Lecture 19

Global Methods

October 28, 2008
Outline

- Line Search Algorithm
  - Gauss-Newton

- Application to a Complementarity Problem

- Trust Region Methods
Line Search for $B$-Differentiable Problems

**Step 0** Choose $x^0 \in X$ and $\gamma \in (0, 1)$. Set $k = 0$.

**Step 1** If $x^k$ is $B$-stationary point, stop.

**Step 2** Choose a symmetric positive definite matrix $H^k$, and determine a vector $d^k$ solving the direction-search problem with $x = x^k$ and $H = H^k$. Set $i = 0$.

**Step 2(i)** If

$$\theta(x^k + d^k/2^i) \leq \theta(x^k) + \frac{\gamma}{2i} \theta'(x^k; d^k),$$

set $i_k = i$ and $\tau_k = 2^i$. Otherwise, set $i := i + 1$ and repeat Step 2(i).

**Step 3** Set $x^{k+1} = x^k + \tau_k d^k$, $k := k + 1$ and go to Step 1.
Convergence Result

**Proposition 8.3.7.** Let $X \subseteq \mathbb{R}^n$ be a closed convex set and $\theta$ be $B$-differentiable on $X$. Let $\{x^k\}$ be a sequence generated by the Line Search Algorithm. Let $\{x^k \mid k \in \mathcal{K}\}$ be a subsequence such that

(a) There exist positive scalars $c_1$ and $c_2$ satisfying for every $k \in \mathcal{K}$,
\[
    c_1\|y\|^2 \leq y^T H^k y \leq c_2\|y\|^2 \quad \text{for all } y \in \mathbb{R}^n
\]

(b) The subsequence $\{x^k \mid k \in \mathcal{K}\}$ converges to a vector $x^*$

(c) The function $\theta$ has a strong $F$-derivative at $x^*$, i.e.,
\[
    \lim_{\substack{y \neq z \to (x^*,x^*)}} \frac{e(y) - e(z)}{\|y - z\|} = 0,
\]
where $e(y) = \theta(y) - \theta(x^*) - \nabla \theta(x^*)^T(y - x^*)$

Then, $x^*$ is a $B$-stationary point of the minimization problem $\min_{x \in X} \theta(x)$. 

Game Theory: Models, Algorithms and Applications
Case of Differentiable $\theta$

Let $X \subseteq \mathbb{R}^n$ be closed and convex and $\theta$ be continuously differentiable. Assume that we choose matrices $H^k$ so that for some scalars $c_1 > 0$ and $c_2 > 0$, and for all $k$, there holds:

$$c_1 \|y\|^2 \leq y^T H^k y \leq c_2 \|y\|^2 \quad \text{for all } y \in \mathbb{R}^n.$$  

Let $\{x^k\}$ be generated by Line Search Algorithm with such matrices $H^k$. Then, by Proposition 8.3.7, every accumulation point $\tilde{x}$ of $\{x^k\}$ is a stationary point of the problem $\min_{x \in X} \theta(x)$:

$$\nabla \theta(\tilde{x})^T (x - \tilde{x}) \geq 0 \quad \text{for all } x \in X.$$  

Caution: Proposition says nothing about the existence of an accumulation point of $\{x^k\}$. It merely says what property an accumulation point would have if it existed.
Gauss-Newton Method

We apply the Line Search Algorithm to find an (unconstrained) zero of the following system

\[ G(x) = 0 \quad x \in \mathbb{R}^n \]

where \( G : \mathbb{R}^n \rightarrow \mathbb{R}^n \) is continuously differentiable.

We define \( \theta(x) \) by

\[ \theta(x) = \frac{1}{2} \| G(x) \|^2 = \frac{1}{2} G(x)^T G(x) \]

We discuss how to exploit the special structure of \( \theta \) for defining the matrix \( H \) in the direction-search problem

\[
\begin{align*}
\text{minimize} & \quad \theta'(x; d) + \frac{1}{2} d^T H d \\
\text{subject to} & \quad d \in X - x
\end{align*}
\]

The modification we are about to see is a modification of well-known Gauss-Newton method for solving the system \( G(x) = 0 \).
Modified Gauss-Newton Method

In the Line Search Algorithm for a given $x^k$, use the matrix $H^k$ given by

$$H^k = JG(x^k)^T JG(x^k) + \|G(x^k)\| I,$$

where $I$ is the $n \times n$-identity matrix.

