STRONG UNIQUENESS
IN SEQUENTIAL LINEAR PROGRAMMING

M. R. OSBORNE[1] AND R. S. WOMERSLEY[2]

Abstract

It is known that strong uniqueness can be used to prove second order convergence of the generalised Gauss-Newton algorithm. Formally this algorithm includes sequential linear programming as a special case. Here we show that the second order convergence result extends when the sequential linear programming algorithm is formulated appropriately. Also this discussion provides an example which shows that the assumption of Lipschitz continuity is necessary for the second order convergence result based on strong uniqueness.

1. Introduction

In this note we explore the use of strong uniqueness in providing a sufficient condition for second order convergence in sequential linear programming algorithms (SLP) for mathematical programming problems. Consider the problem

$$\min_{\mathbf{x} \in S} f(\mathbf{x}); S = \{\mathbf{x} \colon g_i(\mathbf{x}) \geq 0, \ i = 1, 2 \ldots n\}. \tag{1.1}$$

It will be convenient to assume that the g_i are concave so that S is convex (this will actually help to make a point rather than prove a major restriction), and that f has a strong unique minimum at $\mathbf{x} = \mathbf{x}^*$ by which we mean that $\exists \gamma > 0, \rho > 0$ small enough such that

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \gamma \|\mathbf{x} - \mathbf{x}^*\|, \ \forall \mathbf{x} \in \mathscr{T}(S, \mathbf{x}^*) \cap \{\|\mathbf{x} - \mathbf{x}^*\| \leq \rho\} \tag{1.2}$$

[1] Statistics Research Section, Mathematical Sciences School, Australian National University, Canberra, A.C.T. 2601, Australia.
[2] School of Mathematics, University of New South Wales, Kensington, N.S.W. 2033, Australia.

where $\mathcal{T}(S, \mathbf{x}^*)$ is the tangent cone to S at $\mathbf{x}^*$. It is not difficult to show that this result is both implied by and implies that the linear subproblem (LSP)

$$\min_{\mathbf{h} \in H} \nabla f(\mathbf{x}^*)\mathbf{h}; H = \{\mathbf{h}: g(\mathbf{x}^*) + \nabla g(\mathbf{x}^*)\mathbf{h} \geq 0\} \tag{1.3}$$

has a strong unique minimum at $\mathbf{x}^*$ corresponding to $\mathbf{h} = 0$. It follows that $\mathbf{x}^*$ is a vertex of both H and S so that at least $p = \dim(\mathbf{x})$ constraints are active at $\mathbf{x}^*$. In particular, at a strong unique minimum it is the constraint geometry rather than the objective function curvature that is the main determining factor.

2. Strong uniqueness and second order convergence

The first use of strong uniqueness as a sufficient condition in proving a second order convergence result appears to be due to Cromme [1] who showed second order convergence of the generalised Gauss-Newton algorithm in nonlinear maximum norm approximation. This result has been extended by Jittorntrum and Osborne [2] to polyhedral norm problems, and they also consider how close the condition is to being necessary. A significant generalisation was introduced in Womersley [3]. This paper considers the composite nondifferentiable optimization problem

$$\min_{\mathbf{x}} f(\mathbf{x}) + F(g(\mathbf{x})) \tag{2.1}$$

where F is convex and f, g are smooth enough. The generalised Gauss-Newton algorithm now takes the form

(i) solve the linear subproblem

$$\min_{\mathbf{x}} f(\mathbf{y}) + \nabla f(\mathbf{y})(\mathbf{x} - \mathbf{y}) + F(r(\mathbf{x}, \mathbf{y})) \tag{2.2}$$

where

$$r(\mathbf{x}, \mathbf{y}) = g(\mathbf{y}) + \nabla g(\mathbf{y})(\mathbf{x} - \mathbf{y}),$$

(ii) line search in the direction determined by $\mathbf{h} = \mathbf{x} - \mathbf{y}$. It is assumed the test will accept the step $\lambda = 1$ if appropriate, and will otherwise reduce λ, for example by seeking

$$\min_{\lambda} f(\mathbf{y} + \lambda \mathbf{h}) + F(g(\mathbf{y} + \lambda \mathbf{h})).$$

(iii) update $\mathbf{y} := \mathbf{y} + \lambda \mathbf{h}$.

