SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853-3801

TECHNICAL REPORT NO. 964

April 1991

ANOTHER VARIATIONAL DERIVATION OF A SELF-SCALING QUASI-NEWTON UPDATE FORMULA

by

Michael J. Todd*

*Research supported in part by NSF, AFORS and ONR through NSF grant DMS-8920550.

Abstract

This note provides a variation on a recent result of Fletcher showing that the BFGS and DFP quasi-Newton update formulae for unconstrained minimization satisfy a least-change property with respect to a measure introduced by Byrd and Nocedal. We provide an alternative proof based on Fletcher's for a result of Dennis and Wolkowicz that uses a related measure to derive a quasi-Newton update used in the self-scaling variable metric algorithms of Oren and Spedicato.

1. Introduction

Dennis and Wolkowicz [2] investigated a number of issues concerning quasi-Newton update formulae for use in unconstrained optimization using the measure

$$\omega(A) := \frac{\mathrm{trace}(A)}{[\det(A)]^{1/n}} \tag{1}$$

defined on the cone S_n^+ of $n \times n$ symmetric positive-definite matrices. See also Wolkowicz [12]. Clearly, $\omega(A)$ is related to the ℓ_2-condition number of A, but is sensitive to all its eigenvalues. One of the results of Dennis and Wolkowicz gives a variational derivation of a quasi-Newton update (the "inverse-sized BFGS update") used in the self-scaling variable metric algorithms of Oren, Luenberger and Spedicato [7-9]. Here we will provide an alternative proof of this result using a result of Fletcher [3]. (In fact, instead of $\omega(A)$ we will use $n\, \ell n\, \omega(A)$, but this is a monotonic function of $\omega(A)$; $n\, \ell n\, \omega(A)$ is also closely related to Karmarkar's potential function for linear programming [4], as also noted by Dennis and Wolkowicz.) Fletcher gave a new variational derivation of the BFGS and DFP quasi-Newton update formulae using as a measure of closeness to the identity the function

$$\psi(A) := \mathrm{trace}(A) - \ell n \det(A) \tag{2}$$

defined on S_n^+. The function ψ was introduced by Byrd and Nocedal [1] in their convergence analysis of quasi-Newton methods.

The function $-\ell n \det(\cdot)$ is a self-concordant barrier for the cone S_n^+ in the terminology of Nesterov and Nemirovsky [5, Chapter 3]. The associated parameter is n, much less than the dimensionality $n(n+1)/2$ of S_n^+. This allows efficient optimization over the cone S_n^+, for example to find the maximum volume ellipsoid inscribed in a polyhedron [5, Chapter 7]. Moreover, $-\ell n \det(\cdot)$ is also an n-logarithmically homogeneous barrier for S_n^+ (Nesterov and Nemirovsky [6]), meaning simply that

$$-\ell n\ \det(tA) = -\ell n\ \det(A) - n\ \ell n\ t,\ \ t > 0, \tag{3}$$

since det is homogeneous of degree n. Hence it can also be used in potential function methods for optimizing over S_n^+ [6]. This suggests that we consider the function

$$\phi(A) := n\ \ell n\ \mathrm{trace}(A) - \ell n\ \det(A) \tag{4}$$

on S_n^+, where the coefficient n is chosen to match the parameter of the barrier, so that ϕ is homogeneous of degree 0. Note that $\phi(A) = n\ \ell n\ \omega(A)$.

Suppose A has eigenvalues $\mu_1,\ldots,\mu_n$. Then

$$\begin{aligned}\phi(A) &= n\ \ell n\ (\textstyle\sum \mu_j) - \sum \ell n\ \mu_j \\ &= n\ \ell n\ n + n\ \ell n\ \bar{\mu} - n\ \ell n\ \underline{\mu},\end{aligned} \tag{5}$$

where $\bar{\mu}$ denotes the arithmetic and $\underline{\mu}$ the geometric mean of the positive numbers μ_j. The first expression shows that ϕ is closely related to Karmarkar's potential function

$$n\ \ell n\ c^T x - \sum \ell n\ x_j \tag{6}$$

for minimizing $c^T x$ over the intersection of an affine flat and the nonnegative orthant $\mathbb{R}^n_+$ [4]. The second shows that

$$\phi(A) \geq n\ \ell n\ n,\ \text{ with equality iff }\ A = \mu I\ \text{ for some }\ \mu > 0, \tag{7}$$

from the arithmetic-geometric mean inequality. Hence ϕ measures in some sense the distance of A from the set of scaled identities (see Dennis and Wolkowicz [2, Proposition 2.1]), while ψ measures the distance from the identity (Fletcher [3, Theorem 1.1]).