Note that $H^k$ is positive definite when $G(x^k) \neq 0$.

The direction-search problem corresponds to the unconstrained minimization of the (strictly) convex function

$$\nabla \theta(x^k)^T d + \frac{1}{2} dH^k d.$$

The solution $d$ satisfies: $\nabla \theta(x^k) + H^k d = 0 \iff H^k d = -\nabla \theta(x^k)$

Using the expression for $H^k$ and $\nabla \theta(x) = JG(x)^T G(x)$, we obtain

$$\left( JG(x^k)^T JG(x^k) + \|G(x^k)\| I \right) d = -JG(x^k)^T G(x^k)$$

Hence:

$$d = -\left( JG(x^k)^T JG(x^k) + \|G(x^k)\| I \right)^{-1} JG(x^k)^T G(x^k)$$

Gauss-Newton: $d = -\left( JG(x^k)^T JG(x^k) \right)^{-1} JG(x^k)^T G(x^k)$
Modified Gauss-Newton: Convergence Result

Proposition 8.3.8: Let $G : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be continuously differentiable. Then, every accumulation point of the modified Gauss-Newton method is stationary point of the unconstrained problem

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} G(x)^T G(x).$$

Proof: Suppose $x^*$ is an accumulation point of $\{x^k\}$. Suppose that $G(x^*) = 0$. Then, $x^*$ is a global minimum of $\theta(x)$ over $\mathbb{R}^n$, and hence it is a stationary point of $\theta(x)$. When $G(x^*) \neq 0$, we apply Proposition 8.3.7 - Homework 4. □.

NOTE: Here, a stationary point $\tilde{x}$ of $\theta(x)$ satisfies

$$\nabla \theta(\tilde{x}) = 0 \iff JG(\tilde{x})^T G(\tilde{x}) = 0$$

Again, the accumulation points are only characterized; they need not exist.
Application to a Complementarity Problem

Applying the Line Search Method to constrained complementarity problem:  

**Given a closed convex set** $X \subseteq \mathbb{R}^n$, **find an** $x \in X$ **such that**

$$0 \leq F(x) \perp G(x) \succeq 0,$$

where $F, G : \mathbb{R}^n \to \mathbb{R}^n$ are continuously differentiable.

- **We define** $B$-differentiable function $\theta$ **as follows**

$$\theta(x) = \frac{1}{2} \| \Phi(x) \|^2 \quad \text{with} \quad \Phi(x) = \min\{F(x), G(x)\} \quad \text{for} \ x \in \mathbb{R}^n.$$

- **Function** $\Phi$ **has components** $\Phi_i : \mathbb{R}^n \to \mathbb{R}$ **given by** $\Phi_i = \min\{F_i(x), G_i(x)\}$.  

Thus, we have

$$\theta(x) = \frac{1}{2} \sum_{i=1}^{n} [\Phi_i(x)]^2$$
Therefore,
\[ \theta'(x; d) = \sum_{i=1}^{n} \Phi_i(x) \Phi'_i(x; d) \quad \text{for all } x, d \in \mathbb{R}^n \]

where

\[
\Phi'_i(x; d) = \begin{cases} 
F'_i(x; d) & \text{if } F_i(x) < G_i(x) \\
G'_i(x; d) & \text{if } G_i(x) < F_i(x) \\
\min\{F'_i(x; d), G'_i(x; d)\} & \text{if } F_i(x) = G_i(x) 
\end{cases}
\]

By differentiability of \( F \) and \( G \), we have \( F'_i(x; d) = \nabla F_i(x)^T d \) and similarly \( G'_i(x; d) = \nabla G_i(x)^T d \).

