International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

Analogues of Dixon's and Powell's Theorems for Unconstrained Minimization with Inexact Line Searches

Nazareth, J.L.

IIASA Working Paper

WP-82-100

October 1982

Working Paper

ANALOGUES OF DIXON'S AND POWELL'S THEOREMS FOR UNCONSTRAINED MINIMIZATION WITH INEXACT LINE SEARCHES

Larry Nazareth

October 1982
WP-82-100

International Institute for Applied Systems Analysis
A-2361 Laxenburg, Austria

NOT FOR QUOTATION
WITHOUT PERMISSION
OF THE AUTHOR

ANALOGUES OF DIXON'S AND POWELL'S THEOREMS FOR UNCONSTRAINED MINIMIZATION WITH INEXACT LINE SEARCHES

Larry Nazareth

October 1982
WP-82-100

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

ABSTRACT

By modifying the way in which search directions are defined, we show how to relax the restrictive assumption that line searches must be exact in the theorems of Dixon and Powell. We show also that the BFGS algorithm modified in this way is equivalent to the three-term-recurrence (TTR) method for quadratic fuctions.

ANALOGUES OF DIXON'S AND POWELL'S THEOREMS FOR UNCONSTRAINED MINIMIZATION WITH INEXACT LINE SEARCHES

Larry Nazareth

1. Introduction

We are concerned with the problem: minimize $f(x), x \in R^n$, using a variable metric algorithm in the Broydon β-class, see Broydon, 1970. The underlying family of updates is defined as follows: Given an approximation H_k to the inverse hessian of $f(x)$, a step δx_k and gradient change δg_k corresponding to it with $\delta x_k^T \delta g_k \neq 0$ a new approximation H_{k+1}^{β}, which satisfies the quasi-Newton relation $H_{k+1}^{\beta} \delta g_k = \delta x_k$ is defined by

$$H_{k+1}^{\beta} = H_{k+1}^{BFGS} + \beta_k w_k w_k^T \tag{1.1a}$$

where

$$H_{k+1}^{BFGS} = (I - \rho_k \delta x_k \delta g_k^T) H_k (I - \rho_k \delta x_k \delta g_k^T)^T + \rho_k \delta x_k \delta x_k^T \tag{1.1b}$$

$$w_k = H_k \delta g_k - \left[\frac{\delta g_k^T H_k \delta g_k}{\delta g_k^T \delta x_k} \right] \delta x_k \tag{1.1c}$$

β_k is a real number and $\rho_k = 1/\delta g_k^T \delta x_k$.

Dixon's 1972 theorem states that all methods in the Broydon β - class develop identical iterates when line searches are exact, conflicts in choice of minimum along a line are unambiguously resolved and the same initialization is used. Powell's, 1972, theorem which also requires similar assumptions, is closely related. It states that a sequence of updates from the β -class which terminate with a BFGS update give the same hessian approximation matrix regardless of which particular updates were used prior to the last one. By suitably modifying the way in which search directions are defined we show how to relax the restrictive assumption that line searches be exact in both these theorems. We also show that the BFGS algorithm modified in this way reduces to a conjugate direction method known as the three-term-recurrance (TTR). This then bears the same relation to the modified BFGS algorithm as the conjugate gradient method bears to the standard BFGS algorithm (see Nazareth, 1979).

2. Main Results

Henceforth we shall attach the symbol * for the case when line searches are exact. We define search directions by

$$d_{k+1}^* = -H_{k+1}^* g_{k+1}^* \quad , \quad d_{k+1}^{*BFGS} = -H_{k+1}^{*BFGS} g_{k+1}^* \tag{2.1}$$

and iterates by

$$x_{k+2}^* = x_{k+1}^* + \lambda_{k+1}^* d_{k+1}^* \tag{2.2}$$

Lemma 2.1: (Shanno & Kettler, 1970). If line searches are exact, then

$$d_{k+1}^{*BFGS} = -w_k^* \tag{2.3a}$$

$$d_{k+1}^{*} = -(1 + \beta_k g_{k+1}^{*} H_k^{*} g_{k+1}^{*}) d_{k+1}^{*BFGS} \tag{2.3b}$$

Proof: See, for example, Powell, 1972.

