscripts and powers by superscripts. We first consider a problem for which we believe the behavior of the method is typical of what happens in most cases:

Example 1.

$$\text{minimize } f(x) = \tfrac{1}{2} x_{(1)}^2 - \alpha x_{(1)} x_{(2)} + \tfrac{1}{2} x_{(2)}^2$$

$$\text{subject to } c(x) = \frac{1}{(x_{(2)} - 2)^2} - 1 = 0$$

where α is some constant. Note that the constraint is equivalent to requiring $x_{(2)} = 1$, and that there is a strong local minimizer at $x = (\alpha, 1)$. We will apply the Coleman-Conn horizontal vertical method to this problem using exact reduced Hessian information, i.e. $M_k = Z_k^T \nabla^2 L(x_k, \lambda_k) Z_k$ which in this case, since c is a function of $x_{(2)}$ only, is

$$\frac{\partial^2 f}{\partial x_{(1)}^2} = 1.$$

Taking a horizontal step from a point of the form

$$x_k = (\alpha + \delta, 1 + \epsilon)$$

is equivalent to minimizing f subject to $x_{(2)}$ held constant, so

$$x_k + h_k = (\alpha(1+\epsilon), 1+\epsilon). \qquad (5)$$

A vertical step from any point is just a Newton step in $x_{(2)}$ with $x_{(1)}$ held fixed, so

$$x_{k+1} = x_k + h_k + v_k = (\alpha(1+\epsilon), 1+\epsilon^2). \qquad (6)$$

By the same logic, on taking the next step we have

$$x_{k+1} + h_{k+1} = (\alpha(1+\epsilon^2), 1+\epsilon^2),$$

and

$$x_{k+2} = x_{k+1} + h_{k+1} + v_{k+1} = (\alpha(1+\epsilon^2), 1+\epsilon^4).$$

Note that, for x_k an arbitrary point, the subsequent step is linear unless the error component normal to the constraint manifold (ϵ) is much smaller than the error component tangent to the constraint manifold (δ). However after one step the tangent component at x_{k+1} is dominant and the next step and all subsequent steps are quadratic.

This effect of the iterates taking a path approximately tangent to the constraints is well known in practice. What we do next is to modify this example so that at every other horizontal step the tangent component of the error is reduced to zero, and the other horizontal steps and all vertical steps are identical to those in Example 1.

Example 2.

$$\text{minimize } f(x) = \tfrac{1}{2} x_{(1)}^2 - \alpha x_{(1)} x_{(2)} + \tfrac{1}{2} x_{(2)}^2 - \frac{(x_{(1)} - \alpha)^3}{3\alpha}$$

$$\text{subject to } c(x) = \frac{1}{(x_{(2)}-2)^2} - 1 = 0$$

Note that the only difference between this problem and Example 1 is the cubic term added to the objective (which has no effect on the algorithm when $x_{(1)} = \alpha$), and that there is still a strong local minimizer at $x = (\alpha,1)$. Again we apply the Coleman-Conn horizontal vertical method to this problem with $M_k = Z_k^T \nabla^2 L(x_k,\lambda_k) Z_k$ which is now

$$\frac{\partial^2 f}{\partial x_{(1)}^2} = 1 - \frac{2(x_{(1)}-\alpha)}{\alpha}.$$

Suppose the initial point is of the form

$$x_0 = (\alpha, 1+\epsilon).$$

Since $x = \alpha$ the horizontal and vertical steps are just as in equations (5) and (6) so

$$x_0 + h_0 = (\alpha(1+\epsilon), 1+\epsilon),$$

and

$$x_1 = (\alpha(1+\epsilon), 1+\epsilon^2).$$

It is straightforward to show that whenever an iterate has the above form, a horizontal step goes to

$$x_1 + h_1 = (\alpha, 1+\epsilon^2),$$

and the vertical step is in example 1:

$$x_2 = (\alpha, 1+\epsilon^4).$$

Now we have a point of the same form as x_0 and the pattern repeats:

$$x_2 + h_2 = (\alpha + \alpha\epsilon^4, 1+\epsilon^4)$$

$$x_3 = (\alpha + \alpha\epsilon^4, 1+\epsilon^8)$$

$$x_3 + h_3 = (\alpha, 1+\epsilon^8)$$

$$x_4 = (\alpha, 1+\epsilon^{16})$$

$$x_4 + h_4 = (\alpha + \alpha\epsilon^{16}, 1+\epsilon^{16})$$

$$x_5 = (\alpha + \alpha\epsilon^{16}, 1 + \epsilon^{32}).$$

Note that at x_k the error is ϵ^{2^k} if k is even, and $\alpha\epsilon^{2^{k-1}}$ if k is odd. Thus, although the convergence is better than two-step quadratic, the one-step convergence is not even linear when $\alpha > 1$. In view of this example it appears that the convergence rate of the Coleman-Conn horizontal vertical method is in general not superlinear.