- Define the following index sets, for a fixed but arbitrary \( x \),
\[
\mathcal{I}_F(x) = \{ i \mid F_i(x) < G_i(x) \} \\
\mathcal{I}_G(x) = \{ i \mid G_i(x) < F_i(x) \} \\
\mathcal{I}_=(x) = \{ i \mid F_i(x) = G_i(x) \}
\]
Thus, we can write $\theta'(x; d)$ as follows

$$
\theta'(x; d) = \sum_{i \in I_F(x)} F_i(x) \nabla F_i(x)^T d + \sum_{i \in I_G(x)} G_i(x) \nabla G_i(x)^T d \\
+ \sum_{i \in I_\pm(x)} \min \left\{ F_i(x) \nabla F_i(x)^T d, G_i(x) \nabla G_i(x)^T d \right\}
$$

- The direction-search problem now involves a nonconvex function $\theta'(x; d)$ so the **convexity of this problem is lost.**
• To deal with the nonconvexity, we modify the direction-search problem by introducing a convex function \( a(x, \cdot) \) defined as follows:

\[
a(x, d) = \sum_{i \in \mathcal{I}_F(x)} F_i(x) \nabla F_i(x)^T d + \sum_{i \in \mathcal{I}_G(x)} G_i(x) \nabla G_i(x)^T d \\
+ \sum_{i \in \mathcal{I}_=} \max \left\{ F_i(x) \nabla F_i(x)^T d, G_i(x) \nabla G_i(x)^T d \right\}
\]

• In fact, this function is **piecewise linear in** \( d \) for a fixed \( x \).

• Furthermore, it is a **majorant** of the directional derivative \( \theta'(x; \cdot) \) over \( \mathbb{R}^n \), i.e., for every \( x \in \mathbb{R}^n \),

\[
a(x, d) \geq \theta'(x; d) \quad \text{for all } d \in \mathbb{R}^n.
\]

• Alternative convex majorant of \( \theta \) can be obtained by replacing the last summand in the definition of the function \( a(\cdot, \cdot) \) by the average

\[
\frac{1}{2} \left( F_i(x) \nabla F_i(x)^T d + G_i(x) \nabla G_i(x)^T d \right)
\]
**Limsup Property of \( a \)**

**Proposition 8.3.20:** Let \( \{x^k, d^k\} \) be a convergent sequence. Then, for every sequence \( \{t_k\} \) of positive scalars, we have

\[
\limsup_{k \to \infty} \frac{\theta(x^k + t_k d^k) - \theta(x^k) - t_k a(x^k, d^k)}{t_k} \leq 0
\]

**Proof:** We at first write the function \( a(x, d) \) as the following sum:

\[
a(x, d) = \sum_{i \in \mathcal{I}_F(x)} a_i(x, d) + \sum_{i \in \mathcal{I}_G(x)} a_i(x, d) + \sum_{i \in \mathcal{I}_\ast(x)} a_i(x, d)
\]

where

\[
a_i(x, d) = \begin{cases} 
F_i(x) \nabla F_i(x)^T d & \text{if } i \in \mathcal{I}_F(x) \\
G_i(x) \nabla G_i(x)^T d & \text{if } i \in \mathcal{I}_G(x) \\
\max \{F_i(x) \nabla F_i(x)^T d, G_i(x) \nabla G_i(x)^T d\} & \text{if } i \in \mathcal{I}_\ast(x)
\end{cases}
\]

(1)
By definition, we have $\theta(x) = \sum_{i=1}^{n} [\Phi_i(x)]^2$. Fix an index $i$ and for arbitrary $k$, consider the difference $[\Phi_i(x^k + t_k d^k)]^2 - [\Phi_i(x^k)]^2$.

