MINIMIZING A CONVEX FUNCTION BY REFINING A LOWER PIECEWISE LINEAR ENVELOPE

JAN DE LEEUW AND KENNETH LANGE

ABSTRACT. Meet the abstract. This is the abstract.

1. SHARP QUADRATIC MAJORIZATION OF PIECEWISE LINEAR FUNCTIONS

Suppose b_i are n vectors in $\mathbb{R}^m$, and a_i are n scalars. Define the convex piecewise linear function

$$f(x) \triangleq \max_{1\leq i\leq n} (x'b_i + a_i).$$

Without loss of generality we assume that all b_i are different.

For each x there is a subset $\mathcal{I}(x)$ of $\{1, 2, \cdots, n\}$ such that $f(x) = x'b_i + a_i$ if and only if $i \in \mathcal{I}(x)$. Now fix $y \in \mathbb{R}^m$ and $i \in \mathcal{I}(y)$, and consider the quadratic in x

$$g_i(x \mid y) \triangleq \frac{c}{2}\|x - y\|^2 + x'b_i + a_i.$$

Also define index sets $\mathcal{K}(x \mid y) \triangleq \mathcal{I}(x) \cap \overline{\mathcal{I}(y)}$, i.e. the subset of $\mathcal{I}(x)$ that is not in $\mathcal{I}(y)$.

Theorem 1.1. *Suppose $\mathcal{K}(x \mid y) \neq \emptyset$. Then*

$$g_i(x \mid y) = \frac{c_i}{2}\|x - y\|^2 + x'b_i + a_i$$

Date: Sunday 12th September, 2010 — 22h 24min — Typeset in LUCIDA BRIGHT.

Key words and phrases. Template, LaTeX.

with

$$c_i \geq \overline{c}_i(y) \stackrel{\Delta}{=} \max_{k\in\mathcal{K}(x,y)} \frac{\frac12\|b_i-b_k\|^2}{(a_i-a_k)+y'(b_i-b_k)}$$

majorizes $f(x)$ *in* y.

Proof. Clearly $g_i(y,y)=f(y)$ for all y. It remains to show that $g_i(x\mid y)-f(x)>0$ for all $x\neq y$.

For all x and all $k\in\mathcal{I}(x)$,

$$g_i(x\mid y)-f(x)=\frac{c}{2}\|x-y\|^2+x'(b_i-b_k)+(a_i-a_k).$$

Note the right-hand side has the same numerical values, no matter how we choose $k\in\mathcal{I}(x)$. Thus if we show that it is non-negative for any value of k, then it will be non-negative for all values of k.

$$g_i(x\mid y)-f(x)=\frac{c}{2}\|(x-y)+\frac{1}{c}(b_i-b_k)\|^2+\\ +(a_i-a_k)+y'(b_i-b_k)-\frac12\frac1c\|b_i-b_k\|^2.$$

If $k\in\mathcal{K}(x\mid y)$ we have

$$(a_i-a_k)+y'(b_i-b_k)=f(y)-(a_k+y'b_k)>0.$$

It follows that if

$$c\geq\frac{\frac12\|b_i-b_k\|^2}{(a_i-a_k)+y'(b_i-b_k)}$$

we have $g_i(x\mid y)-f(x)\geq 0$.

If $k\notin\mathcal{K}(x\mid y)$, i.e. $k\in\mathcal{I}(x)\cap\mathcal{I}(y)$, we have $(a_i-a_k)+y'(b_i-b_k)=0$. If $k=i$ then $g_i(x\mid y)-f(x)=\frac{c}{2}\|x-y\|^2>0$ for all $c>0$. If $k\neq i$ □

Now let $\mathcal{A}(y)=\{k\mid f(y)>(a_k+y'b_k)\}$. Note that $\mathcal{I}(y)\cup\mathcal{A}(y)=\emptyset$ and $\mathcal{K}(x,y)\subseteq\mathcal{A}(y)$. Define

$$\overline{\overline{c}}_i(y)\stackrel{\Delta}{=}\max_{k\in\mathcal{A}(y)}\frac{\frac12\|b_i-b_k\|^2}{(a_i-a_k)+y'(b_i-b_k)}$$

Corollary 1.2. $g_i(x\mid y)=\frac12\overline{\overline{c}}_i(y)\|x-y\|^2+x'b_i+a_i$ *majorizes* $f(x)$ *at* y.

Proof. It suffices to note that $\overline{\overline{c}}_i(y) \geq \overline{c}_i(y)$. □

2. Examples

If $f(x) = |x-1| = \max(x-1, -x+1)$ we have $b_1 = 1$, $b_2 = -1$, $a_1 = -1$, $a_2 = 1$. If $x > 1$ we have $\mathcal{I}(x) = \{1\}$, if $x < 1$ we have $\mathcal{I}(x) = \{2\}$, and if $x = 1$ we have $\mathcal{I}(x) = \{1, 2\}$.

Now take $y = \frac{1}{2}$. Thus $\mathcal{I}(y) = \{2\}$ and $\mathcal{K}(x, y) \neq \emptyset$ if and only if $x \leq 1$, in which case $\mathcal{K}(x, y) = \mathcal{A}(y) = \{1\}$. Thus $\overline{c}_i(y) = \overline{\overline{c}}_i(y) = 2$, and $g(x, y) = \|x - \frac{1}{2}\|^2 - x + 1$.

If $y = 1$ then $\mathcal{K}(x, y) = \emptyset$ for all x. Thus

$$g_1(x \mid y) = \frac{c}{2}(x-1)^2 + x - 1,$$
$$g_2(x \mid y) = \frac{c}{2}(x-1)^2 - x + 1.$$

If $c > 0$ both are convex quadratics with two distinct real roots. Between their two roots they are negative, and do not majorize.