It is interesting to note that the sequence $\{x_k + h_k\}$ is converging quadratically.

It should be mentioned that this two-step behavior is rather sensitive and that, when a starting point whose first component was not very close to α was used, convergence appeared to be one-step superlinear. For example, consider Example 2 with $\alpha = 5$. The solution is $x = (5,1)$, and we should observe the two-step behavior if we start at a point of the form $x_0 = (5, 1+\epsilon)$. If we instead start at (5.01 , 1.8) the following iterates are produced on a VAX 11/780 in double precision.

$x_{(1)}$	$x_{(2)}$	$\|\|x-(5,1)\|\|_\infty$
5.0100000000000	1.8000000000000	0.8000
9.0160441767068	1.6400000000000	4.0160
5.0424165836447	1.4096000000000	0.4096
7.0829813477127	1.1677721600000	2.0829
4.8267376121833	1.0281474976711	0.1732
5.1260010221621	1.0007922816251	0.1260
5.0008278822104	1.0000006277102	8.2788E-04
5.0000030024674	1.0000000000004	3.0024E-06
5.0000000000002	1.0000000000000	1.6708E-13
5.0000000000000	1.0000000000000	< 1.0E-16

Although initially at every other iterate the $x_{(1)}$ error is smaller than the $x_{(2)}$ error and there is a clear two-step pattern, eventually the steps become nearly tangent to the constraint normals and convergence appears one-step superlinear. Similar behavior was observed for other starting points unless the initial value of $x_{(1)}$ was very close to 5.0. If it is in general so hard to find starting points for which two-step but not one-step superlinear convergence occurs, perhaps it should not be surprising that methods of this type appear one-step superlinear convergent in most experiments.

Next we look at successive quadratic programming applied to our examples. Consider Algorithm 1 (SQP) where the Hessian approximation B_k is defined by

$$\begin{bmatrix} Z_k Y_k \end{bmatrix}^T B_k \begin{bmatrix} Z_k Y_k \end{bmatrix} = \begin{bmatrix} Z_k^T \nabla^2 L(x_k, \lambda_k) Z_k & 0 \\ 0 & I \end{bmatrix}. \tag{7}$$

For our examples this is just a $2x2$ matrix. (Of course the identity in the lower right corner is arbitrary as it does not affect the algorithm but just makes B_k positive definite.) Now note that in the examples, $c(x_k) = c(x_k + h_k)$. Because of this, SQP with B_k given by (7) produces exactly the same iterates for the problems in Examples 1 and 2 as the Coleman-Conn horizontal vertical method does. Therefore all our comments regarding the Coleman-Conn method applied to Examples 1 and 2 also apply to SQP with the above Hessian approximation. In particular Example 2 is an example where SQP, using a Hessian approximation that satisfies Powell's condition (4), generates iterates that do not converge one-step superlinearly.

References

P.T. Boggs, J.W. Tolle and P. Wang[1982]. "On the local convergence of quasi-Newton methods for constrained optimization", *SIAM J. Control and Optimization 20,* pp. 161-171.

T.F. Coleman and A.R. Conn [1982] "Nonlinear programming via an exact penalty function method: asymptotic analysis", *Mathematical Programming 24,* pp. 123-136.

T.F. Coleman and A.R. Conn [1982] "On the local convergence of a quasi-Newton method for a nonlinear programming problem", TR 82-509, Computer Science Department, Cornell University.

R. Fontecilla, T. Steihaug, and R. Tapia [1983]. "A convergence theory for a class of quasi-Newton methods for constrained optimization" Rice University Technical Report 83-15.

M.J.D. Powell [1978]. "The convergence of variable metric methods for nonlinearly constrained optimization calculations", in *Nonlinear Programming 3,* O. Mangasarian, R. Meyer, and S. Robinson, eds., pp.27-63, Academic Press New York.