We will give an alternative proof of the Dennis and Wolkowicz result [2, Theorem 5.1] that ϕ provides a variational derivation of a quasi-Newton update used in self-scaling variable metric algorithms. This result is fairly natural, since, as noted above, ϕ is insensitive to scale.

Two comments are in order before we proceed. First, while ψ is convex, ϕ is not, which somewhat complicates our analysis. (Dennis and Wolkowicz show that ω is pseudo-convex.) Second, Karmarkar's potential function (6) can usually be driven to $-\infty$ when minimizing $c^T x$, while ϕ has a lower bound (7). This is because $c^T x$ can usually achieve its lower bound, assumed to be zero, without all components of x being 0, while $\text{trace}(A) \to 0$ for $A \in S_n^+$ only if $A \to 0$. Hence ϕ is a "centering" potential function, like the part

$$n \,\ell n\, x^T s - \sum \ell n\, x_j - \sum \ell n\, s_j$$

of the primal-dual potential function (e.g. [11]) for linear programming, which strives to keep all products $x_j s_j$ equal (x_j is a primal variable, s_j a dual slack).

2. The Result

In the interpretation of the theorem below, B is viewed as an approximation to the Hessian matrix of a nonlinear function being minimized by an iterative algorithm. A step from x to x_+ has resulted in new information; the change $s = x_+ - x$ in parameter values yields the difference y in gradient values. A line search assures $s^T y > 0$, and then we seek a new approximation B_+ to the Hessian matrix that incorporates the new information yet is close in some sense to the old matrix B.

Theorem (Dennis and Wolkowicz [2]). Let $B \in S_n^+$ with $H = B^{-1}$, and let $s,y \in \mathbb{R}^n$ with $s^T y > 0$. Then the unique solution to

$$\min_{B_+ \in S_n^+} \phi(H^{1/2} B_+ H^{1/2})$$

$$B_+ = B_+^T \tag{8}$$

$$B_+ s = y$$

is

$$B_+ = \Big(B - \frac{Bss^TB}{s^TBs}\Big)\gamma + \frac{yy^T}{s^Ty}, \tag{9}$$

with

$$\gamma = \frac{y^THy}{s^Ty}. \tag{10}$$

Proof. First we show that (8) has a solution. Clearly, (8) has a feasible solution, for example that given by (9), and we may confine our search to those B_+ with $\phi(H^{1/2}B_+H^{1/2})$ at most $\phi(H^{1/2}\hat{B}_+H^{1/2})$ for some fixed feasible $\hat{B}_+$. We show that such B_+ lie in a compact subset of S_n^+, whence the minimum is attained.

Let $\bar{B}_+ := H^{1/2}B_+H^{1/2}$ have eigenvalues $\mu_1,\ldots,\mu_n$, so that

$$\phi(H^{1/2}B_+H^{1/2}) = n\ \ell n \sum \mu_j - \sum \ell n\ \mu_j$$

$$= \sum_j \ell n \frac{\sum_i \mu_i}{\mu_j}.$$

Since each summand is nonnegative, an upper bound on the sum implies that each summand is bounded so that there is $\epsilon > 0$ with

$$\mu_j \geq \epsilon \sum_i \mu_i \text{ for each } j. \tag{11}$$

But $\bar{B}_+\bar{s} = \bar{y}$ (with $\bar{y} = H^{1/2}y$, $\bar{s} = H^{-1/2}s$) shows $||\bar{B}_+|| \geq ||\bar{y}||/||\bar{s}||$. This provides a lower bound on the maximum eigenvalue of $||\bar{B}_+||$, and then (11) shows that each μ_j is bounded away from zero. Similarly, $||\bar{B}^{-1}|| \geq ||\bar{s}||/||\bar{y}||$, which gives an upper bound on the smallest eigenvalue of $\bar{B}_+$ and hence, using (11) again, on each μ_j. But the set of matrices $\bar{B}_+$ in S_n^+ with each eigenvalue in some compact interval in $(0,\infty)$ is compact, and hence so is the corresponding set of B_+'s.

Because the nontrivial linear constraints of (8) are linearly independent, the optimal solution must be a stationary point of the Lagrangian function

$$L(B_+, \Lambda, \lambda) := \phi(H^{1/2}B_+H^{1/2}) + \mathrm{trace}(\Lambda^T(B_+^T - B_+)) + \lambda^T(B_+s - y)$$

where Λ and λ are Lagrange multipliers. Proceeding exactly as in Fletcher [3], we find

$$0 = \frac{\partial L}{\partial (B_+)_{ij}} = \frac{1}{2}\left(\frac{n}{\mathrm{trace}\left(H^{1/2}B_+H^{1/2}\right)} H_{ji} - (B_+^{-1})_{ji}\right) + \Lambda_{ji} - \Lambda_{ij} + (\lambda s^T)_{ij}, \tag{12}$$

for all i,j.

