TMA 4180 Optimeringsteori

The Exact Solution of the Trust Region Quadratic Problem

H. E. Krogstad, Spring 2010/12

The necessary and sufficient conditions for the solution of the spherical trust region quadratic problem are stated in N&W Theorem 4.1, but its proof is not so simple. Actually, the problem is a *constrained optimization problem,* contrary to the unconstrained problems we have looked at so far.

Below is an attempt to summarize the arguments. Recall that we consider

$$\min_{x\in\mathbb{R}^n} f(p). \tag{1}$$

Standing at a point x_k we are in the center of a *trust region*, $\mathcal{D}$, typically a ball with radius Δ, where we have a "trustworthy" approximation m to f:

$$f(x_k+p) \approx m(p) = f(x_k) + b'p + \frac{1}{2}p'Bp. \tag{2}$$

It is reasonable to let $b = \nabla f(x_k)$, and B be somewhat similar to $\nabla^2 f(x_k)$. The simplified problem, which is the core of the Trust Region algorithm, is now

$$\min_{p\in\mathcal{D}} m(p), \tag{3}$$

where

$$m(p) = b'p + \frac{1}{2}p'Bp, \tag{4}$$

$$\mathcal{D} = \{p; \|p\| \le \Delta\}. \tag{5}$$

The object function is *quadratic* and the domain is a *ball*. The matrix B is assumed to be symmetric (since skew symmetric parts will not contribute to $p'Bp$ in any case!).

Since m is continuous and $\mathcal{D}$ is closed and bounded, we know that there *always* exist minima. We also know that

$$\nabla m(p)' = b + Bp, \tag{6}$$

$$\nabla^2 m(p) = B, \tag{7}$$

and the Taylor expansion around a minimum p^* (or any fixed point) has the form

$$m(p^*+\delta) = m(p^*) + \nabla m(p^*)\,\delta + \frac{1}{2}\delta' B\delta. \tag{8}$$

Note that we do not require, or need, that B is positive semi-definite. This, and the constraint in Eqn. 5 complicates the solution considerably compared to the standard unconstrained quadratic model problem.

1 Necessary and sufficient conditions for minima

We first observe that there are two possibilities for a global minimum p^*. Either p^* is in the interior of $\mathcal{D}$, or it is on the boundary.

If p^* is in the interior of $\mathcal{D}$, we clearly need that $\nabla m(p^*) = 0$. Moreover, from the simple form of $m(p)$, it is necessary and sufficient that $\nabla m(p^*) = 0$ and $B \geq 0$.

For a p^* on the boundary of $\mathcal{D}$ we may still have that $\nabla m(p^*) = 0$, and then, from Eqn. 8, we must have $\delta' B \delta \geq 0$, for all vectors δ pointing into $\mathcal{D}$. However, since $\delta' B \delta = (-\delta)' B (-\delta)$, this must actually be the case for all δ-s, and hence B has to be positive semi-definite (*No*, it can not be strictly negative even for vectors that are tangents to the boundary! Prove it yourself, or see the argument in the appendix).

We then look at the case when p^* is on the boundary and $\nabla m(p^*)' = b + Bp^*$ is different from 0. In general, all directions within $\pi/2$ of the negative gradient are descent directions, and since the vector $p^* \in \partial\mathcal{D}$ at the same time is an outward normal vector to the boundary, $\nabla m(p^*)$ has to point *exactly* opposite to p^* at a minimum (Make a simple sketch for $\mathbb{R}^2$ if you do not see this!). Thus, p^* is a minimum *only* if $\nabla m(p^*)$ is proportional to $-p^*$, or, equivalently; there is a $\lambda > 0$ so that

$$b + Bp^* = -\lambda p^*. \tag{9}$$

This may also be written $(B + \lambda I) p^* = -b$. It is even possible to show that in this case, $B + \lambda I \geq 0$, but the argument in N&W is tricky and reproduced in the Appendix at the end of the note.

For the converse, let us assume that we have found a $\lambda > 0$ such that Eqn. 9 holds, $B + \lambda I \geq 0$, and $\|p^*\| = \Delta$. We need to prove that p^* is a global minimum. Consider the function

$$m_\lambda(p) = m(p) + \frac{\lambda}{2}\left(p'p - p^{*\prime}p^*\right), \tag{10}$$

which is quadratic, and

$$\nabla m_\lambda(p^*) = b + (B + \lambda I) p^* = 0, \tag{11}$$

$$\nabla^2 m_\lambda(p^*) = B + \lambda I \geq 0. \tag{12}$$

Thus, according to what we already have proved above, p^* *is* indeed a minimum for $m_\lambda(p)$ in $\mathcal{D}$. But then

$$m(p^*) = m_\lambda(p^*) \leq m_\lambda(p) = m(p) + \frac{\lambda}{2}\left(p'p - p^{*\prime}p^*\right) \leq m(p), \tag{13}$$

since $\|p\| \leq \|p^*\|$ for all $p \in \mathcal{D}$. Thus, p^* is a global minimum for $m(p)$ as well!