The result we need is that

(a) if (2.1) has a strong unique minimum at $\mathbf{x}^*$, and
(b) if $g(\mathbf{x}^*)$ is properly in the interior of $\mathrm{dom}\, F$,

then the generalised Gauss-Newton algorithm converges to $\mathbf{x}^*$ from all close enough starting points, $\lambda = 1$ will be accepted eventually, and the ultimate rate of convergence is second order.

3. Sequential linear programming

SLP is used as a general name for a family of methods which seek an improved estimate to the solution of (1.1) by linearising the objective function and constraints about the current point (early examples include the method of feasible directions and the reduced and projected gradient methods). A recent example of the use of SLP is discussed in [5] where convergence but not second order convergence is shown. This class of method will have second order convergence only when a property akin to strong uniqueness holds at the solution (in all other cases curvature information must be included—this point is discussed in detail in [3]). A first attempt to apply strong uniqueness to SLP might consider setting

$$P(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^{n} \delta(g_i(\mathbf{x})) \tag{3.1}$$

where δ is the elementary indicator function defined by

$$\delta(t) = 0, \qquad t \geq 0,$$
$$= \infty, \qquad t < 0.$$

It is tempting to apply the generalised Gauss-Newton algorithm directly to P, but this will likely be unsuccessful unless S is polyhedral. The second order convergence result does not apply as stated if $g(\mathbf{x}^*) \notin \text{int dom } F$ (so that P is not Lipschitz and $\partial P(\mathbf{x}^*)$ is unbounded). One possible form of difficulty that results is that $\mathbf{h}$ need not be a feasible direction. In general this can happen if $\mathbf{y}$ is on the boundary of S for then $S \subseteq H$ follows from the convexity of S. As the solution of the LSP is at a vertex of H, it follows that $\mathbf{h}$ may be tangential to the boundary of S so that progress in the algorithm cannot be guaranteed. But the following example due to Stephen Wright [4] shows that SLP can be at best linearly convergent even if the successive iterates are in the interior of S and the solution to (1.1) strongly unique. Consider the problem

$$\min f(\mathbf{x}) = x_2 \tag{3.2}$$

subject to

$$g_1(\mathbf{x}) = x_1 + x_2 - \geq 0,$$
$$g_2(\mathbf{x}) = -x_1 + x_2 \geq 0,$$
$$g_3(\mathbf{x}) = x_2^2 \geq 0.$$

If the point (y_1, y_2) is feasible then LSP is

$$\min x_2$$

subject to

$$x_1 + x_2 \geq 0,$$
$$-x_1 + x_2 \geq 0,$$
$$x_2 \geq y_2/2.$$

It follows that $\{x^{(k)}\} \rightarrow 0$ like $[1/2]^k$, demonstrating the linear convergence. The constraint g_3 is redundant in the original problem which has a strongly unique minimum at $x = 0$ (essentially it is equivalent to a linear program with a unique minimum). In LSP the redundant constraint now plays an essential role.

4. An exact penalty function

One remedy is simple, and corresponds to the procedure adopted in [5]. The idea is to permit limited penetration of the infeasible region, and this can be done by introducing the penalty function

$$Q(\mathbf{x}) = f(\mathbf{x}) + \mu \sum_{i=1}^{n} \chi(g_i(\mathbf{x})) \tag{4.1}$$

where

$$\chi(t) = |t|, \qquad t < 0$$
$$= 0, \qquad t \geq 0.$$

This penalty function is exact for (1.1) provided μ is large enough (μ must be larger than the largest of the Kuhn-Tucker multipliers at x^*); but if μ is large enough this also ensures that the LSP (2.2) provides an equivalent statement to the linear program (1.3). Thus the resulting algorithm is just SLP with the modified objective function (4.1) being used in the line search. Presumably, μ should be chosen reasonably small compatible with the requirement of exactness to permit steps with $\lambda = 1$ to be taken readily in order to encourage the onset of second order convergence.

