On Methods for a Trust Region SR1 Algorithm

Johannes Brust and Prof. Roummel Marcia
Applied Mathematics, UC Merced

SDSC, California, USA

July 9, 2015
Problem Formulation

We consider solving the minimization problem

\[
\min_{x \in \mathbb{R}^n} f(x),
\]

where \( f : \mathbb{R}^n \to \mathbb{R} \), using a Trust-Region algorithm.
Notation

- A bold lower letter will be a vector: \( \mathbf{x} \).
- A bold upper letter will be a matrix: \( \mathbf{X} \).
- \( k \) represents an iteration index.
- \( n \) represents the problem dimension.
- \( \| (\cdot) \|_2 \) represents the Euclidean norm:
  \[
  \| \mathbf{x} \|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}.
  \]
- positive-semidefinite: \( \mathbf{x}^T \mathbf{X} \mathbf{x} \geq 0 \), for any \( \mathbf{x} \).
A Trust-Region Algorithm

For $k = 1, 2, \ldots$

1. Solve the Trust-Region subproblem to obtain $s_k$:

   $$\min_{||s||_2 \leq \Delta_k} s^T g_k + \frac{1}{2} s^T B_k s.$$

2. Set $x_{k+1} = x_k + s_k$.

3. Update $\Delta_k$ (Radius).

4. Update $B_k, x_k$.

Here

$$g_k = \nabla f(x_k), \quad B_k \approx \nabla^2 f(x_k), \quad \text{and} \quad \Delta_k > 0.$$
Subproblem Optimality Conditions

Moré and Sorensen (1983): For some $\sigma^*$, a solution $s^*$ of the Trust-Region subproblem satisfies

$$(B_k + \sigma^* I_n)s^* = -g_k,$$

$$\sigma^* \cdot (||s^*||_2 - \Delta_k) = 0,$$

$$\sigma^* \geq 0,$$

$$B_k + \sigma^* I_n \text{ is positive semidefinite}.$$ 

In the remainder we solve for the optimal pair $(s^*, \sigma^*)$, when $B_k$ is not positive semidefinite.
Pro et Contra of the Method

- By introducing a constrained subproblem the method allows for flexibility, e.g:
  - The model Hessian, $B_k$, may have any definiteness state.
  - The step, $s_k$, may take any direction.

- However, each step is required to satisfy a set of coupled conditions.

- It is a challenge to satisfy these conditions efficiently.
The Symmetric Rank 1 Update

Let

$$s_k = x_{k+1} - x_k,$$
$$y_k = g_{k+1} - g_k.$$

Then the SR1 update is defined as

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k},$$

provided $$(y_k - B_k s_k)^T s_k \neq 0.$$ 

Observe that $B_{k+1}$ is

(1) symmetric if $B_k$ is symmetric.
(2) generated by a rank-1 update.
(3) not guaranteed to be positive definite.
Representation of $B_k$

Nocedal, Byrd & Schnabel (1994): The compact SR1 representation is given by

$$B_k = \gamma_k I + V_k W_k V_k^T,$$

where $\gamma_k \in \mathbb{R}$, and

$$V_k = (Y_k - \gamma_k S_k), \quad \in \mathbb{R}^{n \times m},$$

$$W_k = (D_k + L_k + L_k^T - \gamma_k S_k^T S_k)^{-1}, \quad \in \mathbb{R}^{m \times m},$$

$$S_k = [s_{k-1-m} \ s_{k-m} \ldots \ s_{k-1}], \quad \in \mathbb{R}^{n \times m},$$

$$Y_k = [y_{k-1-m} \ y_{k-m} \ldots \ y_{k-1}], \quad \in \mathbb{R}^{n \times m},$$

$$(L_k)_{i,j} = s_i^T y_j, \quad \text{for } i > j, \quad \in \mathbb{R}^{m \times m},$$

$$(D_k)_{i,j} = s_i^T y_j, \quad \text{for } i = j, \quad \in \mathbb{R}^{m \times m}.$$

For notational simplicity we will drop the index $k$. 
An Eigendecomposition of $B$

