Derivation of Error Bounds for Quadratic Regression (01/27/2011)

A. R. Conn, K. Scheinberg, and L. N. Vicente, Introduction to Derivative-Free Optimization, MPS-SIAM Book Series on Optimization, SIAM, Philadelphia, 2009

Approx. xii + 277 pages / Softcover / ISBN: 978-0-898716-68-9

The error bounds for quadratic regression models (based on the condition number of a scaled version of the regression matrix), given in Theorem 4.13, can be derived as follows.

First, note that one can assume $y^0 = 0$ without loss of generality (see the argument at the end of the proof of Theorem 3.16). Recall that the regression model is written as $m(y) = c + g^\top y + (1/2)y^\top Hy$, the sample set is $Y = \{f(y^0), f(y^1), \ldots, f(y^p)\}$, $\bar{\phi}$ is the canonical basis in (3.1), and $M = M(\bar{\phi}, Y)$ is the regression matrix. Under the Lipschitz continuity of the Hessian of f (with constant $\nu_2 > 0$), one has

$$M \begin{bmatrix} f(y^0) \\ \nabla f(y^0) \\ \text{vec}(\nabla^2 f(y^0)) \end{bmatrix} - f(Y) \;=\; r,$$

with $f(Y) = (f(y^0), f(y^1), \ldots, f(y^p))^\top$ and $|r_i| \leq (\nu/2)\Delta^3$, $i = 0, \ldots, p$. Here we used the notation vec(A) for the vectorial representation of A that uses the ordering of the columns of M corresponding to the quadratic terms. Thus, one obtains[1]

$$\begin{bmatrix} f(y^0) \\ \nabla f(y^0) \\ \text{vec}(\nabla^2 f(y^0)) \end{bmatrix} - M^\dagger f(Y) \;=\; \begin{bmatrix} f(y^0) \\ \nabla f(y^0) \\ \text{vec}(\nabla^2 f(y^0)) \end{bmatrix} - \begin{bmatrix} c \\ g \\ \text{vec}(H) \end{bmatrix} \;=\; M^\dagger r. \tag{0.1}$$

Letting I_n and $I_{\bar{p}}$ be the identity matrices of order n and $\bar{p} = n(n+1)/2$, respectively, one can write

$$M^\dagger \;=\; \begin{bmatrix} 1 & 0 & 0 \\ 0 & (1/\Delta)I_n & 0 \\ 0 & 0 & (1/\Delta^2)I_{\bar{p}} \end{bmatrix} \hat{M}^\dagger, \tag{0.2}$$

where $\hat{M} = M(\bar{\phi}, \hat{Y})$ and $\hat{Y} \subset B(0;1)$ is the scaled sample set.

Now, from (0.1) and (0.2),

$$\| \text{vec}(\nabla^2 f(y^0)) - \text{vec}(H)\| \;\leq\; \frac{\bar{p}^{\frac{1}{2}}\nu_2}{2}\|M^\dagger\|\Delta,$$

from which we deduce

$$\|\nabla^2 f(y^0) - H\| \;\leq\; \frac{\sqrt{2}\bar{p}^{\frac{1}{2}}\nu_2}{2}\|M^\dagger\|\Delta, \tag{0.3}$$

[1] $A^\dagger$ denotes the Moore-Penrose generalized inverse of a matrix A, which can be expressed by the singular value decomposition of A for any real or complex matrix A. In the current context where M is full column rank, we obtain the left inverse $M^\dagger = (M^\top M)^{-1}M^\top$.

and thus

$$\|\nabla^2 f(y) - H\| \ \le \ \left(\nu_2 + \frac{\sqrt{2}\bar{p}^{\frac{1}{2}}\nu_2}{2}\|M^\dagger\|\right)\Delta.$$

For the bound on the gradient, also from (0.1) and (0.2), we have

$$\|\nabla f(y^0) - g\| \ \le \ \frac{n^{\frac{1}{2}}\nu_2}{2}\|M^\dagger\|\Delta^2. \tag{0.4}$$

Hence, using $y^0 = 0$, (0.3), and (0.4), for some $t \in (0, 1)$,

$$\begin{aligned} \|\nabla f(y) - \nabla m(y)\| &= \|\nabla f(y^0) + \nabla^2 f(y^0 + ty)y - g - Hy - \nabla^2 f(y^0)y + \nabla^2 f(y^0)y\| \\ &\le \nu_2\Delta^2 + \frac{n^{\frac{1}{2}}\nu_2}{2}\|M^\dagger\|\Delta^2 + \frac{\sqrt{2}\bar{p}^{\frac{1}{2}}\nu_2}{2}\|M^\dagger\|\Delta^2 \\ &= \left(\nu_2 + \left(\frac{n^{\frac{1}{2}} + \sqrt{2}\bar{p}^{\frac{1}{2}}}{2}\right)\nu_2\|M^\dagger\|\right)\Delta^2. \end{aligned}$$

For the bound on function values, noting again that $y^0 = 0$,

$$|f(y) - m(y)| \ = \ \left|f(y^0) + \nabla f(y^0)^\top y + \frac{1}{2}y^\top\nabla^2 f(y^0)y - c - g^\top y - \frac{1}{2}y^\top Hy\right| + \frac{\nu_2}{2}\Delta^3.$$

Thus, using (0.1), (0.3), and (0.4),

$$|f(y) - m(y)| \ \le \ \left(\frac{\nu_2}{2} + \left(\frac{1}{2} + \frac{n^{\frac{1}{2}}}{2} + \frac{\sqrt{2}\bar{p}^{\frac{1}{2}}}{4}\right)\nu_2\|M^\dagger\|\right)\Delta^3.$$