But now $Q(\mathbf{x})$ meets the requirements of the second order convergence result provided μ is finite. It is necessary to verify that $Q(\mathbf{x})$ has a strong unique minimum at x^*, but this is a direct consequence of the LSP having a strong unique minimum at x^*. Thus this version of SLP has a second order convergence rate at a strong unique minimum.

Consider the problem (3.2). Then penalty function is

$$Q(\mathbf{x}) = x_2 + \mu\chi(x_1 + x_2) + \mu\chi(-x_1 + x_2) + \mu\chi(x_2^2);$$

and the penalty function for the linearised problem corresponding to (2.2) is

$$Q_L(\mathbf{x}) = x_2 + \mu\chi(x_1 + x_2) + \mu\chi(-x_1 + x_2) + \mu\chi(2y_2x_2 - y_2^2).$$

To minimize Q_L it is necessary to determine $\mathbf{x}$ such that $0 \in \partial Q_L(\mathbf{x})$. Noting that

$$\begin{aligned}
\partial\chi(t) &= 0, \quad && t > 0, \\
&= [-1, 0], \quad && t = 0, \\
&= -1, \quad && t < 0,
\end{aligned}$$

this gives the following possibilities:
(i) $g_1 > 0, g_2 > 0, 2x_2y_2 - y_2^2 = 0$,

$$\partial Q_L(\mathbf{x}) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} - 2y_2\mu\theta \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad 0 < \theta < 1,$$

and $0 \in \partial Q_L(\mathbf{x})$ is possible if $y_2 > \mu/2$.
(ii) $g_1 < 0, g_2 < 0, 2x_2y_2 - y_2^2 = 0$,

$$\partial Q_L(\mathbf{x}) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \mu\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \mu\begin{bmatrix} -1 \\ 1 \end{bmatrix} - 2y_2\mu\theta\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad 0 < \theta < 1,$$

and $0 \in \partial Q_L(\mathbf{x})$ is possible if $y_2 < -1 + \mu/2$
(iii) $g_1 = 0, g_2 = 0, 2x_2y_2 - y_2^2 < 0$,

$$\partial Q_L(\mathbf{x}) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \mu\theta_1\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \mu\theta_2\begin{bmatrix} -1 \\ 1 \end{bmatrix} - 2y_2\mu\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad 0 < \theta_1, \theta_2 < 1,$$

and $0 \in \partial Q_L(\mathbf{x})$ is possible if $-1 + \mu/2 < y_2 < \mu/2$. In either of the first two cases $x_2 = y_2/2$ and shows the linear decrease observed previously. But if $\mu > 1/2$ then the range of values for y_2 in case (iii) includes $y_2 = 0$ in its interior. Thus the third possibility applies for all small enough y_2 when μ large enough. This case gives the second order convergence (in fact finite termination). This pattern of behaviour, with the fast ultimate convergence setting in only when the current iterate is close enough to the strong unique minimum to be influenced by the local geometry, seems typical for these methods.

References

[1] L. Cromme, "Strong uniqueness. A far reaching criterion for the convergence analysis of iterative procedures," *Numer. Math.* **29** (1978) 179–193.

[2] K. Jittorntrum and M. R. Osborne, "Strong uniqueness and second order convergence in nonlinear discrete approximation," *Numer. Math.* **34** (1980) 439–455.

[3] R. S. Womersley, "Local properties of algorithms for minimizing nonsmooth composite functions," *Math. Prog.* **32** (1985) 69–89.

[4] Stephen Wright, "Convergence of *SQP*-like methods for constrained optimization," *SIAM J. Cont. and Optim.* (to appear).

[5] Jianzhong Zhang, Nae-heon Kim, and L. Lasdon, "An improved successive linear programming algorithm," *Manag. Sci.* **31** (1985) 1312–1331.