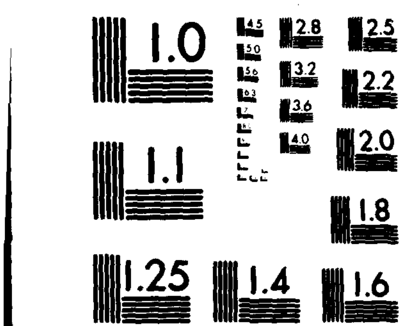
AD-A094 257 MARYLAND UNIV COLLEGE PARK DEPT OF COMPUTER SCIENCE F/6 12/1
CONSTRAINED DEFINITE HESSIANS TEND TO BE WELL CONDITIONED.(U)
JUN 80 G W STEWART N00014-76-C-0391
UNCLASSIFIED TR-908 NL

END
DATE
FILMED
2-81
DTIC



MICROCOPY RESOLUTION TEST CHART

LEVEL II

(2)

AD A094257

DTIC ELECTE JAN 29 1981 S D

UNIVERSITY OF MARYLAND
COMPUTER SCIENCE CENTER

COLLEGE PARK, MARYLAND
20742

81 1 27 015

9 Technical Report 908   11 June 1980

6 Constrained Definite Hessians Tend to be Well Conditioned.

11 G. W. Stewart*

TR-9??   12 11

Abstract

The testing of optimization algorithms requires the running of problems with ill-conditioned Hessians. For constrained problems, it is the projection of the Hessian onto the space determined by the active constraints that must be ill conditioned. In this note it is argued that unless the Hessian and the constraints are constructed together, the constrained Hessian is likely to be well conditioned. The approach is to examine the effects of random constraints on a singular Hessian.

page - A-

*Department of Computer Science and the Institute for Physical Science and Technology. This work was supported in part by the Office of Naval Research under Contract No. N00014-76-C-0391.

15

401 C22

JOB

Constrained Definite Hessians Tend to be Well Conditioned

G. W. Stewart

In testing and comparing optimization algorithms, it is important to include test problems for which the Hessian matrix of the objective function is ill conditioned, both at the optimum point and away from it. For unconstrained optimization this is easy enough to do, and problems with ill-conditioned or even singular Hessians appear frequently in the literature.

For constrained optimization problems, however, the operative condition number is usually that of the projection of the Hessian onto the space of active constraints (or the tangent space in the case of nonlinear constraints). It is the purpose of this note to show that such a projection will tend to be well conditioned, even when the underlying Hessian is singular.

It is easy to see that projection can only improve the condition of a definite matrix. Specifically, for a positive definite matrix H of order n, define the condition number $\kappa(H)$ by

$$\kappa(H) = \lambda_{max}/\lambda_{min} \tag{1}$$

where λ_{max} and λ_{min} are the largest and smallest eigenvalues of H (n.b. this definition is appropriate only for positive definite matrices; it does not generalize to indefinite or nonsymmetric matrices). A set of constraints may be specified by a set $\{q_1, q_2, \ldots, q_p\}$ of independent vectors, which, without loss of generality, may be taken to be orthonormal. The constraint space is then the orthogonal complement of the space spanned by $q_1, q_2, \ldots, q_p$. Thus if we set

$$Q_1 = (q_1, q_2, \ldots, q_p)$$

and determine an orthogonal matrix

$$Q = (Q_1 \; Q_2) \tag{2}$$

whose first p columns are the vectors $q_1, q_2, \ldots, q_p$, then the constraint space will be spanned by the columns of Q_2.

The constrained Hessian is

$$H_c = Q_2^T H Q_2 \quad . \tag{3}$$

It follows from standard results of matrix theory [2] that

$$\lambda_{max}(H_c) \leq \lambda_{max}(H)$$

and

$$\lambda_{min}(H_c) \geq \lambda_{min}(H) \quad .$$

Hence from (1)

$$\kappa(H_c) \leq \kappa(H) \quad .$$

Moreover, equality will be attained only if the constraint space contains eigenvectors of H corresponding to $\lambda_{max}(H)$ and $\lambda_{min}(H)$. This suggests that unless Q_1 is specially chosen, $\kappa(H_c)$ can be appreciably smaller than $\kappa(H)$. The rest of this paper is an attempt to give some quantitative substance to this conjecture by examining the behavior of $\kappa(H_c)$ when the constraint space is chosen at random.

For definiteness we shall consider the singular matrix

$$H = \begin{bmatrix} I_{n-1} & 0 \\ 0 & 0 \end{bmatrix}, \tag{4}$$

where I_{n-1} is the identity matrix of order n-1. We shall determine a randomly constrained Hessian by choosing a random orthogonal matrix Q from the Haar distribution on the group of orthogonal matrices [1], partitioning Q as in (2), and defining H_c by (3). The distribution from which Q is chosen is analogous to the uniform distribution. Computationally such a Q may be obtained by orthogonalizing a set of n vectors whose components are identically distributed, independent, normal random variables [3]. In particular, any row or column of such a matrix has the distribution of a normalized vector of identically distributed, independent, random normal variables.

The principal result is contained in the following theorem.