Since $B$ is symmetric it may be represented as

$$B = P \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} P^T = P \Lambda P^T,$$

where

$$PP^T = P^T P = I, \quad \Lambda = \text{diag}(\lambda_i), \quad \lambda_1 \leq \lambda_2 \cdots \leq \lambda_n.$$

Observe that when $B$ is indefinite, then

$$\lambda_1 < 0.$$
Subproblem Optimality

For the first subproblem optimality condition we have

\[-g = (B + \sigma I)s\]
\[= (P\Lambda P^T + \sigma I)s\]
\[= (P\Lambda P^T + \sigma PP^T)s\]
\[= P(\Lambda + \sigma I)P^T s\]
\[= P(\Lambda + \sigma I)v,\]

where \(P^Ts = v\). Then

\[-P^Tg = (\Lambda + \sigma I)v.\]

For the second condition, since \(P\) is orthogonal and since \(s = Pv\),

\[\|s\|_2 = \|Pv\|_2 = \|v\|_2.\]
Subproblem Optimality

Now the subproblem optimality conditions can be expressed equivalently in terms of \((s^*, \sigma^*)\) or \((v^*, \sigma^*)\):

\[
\begin{align*}
(B + \sigma^* I)s^* &= -g \\
\sigma^*(||s^*||_2 - \Delta) &= 0 \\
\sigma^* &\geq 0 \\
B + \sigma^* I &\geq 0
\end{align*}
\quad \begin{align*}
(\Lambda + \sigma^* I)v^* &= -P^T g \equiv a \\
\sigma^*(||v^*||_2 - \Delta) &= 0 \\
\sigma^* &\geq 0 \\
\lambda_i + \sigma^* &\geq 0 \text{ for all } i.
\end{align*}

Notice that the optimality conditions on the right have a simplified form, mainly because \(\Lambda + \sigma^* I\) is a diagonal matrix.
Method Of The Largest Root

The solution of the subproblem will be obtained by satisfying the optimality conditions, with $\lambda_1 < 0$.

Now the condition

$$\lambda_i + \sigma^* \geq 0,$$

for all $i$

is implied by

$$\lambda_1 + \sigma^* \geq 0,$$

and therefore

$$\sigma^* > 0.$$

With this, the 2nd condition

$$\sigma^* \cdot (\|v^*\|_2 - \Delta) = 0,$$

requires

$$\|v^*\|_2 = \Delta.$$
Method Of The Largest Root

From the first condition

$$(\Lambda + \sigma^* I)v^* = a,$$

we have

$$\|v^*\|_2 = \|(\Lambda + \sigma^* I)^{-1}a\|_2 = \sqrt{\sum_{i=1}^{n} \left( \frac{a_i}{\lambda_i + \sigma^*} \right)^2},$$

and therefore, using the 2nd condition,

$$\Delta = \sqrt{\sum_{i=1}^{n} \left( \frac{a_i}{\lambda_i + \sigma^*} \right)^2}.$$

To solve the nonlinear equation let

$$\tilde{\phi}(\sigma) = \sqrt{\sum_{i=1}^{n} \left( \frac{a_i}{\lambda_i + \sigma} \right)^2} - \Delta,$$

and seek its zeros.
Typical Graph of $\tilde{\phi}(\sigma)$

Note that $\lim_{\sigma \to -\lambda_i} \tilde{\phi}(\sigma) = \infty$ and $\lim_{|\sigma| \to \infty} \tilde{\phi}(\sigma) = -\Delta$. This function is singular at the negative eigenvalues.
Method of The Largest Root : Alternative objective

To avoid the singularities, and since

\[ \tilde{\phi}(\sigma^*) = \sqrt{\sum_{i=1}^{n} \left( \frac{a_i}{\lambda_i + \sigma^*} \right)^2} - \Delta = 0, \]

implies

\[ \sqrt{\sum_{i=1}^{n} \left( \frac{a_i}{\lambda_i + \sigma^*} \right)^2} = \frac{1}{\Delta}, \]

introduce the alternative objective

\[ \phi(\sigma) = \frac{1}{\sqrt{\sum_{i=1}^{n} \left( \frac{a_i}{\lambda_i + \sigma} \right)^2}} - \frac{1}{\Delta}. \]