1900 75.995 1910 91.972 1900 105.711 1930 123.203 1940 131.669 1950 150.697 1960 179.323 1970 203.212 1980 226.505 1990 249.633 2000 281.422

3. Refining the Lower Envelope

If h is any convex function and $y_1, \ldots, y_n$ are points then

$$h(x) \geq f(x) = \max_{1 \leq i \leq n} h(y_i) + \mathcal{D}h(y_i)(x - y_i),$$

where $\mathcal{D}h(y_i)$ is any point in the subgradient. The result above can be used to minimize f (or just to do one single quadratic step), which generates an new point y_{n+1} approximating the minimum of h. Use this point as an additional linearization point. Updating the c_i can be done fairly simply.

4. Examples

4.1. WLS. Let

$$h(\beta)=\frac{1}{2}(y-X\beta)'W(y-X\beta),$$

so that

$$\mathcal{D}h(\beta)=X'W(y-X\beta).$$

4.2. LAV approximation.

4.3. Maximum Likelihood Logistic Regression. We have

$$h(\beta)=\sum_{i=1}^{n}N_i(p_ix_i'\beta+\log(1-\pi_i(\beta)),$$

where

$$\pi_i(\beta)=\frac{1}{1+\exp(-x_i'\beta)}.$$

Thus

$$\mathcal{D}h(\beta)=\sum_{i=1}^{n}N_i(\pi_i(\beta)-p_i)x_i.$$

4.4. Indefinite QP. Minimize

$$h(x)=a+b'x+\frac{1}{2}x'Cx$$

over $x\in\mathcal{X}$. Separating the positive and negative eigenvalues of C we can write $C=C_+-C_-$, with both C_+ and C_- positive semi-definite.

4.5. MDS. MDS can be formulated as minimizing

$$h(x)=\frac{1}{2}x'x-\sum_{i=1}^{n}w_i\delta_i\sqrt{x'A_ix},$$

where $\sum_{i=1}^{n}w_iA_i=I$.

5. Generalization

$$Lf(x) \triangleq \max_{t \in \mathcal{T}}(x'b(t) + a(t)),$$

with t some compact set.

Appendix A. Result

Theorem A.1. *Suppose $f(x) = \max(x'p, x'q)$, with $p \neq q$, and y satisfies $y'p = y'q = f(y)$. There is no quadratic majorizing f at y.*

Proof. Define

$$g_1(x) \triangleq \frac{1}{2}x'Ax + (b-p)'x + c,$$
$$g_2(x) \triangleq \frac{1}{2}x'Ax + (b-q)'x + c.$$

and

$$h(x) = g(x) - f(x) = \min_x(g_1(x), g_2(x)).$$

We want $h(x)$ to have a minimum equal to zero at y.

Of course h is not differentable at y, but we can compute the directional derivatives.

$$h(y + \epsilon\delta) = h(y) + \epsilon\left(\delta'Ay + \min(\delta'(b-p), \delta'(b-q))\right) + o(\epsilon).$$

For y to be a minimum we must have $\delta'Ay + \min(\delta'(b-p), \delta'(b-q)) \geq 0$ for all δ. In order words we must have $\delta'(Ay + (b-p)) \geq 0$ for all δ such that $\delta'(b-q) \geq \delta'(b-p)$, i.e. for all δ such that $\delta'(p-q) \geq 0$. This means there must be a $\lambda > 0$ such that

$$Ay + (b-p) = \lambda(p-q).$$

In the same way there must be a $\gamma > 0$ such that

$$Ay + (b-q) = \gamma(q-p).$$

It follows that

$$\lambda = \frac{(p-q)'A^{-1}(b-p)}{(p-q)A^{-1}(p-q)},$$
$$\gamma = -\frac{(p-q)'A^{-1}(b-q)}{(p-q)A^{-1}(p-q)},$$

and $\lambda + \gamma = -1$. □

Appendix B. Ben-Tal and Zowe

A convex quadratic $g(x) = \frac{1}{2}x'Ax + b'x + c$ majorizes $f(x) = \max_{i=1}^{n} p_i'x + q_i$ at y if $g(x) \geq f(x)$ for all x and $g(y) = f(y)$. In other words $h(x) \triangleq g(x) - f(x)$ has a minimum equal to zero at y.

If

$$h_i(x) = \frac{1}{2}x'Ax + (b - p_i)'x + (c - q_i),$$

then $h(x) = \min_{i=1}^{n} h_i(x)$. Define

$$I(x) \triangleq \{i \mid h(x) = h_i(x)\},$$
$$h'(x,d) \triangleq \min_{i \in I(x)} d'(Ax + (b - p_i)),$$
$$I(x,d) \triangleq \{i \in I(x) \mid h'(x,d) = d'(Ax + (b - p_i))\}.$$

A necessary condition for x to be a local minimum is that for all d for which $h'(x,d) \leq 0$ there exist non-negative multipliers λ_i, with $i \in I(x,d)$, adding up to one, such that

$$Ax + b = \sum_{i \in I(x,d)} \lambda_i p_i.$$

If $n = 2$ we want to find out if y with $f(y) = p_1'y + q_1 = p_2'y + q_2$ is a local minimum. Thus $I(y) = \{1,2\}$, and

$$h'(y,d) = \min(d'(Ax + (b - p_1)), d'(Ax + (b - p_2)))$$

Department of Statistics, University of California, Los Angeles, CA 90095-1554

Department of Statistics, University of California, Los Angeles, CA 90095-1554

E-mail address, Jan de Leeuw: deleeuw@stat.ucla.edu

URL, Jan de Leeuw: http://gifi.stat.ucla.edu