Lemma 2.1 says that $\delta x_k^{*} || d_k^{*} || w_{k-1}^{*}$ provided that $\beta_k \neq 1/(g_{k+1}^{*T} H_k^{*} g_{k+1}^{*}\quad)$. If we write $M_k^{*} = (I - \rho_k^{*} \delta x_k^{*} \delta g_k^{*})^T$ then

$$H_{k+1}^{*BFGS} = M_k^{*}(H_k^{*BFGS} + \beta_{k-1} w_{k-1}^{*} w_{k-1}^{*T}) M_k^{*T} + \rho_k^{*} \delta x_k^{*} \delta x_k^{*T} \tag{2.4}$$

$\delta x_k^{*} || w_{k-1}^{*}$ and Lemma 2.1 together imply that

$$M_k^{*}(\beta_{k-1} w_{k-1}^{*} w_{k-1}^{*T}) M_k^{*T} = 0$$

Hence,

$$H_{k+1}^{*BFGS} = M_k^{*} H_k^{*BFGS} M_k^{*T} + \rho_k^{*} \delta x_k^{*} \delta x_k^{*T}$$

This provides the basis for an inductive proof of the results quoted above. We should mention that the value $\beta_k = 1/g_{k+1}^{*T} H_k^{*} g_{k+1}^{*}$ is outlawed since it would give $w_k^{*} = 0$.

Motivated by these results, we turn to the case when line searches are no longer required to be exact. We shall now define search directions by

$$d_1 = -H_1 g_1 \tag{2.5a}$$

$$d_{k+1} = -w_k = -(H_k \delta g_k - \frac{\delta g_k^T H_k \delta g_k}{\delta g_k^T \delta x_k} \delta x_k) \ , \ k \geq 1 \ , \tag{2.5b}$$

and iterates by

$$x_{k+2} = x_{k+1} + \lambda_{k+1} d_{k+1} \tag{2.6}$$

This is certainly not the conventional way in which variable metric methods develop a search direction. However, we can note the following:

1. When line searches are exact $d_{k+1} || d^*_{k+1}$. This follows directly from Lemma 2.1.

2. d_{k+1} is a conjugate direction, since $d^T_{k+1}\delta g_k = 0$.

3. As we shall see in Section 3, the resulting method is equivalent to a standard conjugate direction method when applied to a quadratic function.

We now have the following theorem which is the natural extension of the results of Powell, 1972 and Dixon 1972 quoted above.

Theorem 2.1: If the method based upon (1.1a-c) and (2.5a-b) with x_1 and $H_1 > 0$ given, is used to minimize a differentiable function $f(x)$ and if the steps are defined unambiguously, for example, using normalized search directions and given values of λ_k in (2.6), then the sequence of points x_k and the sequence of matrices H_k^{BFGS}, $k = 1,2,3,\ldots$, are independent of the parameter values β_k, $k = 1,2,3,\ldots$, provided the search directions defined by (2.5) do not vanish.

Proof: Since H_1 is given, d_1 is obviously independent of the parameters β_k, $k = 1,2,3,\ldots,$. x_2 is then independent of the parameters and so is H_2^{BFGS}. $d_2 \in \left[H_1\delta g_1, \delta x_1\right]$ and $d_2^T\delta g_1 = 0$, and thus d_2 is independent of the parameters.

We now use induction. Suppose that for $k = 2,3,\ldots,$ x_{k+1} and H_{k+1}^{BFGS} are independent of the parameters. We must show this to be true for x_{k+2} and H_{k+2}^{BFGS} . From (2.5) we have

$$d_{k+1} = -\left[(H_k^{BFGS} + \beta_k w_{k-1} w_{k-1}^T)\delta g_k - \frac{\delta g_k^T H_k \delta g_k}{\delta g_k^T \delta x_k}\delta x_k\right]$$

Also $w_{k-1} \,/\!/\, \delta x_k$.