Thus the largest root, \( \sigma^* \), satisfies \( \tilde{\phi}(\sigma^*) = \phi(\sigma^*) = 0. \)
Note that \( \lim_{\sigma \to -\lambda_i} \phi(\sigma) = -\frac{1}{\Delta} \) and \( \lim_{|\sigma| \to \infty} \phi(\sigma) = \infty \). Here the largest root, \( \sigma^* \), satisfies \( \sigma^* \geq -\lambda_1 \).
Method Of The Largest Root

For $\sigma \in (-\lambda_1, \infty)$

$$
\phi'(\sigma) = \frac{\sum_{i=1}^{n} \frac{a_i^2}{(\lambda_i+\sigma)^3}}{\left[ \sum_{i=1}^{n} \left( \frac{a_i}{\lambda_i+\sigma} \right)^2 \right]^{\frac{3}{2}}} > 0,
$$

and when

$$
\phi(-\lambda_1) = -\frac{1}{\Delta},
$$

there will consequently be a largest root, $\sigma^* > -\lambda_1$, such that

$$
\phi(\sigma^*) = 0, \quad \lambda_1 + \sigma^* > 0.
$$

This method reduces the Trust-Region subproblem to a root finding problem, whenever applicable.
The Hard Case

Suppose that the largest root satisfies

$$\sigma^* > 0 \quad \text{but} \quad \sigma^* + \lambda_1 < 0.$$  

Hence the optimality condition

$$\sigma^* + \lambda_i \geq 0 \quad \text{for all} \ i,$$

is not satisfied.

$$\Rightarrow \quad \text{The Method Of The Largest Root is not applicable.}$$

This case will be called the Hard Case, in accordance with a description of Moré and Sorensen (1983)
The Hard Case

A necessary and sufficient condition for the Hard Case is

\[ \phi(-\lambda_1) > 0. \]

**Sufficiency :**

(1) By the monotonicity of \( \phi' \in (-\lambda_1, \infty) \) no root larger than \(-\lambda_1\) occurs.

(2) If not \( a_i = 0, i = 2, 3, \ldots, n \), then

\[ \phi(-\lambda_i) = -\frac{1}{\Delta}. \]

(3) Thus the largest root, \( \sigma^* \in (-\infty, -\lambda_1) \).
The Hard Case

A necessary and sufficient condition for the Hard Case is

$$\phi(-\lambda_1) > 0.$$  

Necessity:

(1) By contradiction assume

$$\phi(-\lambda_1) \leq 0.$$  

(2) Then, from the monotonicity of $\phi'$, there will be a $\sigma^*$ with $\phi(\sigma^*) = 0$ such that

$$\sigma^* \geq -\lambda_1 \quad \text{or} \quad \sigma^* + \lambda_1 \geq 0.$$  

(3) Conclude that the Hard Case, $\sigma^* + \lambda_1 < 0$, does not occur.
The Hard Case

If we recall

\[ \phi(\sigma) = \frac{1}{\sqrt{\sum_{i=1}^{n} \left( \frac{a_i}{\lambda_i + \sigma} \right)^2}} - \frac{1}{\Delta}, \]

then the condition

\[ \phi(-\lambda_1) > 0, \]

implies

\[ a_1 = 0. \]

In this case, we propose a method to solve the optimality conditions based on the smallest eigenvalue.
Method Of The Smallest Eigenvalue

First, we solve the condition $\sigma^* + \lambda_i \geq 0$ by setting

$$\sigma^* = -\lambda_1.$$

Then the 1st condition

$$(\Lambda + \sigma^* I)v^* = a,$$

becomes

$$0 \cdot v_1^* = 0$$

$$(\lambda_2 - \lambda_1) \cdot v_2^* = a_2$$

$$\vdots$$

$$(\lambda_n - \lambda_1) \cdot v_n^* = a_n,$$

or

$$v_i^* = \frac{a_i}{\lambda_i - \lambda_1} \quad \text{for} \quad i = 2, 3, \ldots, n.$$
Method Of The Smallest Eigenvalue