<u>Theorem</u>. Let Q be a random orthogonal matrix from the Haar distribution on the group of orthogonal matrices. Let Q be partitioned as in (2), where Q_1 has p columns. Let H be defined by (4) and

H_c by (3). If $p \leq n-2$, then

$$\kappa(H_c) = 1 + \frac{n-p}{p} F \tag{5}$$

where F has an F-distribution with $n-p$ and p degrees of freedom.

Proof. Since Q_2 has at least two colums,

$$\lambda_{max}(H_c) = 1 \quad .$$

Thus the problem becomes that of determining the distribution of $\lambda_{min}(H_c)$. We shall use the characterization

$$\lambda_{min}(H_c) = \min_{\|x\|=1} x^T H_c x \quad ,$$

where $\|\cdot\|$ denotes the usual Euclidean norm.

Let Q be partitioned in the form

$$Q = \begin{array}{cc} p & n-p \\ \left[\begin{array}{c} Q_{11} \\ q_{21}^T \end{array}\right. & \left.\begin{array}{c} Q_{12} \\ q_{22}^T \end{array}\right] \end{array}$$

Then $H_c = Q_{12}^T Q_{12}$. Hence

$$\begin{aligned} \lambda_{min}(H_c) &= \min_{\|x\|=1} x^T Q_{12}^T Q_{12} x = \min_{\|x\|=1} \| Q_{12} x \|^2 \\ &= \min_{\|x\|=1} 1 - (q_{22}^T x)^2 \quad , \end{aligned} \tag{6}$$

the last inequality following from the fact that the vector

$$\begin{bmatrix} Q_{12} \\ q_{22}^T \end{bmatrix} x$$

has norm one. The last expression in (6) is clearly minimized when $x = q_{22}/\|q_{22}\|$, in which case

$$\lambda_{min}(H_c) = 1 - \|q_{22}\|^2 = \|q_{21}\|^2 ,$$

since $\|(q_{21}^T, q_{22}^T)\|^2 = 1$.

Let y be an n_1 vector of identically distributed, independent normal random variables, and partition $y^T = (y_1^T, y_2^T)$, where y_1 is a p-vector. Then by the observations made before the theorem, $\lambda_{min}(H_c)$ has the distribution of

$$\frac{\|y_1\|^2}{\|y\|^2} = \frac{\|y_1\|^2}{\|y_1\|^2 + \|y_2\|^2}$$

Thus $\kappa(H_c) = 1/\lambda_{min}(H_c)$ has the distribution of

$$\frac{\|y_1\|^2 + \|y_2\|^2}{\|y_1\|^2} = 1 + \frac{n-p}{p} \frac{p\|y_2\|^2}{(n-p)\|y_1\|^2}$$

$$= 1 + \frac{n-p}{p} F ,$$

where F has an F-distribution with $n-p$ and p degrees of freedom.

Note. The proof of the theorem can easily be extended to cover the case $H_c = \mathrm{diag}(\lambda_1,\lambda_2,\ldots,\lambda_{n-1},0)$ where $\lambda_1 > \lambda_2 > \ldots > \lambda_{n-1} > 0$. For this matrix

$$\kappa(H_c) \le \frac{\lambda_1}{\lambda_{n-1}} \left[1 + \frac{n-p}{p} F\right]$$

where again F has an F-distribution with p and $n-p$ degrees of freedom. We do not pursue this embellishment here because the simpler case (4) adequately illustrates how likely projection is to produce a well-conditioned matrix.

The well known properties of the F distribution along with (5) can be used to determine what is a probable value of the condition number of H_c. Table 1 gives values of μ_p such that for $n \ge p+2$

(7) $$P\{\kappa(H_c) \le 1 + (n-p)\mu_p\} \ge 0.99 \quad .$$

Thus for $p = 3$, we shall observe $\kappa(H_c) \le 1 + 9.4(n-3)$ at least ninety nine percent of the time, and n will have to be very large indeed to produce the degree of ill-conditioning that would seriously discommode a well constructed algorithm.

It would be wrong to conclude from this analysis that ill-conditioned constrained Hessians do not occur in practical problems. Nature has a way

Table 1

μ_p from (7)

p	μ_p
1	6370
2	49.8
3	9.40
4	3.80
5	2.10
6	1.35
7	0.960
8	0.726
9	0.576
10	0.471

of confounding naive randomness assumptions by behaving in a distinctly nonrandom and frequently perverse manner. However, to the extent that ill-conditioned constrained Hessians occur in practice, to that extent there is a need for test problems with such Hessians; and the above analysis has implications for the construction of these problems. Namely, it is not enough to choose an objective function with an ill-conditioned Hessian and hope that unsystematically chosen constraints will preserve the ill-conditioning; rather the constraints and the Hessian must be constructed together.

References

1. P.R. Halmos, <u>Measure Theory</u>, Van Nostrand, Princeton, New Jersey (1950).

2. A. S. Householder, <u>The Theory of Matrices in Numerical Analysis</u>, Blaisdell, New York (1964).

3. G. W. Stewart, <u>The efficient generation of random orthogonal matrices with an application to condition estimators</u>, to appear SIAM J. Numer. Anal.


-8