Provided d_{k+1} does not vanish, we have

$$d_{k+1} \in \left[H_k^{BFGS}\delta g_k, \delta x_k\right] \ , \ d_{k+1}^T \delta g_k = 0$$

Thus d_{k+1} is independent of the parameters. Therefore, δx_{k+1} and δg_{k+1} are also independent of the parameters, and so is x_{k+2}.

We must now show that H_{k+2}^{BFGS} is independent of the parameters.

Writing

$$M_{k+1} = (I - \rho_{k+1}\delta x_{k+1}\delta g_{k+1}^T)$$

we have

$$H_{k+2}^{BFGS} = M_{k+1}H_{k+1}M_{k+1}^T + \rho_{k+1}\delta x_{k+1}\delta x_{k+1}^T$$

$$= (M_{k+1}H_{k+1}^{BFGS}M_{k+1}^T) + M_{k+1}(\beta_k w_k w_k^T)M_{k+1}^T + \rho_{k+1}\delta x_{k+1}\delta x_{k+1}^T$$

But $w_k \,/\!/\, \delta x_{k+1}$, and hence $M_{k+1} w_k w_k^T M_{k+1}^T = 0$. It follows that H_{k+2}^{BFGS} is independent of the parameters. This completes the proof of the theorem.

3. Specialization to Quadratic Functions

We now show that for a quadratic function, the algorithm defined by (1.1) and (2.5) using the BFGS option is the three-term-recurrence (TTR) algorithm given in Nazareth, 1977. In this method, which employs the metric defined by $H > 0$, search directions are given by

$$d_1 = -Hg_1 \tag{3.1a}$$

$$d_2 = -H\delta g_1 + \frac{\delta g_1^T H \delta g_1}{\delta g_1^T \delta x_1}\delta x_1 \tag{3.1b}$$

$$d_{k+1} = -H\delta g_k + \frac{\delta g_{k-1}^T H \delta g_k}{\delta g_{k-1}^T \delta x_{k-1}}\delta x_{k-1} + \frac{\delta g_k^T H \delta g_k}{\delta g_k^T \delta x_k}\delta x_k \tag{3.1c}$$

Theorem 3.1: Consider the algorithm defined by (1.1) with $\beta_k = 0$, i.e., using the BFGS option. Let x_1 and $H_1 = H > 0$ be given and suppose the algorithm is applied to quadratic function $\psi(x)$. Then search directions are conjugate, H_{k+1} satisfies $H_{k+1}\delta g_j = \delta x_j$, $j = 1,2,\ldots,k$, and the search directions d_{k+1} are the same as those given by (3.1), in length and direction.

Proof: (2.5a) and (3.1a) define the same search directions. $H_2\delta g_1 = \delta x_1$ and d_2 is conjugate to $d_1 = -H_1 g_1$. Also $H_3\delta g_j = \delta x_j$, $j = 1,2$.

We now use induction to complete the proof. Suppose the claims of the lemma hold for iterates upto x_{k+1}, i.e., $d_1,\ldots,d_k$ are conjugate, $H_{k+1}\delta g_j = \delta x_j$, $j = 1,2,\ldots,k$ and search directions defined by (2.5) and (3.1) are the same for $d_1,\ldots,d_k$.

For $j \le (k-1)$

$$\delta g_j^T d_{k+1} = -\delta g_j^T H_k \delta g_k + \frac{\delta g_k^T H_k \delta g_k}{\delta g_k^t \delta x_k}\delta g_j^T \delta x_k$$

Using $\delta g_j^T H_k = \delta x_j^T$ and $\delta g_j^T \delta x_k = 0$ we have

$$\delta g_j^T d_{k+1} = 0 \ , \ j \le k-1$$

Since $\delta g_k^T d_{k+1} = 0$, by the definition of d_{k+1}, we have d_{k+1} conjugate to all previous search directions. $(H_{k+1}\delta g_{k+1} - \delta x_{k+1})$ and δx_{k-1} are conjugate to δx_j , $j = 1,2,\ldots,k$. Thus $H_{k+1}\delta g_{k+1}$ is conjugate to δx_j , $j = 1,2,\ldots,k$. $H_{k+2}\delta g_{k+1} = \delta x_{k+1}$ by definition. Because H_{k+2} is obtained by updating H_{k+1} using rank-1 matrices composed from $H_{k+1}\delta g_{k+1}$ and δx_{k+1} it has the hereditary property, i.e., $H_{k+2}\delta g_j = \delta x_j$, $j = 1,2,\ldots,k+1$.