Let

$$\nu := \frac{\mathrm{trace}(H^{1/2}B_+H^{1/2})}{n} . \tag{13}$$

Multiplying (12) by ν, we obtain

$$0 = \frac{1}{2}\,(H_{ji} - ((B_+/\nu)^{-1})_{ji}) + \nu\Lambda_{ji} - \nu\Lambda_{ij} + ((\nu\lambda)s^T)_{ij}$$

for all i,j. This shows that B_+/ν is the unique solution (with multipliers $\nu\Lambda$, $\nu\lambda$) found by Fletcher for the problem

$$\begin{aligned} &\min_{C_+\in S_n^+} \psi(H^{1/2}C_+H^{1/2}) \\ &C_+ = C_+^T \\ &C_+s = y/\nu, \end{aligned}$$

so that Theorem 2.1 of [3] gives

$$B_+/\nu = B - \frac{Bss^TB}{s^TBs} + \frac{(y/\nu)(y/\nu)^T}{s^T(y/\nu)}$$

or

$$B_+ = \left(B - \frac{Bss^TB}{s^TBs}\right)\nu + \frac{yy^T}{s^Ty} . \tag{14}$$

The only possible freedom we have in choosing B_+ is the choice of ν. But if we pre- and postmultiply (14) by $H^{1/2}$ and take the trace we find

$$\begin{aligned} n\nu &= \left(\operatorname{trace}(I) - \operatorname{trace}\left(\frac{B^{1/2}s^Ts^TB^{1/2}}{s^TBs}\right)\right)\nu + \operatorname{trace}\left(\frac{H^{1/2}yy^TH^{1/2}}{s^Ty}\right) \\ &= (n-1)\nu + \frac{y^THy}{s^Ty} , \end{aligned}$$

so that ν must equal γ in (10). Since (8) has a solution, since any such solution is a stationary point of L, and since the only stationary point is given by (9), that must be the unique solution to (8). The proof is complete.

The update (9) with γ given by (10) is the first case of the update resulting from the switching strategy in Oren [7], and is the optimally-conditioned one from a class of updates considered by Oren and Spedicato [9] when γ is fixed as in (10). It is also an update suggested for the initial update by Shanno and Phua [10]. Another update suggested in [7,9] arises in the version of the theorem where B and H, B_+ and H_+, and s and y are interchanged.

References

[1] R.H. Byrd and J. Nocedal, "A tool for the analysis of quasi-Newton methods with application to unconstrained minimization," SIAM J. Numer. Anal., 26 (1989), pp. 727-739.

[2] J.E. Dennis and H. Wolkowicz, "Sizing and least-change secant methods," Research Report, Department of Combinatorics and Optimization, University of Waterloo (Waterloo, Ontario, 1989).

[3] R. Fletcher, "A new variational result for quasi-Newton formulae," SIAM J. Opt., 1 (1991), pp. 19-21.

[4] N. Karmarkar, "A new polynomial-time algorithm for linear programming," Combinatorica, 4 (1984), pp. 373-395.

[5] Ju. E. Nesterov and A.S. Nemirovsky, Self-Concordant Functions and Polynomial-Time Methods in Convex Programming, Central Economic and Mathematical Institute, USSR Academy of Sciences, Moscow, 1989.

[6] Ju. E. Nesterov and A.S. Nemirovsky, "Logarithmically homogeneous self-concordant barriers and polynomial time potential function methods for convex programming problems," preprint, Central Economic and Mathematical Institute, USSR Academy of Sciences, Moscow, 1990.

[7] S.S. Oren, "On the selection of parameters in self-scaling variable metric algorithms," Math. Programming, 7 (1974), pp. 351-367.

[8] S.S. Oren and D.G. Luenberger, "Self-scaling variable metric (SSVM) algorithms, Parts I and II," Management Science, 20 (1974), pp. 845-862, 863-874.

[9] S.S. Oren and E. Spedicato, "Optimal conditioning of self-scaling variable metric algorithms," Math. Programming, 10 (1976), pp. 70-90.

[10] D.F. Shanno and K.-H. Phua, "Matrix conditioning and nonlinear optimization," Math. Programming, 14 (1978), pp. 149-160.

[11] M.J. Todd and Y. Ye, "A centered projective algorithm for linear programming," Math. Oper. Res., 15 (1990), pp. 508-529.

[12] H. Wolkowicz, "Measures for symmetric rank-one updates," Research Report, Department of Combinatorics and Optimization, University of Waterloo (Waterloo, Ontario, 1990).