In summary, we obtain Thm. 4.1 in N&W in the alternate formulation:

The vector p^* is a global minimum of the Trust Region Quadratic Problem if and only if one of the following conditions hold:

$$\begin{array}{ll} (i) & b + Bp^* = 0,\ B \geq 0, \|p^*\| \leq \Delta, \\ (ii) & \text{There exists a } \lambda > 0 \text{ such that } (B + \lambda I) p^* = -b, \\ & B + \lambda I \geq 0,\ \|p^*\| = \Delta. \end{array} \tag{14}$$

2 How to find the solution

First of all, we may be lucky and find that (i) holds, e.g. $B > 0$ and $\|p^*\| = \|B^{-1}(-b)\| \leq \Delta$. If we are not so lucky, perhaps the equation for p^* in (i) cannot be solved, we need to consider (ii). Then there is an extra parameter in our problem, namely λ, which we later will identify as a *Lagrange parameter* coming from the inequality constraint (Eqn. 5).

In order to analyze these cases, we recall that B was supposed to be symmetric, so it has n real eigenvalues, $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$, and a corresponding set of normalized, orthogonal eigenvectors $\{e_j\}_{j=1}^n$.

When $\lambda_j + \lambda \neq 0$ for all j, $B + \lambda I$ will be non-singular and we can solve Eqn. 9,

$$p_\lambda = (B + \lambda I)^{-1}(-b). \tag{15}$$

Show that the solution in Eqn. 15 may be expressed as

$$p_\lambda = \sum_{j=1}^{n} \left(\frac{e_j'(-b)}{\lambda_j + \lambda} \right) e_j \tag{16}$$

and hence

$$\|p_\lambda\|^2 = \sum_{j=1}^{n} \left(\frac{e_j'(-b)}{\lambda_j + \lambda} \right)^2. \tag{17}$$

We have to look for a solution λ^* in the interval $[-\lambda_1, \infty)$, since this will ensure that $B + \lambda I \geq 0$. Moreover, it is required that $\|p_{\lambda^*}\| = \Delta$. What happens with $\|p_\lambda\|^2$ when $\lambda \to \infty$? And, assuming that $e_1'(-b) \neq 0$, when λ approaches $-\lambda_1$ from above? Draw a sketch of the RHS of Eqn. 17, or look at Fig.4.5 in N&W. Since the right hand side of Eqn. 17 is a continuous function in the interval $(-\lambda_1, \infty)$, there has to be a λ^* so that $\|p_\lambda\| = \Delta$. Finding λ^* thus amounts to find a solution of a non-linear equation. There is one little snag left in this argument: What happens if $B > 0$, so that $-\lambda_1 < 0$? Our solution requires $\lambda^* > 0$. Try to fill in the details by considering p_λ for $\lambda = 0$.

In exceptional cases, $e_1'(-b) = 0$. We then have to solve Eqn. 9 on the subspace spanned by $\{e_2, \cdots, e_n\}$, or, if necessary, choose $\lambda^* = -\lambda_1$ and add some contribution from e_1 so that $\|p_\lambda\| = \Delta$ (This is *The Hard Case*, p. 87 in N&W).

The above covers much of the discussion in Chapter 4, and we shall leave the general problem here. In practice, it appears to be more reasonable to apply some of the simplified, but approximate solutions to the quadratic problem, as described in N&W §4.1.

3 Appendix: $B + \lambda I \geq 0$.

We assume that p^* is a global minimum for $m(p)$, $\|p^*\| = \Delta$, $\lambda > 0$, and $(B + \lambda I) p^* = -b$. Following N&W, we consider the following expression for $p \in \partial\mathcal{D}$ (and note that then, $p^{*\prime} p^* - p' p = 0$):

$$\begin{aligned}
0 &\leq m(p) - m(p^*) - \frac{\lambda}{2}\left(p^{*\prime}p^* - p'p\right) \\
&= b'p + \frac{1}{2}p'Bp - b'p^* - \frac{1}{2}p^{*\prime}Bp^* - \frac{\lambda}{2}\left(p^{*\prime}p^* - p'p\right) \\
&= -p'(B+\lambda I)p^* + \frac{1}{2}p'Bp + p^{*\prime}(B+\lambda I)p^* - \frac{1}{2}p^{*\prime}Bp^* - \frac{\lambda}{2}\left(p^{*\prime}p^* - p'p\right) \\
&= -p'(B+\lambda I)p^* + \frac{1}{2}p'(B+\lambda I)p + \frac{1}{2}p^{*\prime}(B+\lambda I)p^* \\
&= \frac{1}{2}(p-p^*)'(B+\lambda I)(p-p^*)
\end{aligned}$$

The proof will be complete if we can show that $x'(B + \lambda I)x \geq 0$ for all $x \in \mathbb{R}^n$. The inequality will hold for all vectors proportional to $w = \pm(p - p^*)$ for some $p \in \partial\mathcal{D}$, and this includes *all* vectors that are not orthogonal to p^*: Assume that $x'p^* \neq 0$. Then

$$p = p^* - \frac{2x'p^*}{\|x\|^2}x \in \partial\mathcal{D},$$

and

$$x = \frac{\|x\|^2}{2x'p^*}(p^* - p).$$

Finally, let $y'p^* = 0$. Then $y'(B + \lambda I)y = \lim_{\varepsilon\to 0}(y + \varepsilon p^*)'(B + \lambda I)(y + \varepsilon p^*)$. However, $(y + \varepsilon p^*)'(B + \lambda I)(y + \varepsilon p^*) \geq 0$ for all $\varepsilon \neq 0$, so in the limit, $y'(B + \lambda I)y \geq 0$!

May be you see a simpler argument?