Finally we can readily show that

$$H_k = (I - \sum_{j=1}^{k-1}\rho_j \delta x_j \delta g_j^T)H(I - \sum_{j=1}^{k-1}\rho_j \delta x_j \delta g_j^T)^T + \sum_{j=1}^{k-1}\delta x_j \delta x_j^T$$

Substituting into (2.5) and using $\delta x_j^T \delta g_k = 0$, $j = 1,2,\ldots,k-1$

$$d_{k+1} = -(I - \sum_{j=1}^{k-1}\rho_j \delta x_j \delta g_j^T)H\delta g_k + \frac{\delta g_k^T H_k \delta g_k}{\delta g_k^T \delta x_k}\delta x_k \tag{3.2}$$

$$d_{k+1} = -H\delta g_k + \sum_{j=1}^{k-1}\rho_j(\delta g_j^T H \delta g_k)\delta x_j + \frac{\delta g_k^T H_k \delta g_k}{\delta g_k^T \delta x_k}\delta x_k \tag{3.3}$$

Since the induction hypothesis and (3.1) imply that

$$[\delta g_1, \ldots, \delta g_{k-2}] \subseteq [\delta x_1, \ldots, \delta x_{k-1}]$$

it follows that

$$\sum_{j=1}^{k-1}\rho_j(\delta g_j^T H \delta g_k)\delta x_j = \rho_{k-1}(\delta g_{k-1}^T H \delta g_k)\delta x_{k-1} \tag{3.4}$$

Therefore

$$d_{k+1} = -H\delta g_k + \frac{\delta g_{k-1}^T H \delta g_k}{\delta g_{k-1}^T \delta x_{k-1}}\delta x_{k-1} + \frac{\delta g_k^T H \delta g_k}{\delta g_k^T \delta x_k}\delta x_k \tag{3.5}$$

This completes the proof.

One should note that the search vectors for the algorithm defined by the BFGS update and (2.5) are the *same in length and direction* as those of the TTR method. If other updates were used in place of the

BFGS, then we would obtain search vectors that coincide in direction but not in length. We see that the modified BFGS algorithm stands in relation to the TTR method, in the same way as the standard BFGS method is related to the conjugate gradient method, see Nazareth, 1979. It is also interesting to note that Theorem 3.1 suggests a new way to implement the TTR method based upon a limited memory BFGS update and definition of search directions by (2.5b).

REFERENCES

Broyden, C.G. (1970), "The convergence of a class of double-rank minimization algorithms", Journal of the Institute of Mathematics and its Applications, 6, 76-90.

Dennis, J.E. and J.J. More (1977), "Quasi-Newton methods, motivation and theory", SIAM Review, 19, 46-89.

Dixon, L.C.W. (1972), "Quasi-Newton algorithms generate identical points", Mathematical Programming, 2, 383-387.

Goldfarb, D. (1970), "A family of variable metric methods derived by variational means", Mathematics of Computation, 24, 23-26.

Greenstadt, J. (1970), "Variations on variable metric methods", Mathematics of Computation, 24, 1-18.

Murray, W. (Ed.) (1972), Numerical Methods for Unconstrained Optimization, Academic Press, London and New York.

Nazareth, L. (1977), "A conjugate direction algorithm without line series", Journal of Optimization Theory and Applications, 23, 373-387.

Nazareth, L. (1979), "A relationship between the BFGS and conjugate gradient algorithms and its implication for new algorithms", SIAM Journal of Numerical Analysis, 16, 794-800.

Powell, M.J.D. (1972), "Unconstrained minimization and extensions for constraints", Report T.P. 495, Atomic Energy Research Establishment, Harwell, England.

Shanno, D.F. and P.C. Kettler (1970), "Optimal conditioning of quasi-Newton methods", Mathematics of Computation, 24, 657-664.