Given $v_2^*, v_3^*, \ldots, v_n^*$, we use the 2nd condition

$$\|v^*\|_2 = \Delta,$$

to obtain

$$(v_1^*)^2 + (v_2^*)^2 + \cdots + (v_n^*)^2 = \Delta^2.$$

Therefore we may compute

$$v_1^* = \sqrt{\Delta^2 - (v_2^*)^2 - \cdots - (v_n^*)^2}.$$  

This yields a solution for the optimality conditions in the Hard Case.
Method Of The Smallest Eigenvalue

The computation of $v^*$ can be summarized by introducing a pseudoinverse, $(.)^\dagger$. That is

$$v^* = (\Lambda - \lambda_1 I)^\dagger a,$$

where the elements of $v^*$ are computed as described.

In conclusion an optimal Trust Region step can now be computed as

$$s^* = Pv^* = P(\Lambda - \lambda_1 I)^\dagger a,$$

using the just described method.
Numerical Experiments

In our experiments $S, Y, \in \mathbb{R}^{n \times m}$ and $g \in \mathbb{R}^{n}$ are generated randomly from a zero-mean Gaussian distribution.

Parameters:

- Convergence tolerance: $\epsilon = 1.0 \times 10^{-10}$.
- Trust-Region radius: $\Delta = 1$.
- $\gamma = 0.5$ in $B = \gamma I + V \Sigma V^T$.
- $m = 5$.

For the experiments, $n$ varies from $n = 10^1$ to $n = 10^6$.

We ran the experiment 10 times and report one representative experiment.
Results of Numerical Experiments

Let $\tau = \|(B + \sigma^* I)s^* - (-g)\|_2$.

| $n$     | $\tau$       | $|\phi(\sigma^*)|$ | $\lambda_1 + \sigma^*$ | Iter | Time  |
|---------|--------------|---------------------|-------------------------|------|-------|
| 10      | $2.3 \times 10^{-15}$ | $1.0 \times 10^{-15}$ | 1.1                     | 4    | 0.04  |
| 100     | $2.2 \times 10^{-15}$ | $1.0 \times 10^{-15}$ | 4.9                     | 5    | 0.02  |
| 1000    | $3.6 \times 10^{-15}$ | $4.0 \times 10^{-11}$ | 24.1                    | 4    | 0.03  |
| 10000   | $1.1 \times 10^{-14}$ | $1.0 \times 10^{-14}$ | 94.5                    | 4    | 0.03  |
| 100000  | $3.4 \times 10^{-14}$ | $1.0 \times 10^{-15}$ | 312.2                   | 4    | 0.07  |
| 1000000 | $1.2 \times 10^{-13}$ | $1.0 \times 10^{-15}$ | 995.0                   | 4    | 0.42  |

Note that

- $\|(B + \sigma^* I)s^* - (-g)\|_2 \approx 0$,
- $|\phi(\sigma^*)| \approx 0$,
- $\lambda_1 + \sigma^* \geq 0$. 
Conclusion

- We presented an SR1 Trust-Region algorithm.

- We developed a method for computing the solution to the optimality conditions when $\lambda_1 < 0$, The Method Of The Largest Root.

- We developed a method to compute the solution to the optimality conditions in the Hard-Case, The Method Of The Smallest Eigenvalue.
We derive formulas for the computation of $\gamma_k$ in the compact representation of $\mathbf{B}_k$ based on minimizing $\kappa(\mathbf{B}_{k+1})$.

Given satisfactory $\gamma_k$ we intend to implement and test the full Trust-Region method for the compact SR1 matrix.

Currently we study mathematical techniques for simulations in molecular dynamics.
Questions?

jbrust@ucmerced.edu

\[ \mathbf{Q}^? = \mathbf{Q} \begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix} \mathbf{Q}^T. \]
References

(1) Burdakov, Gong, Yuan, Zikrin (2013), "On Efficiently Combining Limited Memory And Trust Region Techniques".

(2) Byrd, Nocedal, Schnabel (1994), "Representations Of Quasi-Newton Matrices And Their Use In Limited Memory Methods".

(3) Moré, Sorensen (1983), "Computing A Trust Region Step".