Suppose $i \in I = (x^k)$, i.e., $F_i(x^k) = G_i(x^k)$ and without loss of generality assume that $\Phi_i(x^k) = F_i(x^k)$. Then, by relation (1) for $a_i$, we have

$$
[\Phi_i(x^k + t_k d^k)]^2 - [\Phi_i(x^k)]^2 = [F_i(x^k + t_k d^k)]^2 - [F_i(x^k)]^2
= 2t_k F_i(x^k) \nabla F_i(x^k)^T d^k + o(t_k)
\leq 2t_k a_i(x^k, d^k) + o(t_k)
$$

Suppose $i \in I_F(x^k)$, so that $\Phi_i(x^k) = F_i(x^k)$. Then, similar to the preceding by relation (1) for $a_i$, we obtain

$$
[\Phi_i(x^k + t_k d^k)]^2 - [\Phi_i(x^k)]^2 = [F_i(x^k + t_k d^k)]^2 - [F_i(x^k)]^2
= 2t_k F_i(x^k) \nabla F_i(x^k)^T d^k + o(t_k)
= 2t_k a_i(x^k, d^k) + o(t_k)
$$
Suppose \( i \in \mathcal{I}_G(x^k) \). By the same line of argument with \( G_i \) in place of \( F_i \), we conclude that

\[
[\Phi_i(x^k + t_kd^k)]^2 - [\Phi_i(x^k)]^2 = 2t_k a_i(x^k, d^k) + o(t_k).
\]

Therefore, for any \( i \) and \( k \), there holds

\[
[\Phi_i(x^k + t_kd^k)]^2 - [\Phi_i(x^k)]^2 \leq 2t_k a_i(x^k, d^k) + o(t_k),
\]

implying that for all \( i \) and \( k \),

\[
\frac{1}{2} \left( [\Phi_i(x^k + t_kd^k)]^2 - [\Phi_i(x^k)]^2 - t_k a_i(x^k, d^k) \right) \leq \frac{1}{2} o(t_k).
\]

By summing over \( i \), we obtain for all \( k \),

\[
\theta(x^k + t_kd^k) - \theta(x^k) - t_k a(x^k, d^k) \leq \frac{1}{2} o(t_k).
\]

By dividing with \( t_k \) and taking \( \limsup \) as \( k \to \infty \), we obtain the desired result. \( \square \)
**Line Search Algorithm**

Suppose we use Line Search Algorithm with

- Symmetric positive definite matrices $H^k$ with uniform bounds on the minimum and the maximum eigenvalues, i.e., such that for some scalars $c_1 > 0$ and $c_2 > 0$, we have for all $k$
  
  $$c_1\|y\|^2 \leq y^T H^k y \leq c_2\|y\|^2.$$  

- Forcing function $\sigma(x, d) = -a(x, d)$

- At iteration $k$, the direction $d^k$ determined by solving the following direction-search (convex) problem:
  
  \[
  \begin{align*}
  \text{minimize} & \quad a(x^k, d) + \frac{1}{2}d^T H^k d \\
  \text{subject to} & \quad d \in X - x
  \end{align*}
  \]

- Stopping rule $a(x^k, d^k) = 0$

- Otherwise, we must have $a(x^k, d^k) < 0$, and we determine the next iterate $x^{k+1}$ by using Armijo line search rule
Some Properties

Suppose that \( \{x^k \mid k \in \mathcal{K}\} \) converges to \( x^* \), which belongs to \( X \) by closedness of \( X \). In view of the direction search problem, it follows that

\[
\alpha(x^k, d^k) + \frac{1}{2} (d^k)^T H^k d^k \leq 0,
\]

implying \((d^k)^T H^k d^k \leq -2 \alpha(x^k, d^k)\). By the choice of \( H^k \), we have

\[
c_1 \|d^k\|^2 \leq -2 \alpha(x^k, d^k)
\]

Therefore

\[
\|d^k\| \leq \sqrt{-2c_1^{-1} \alpha(x^k, d^k)} \tag{2}
\]

Hence, \( \{d^k \mid k \in \mathcal{K}\} \) is bounded, and let \( \mathcal{K}' \subseteq \mathcal{K} \) be such that \( \{d^k \mid k \in \mathcal{K}'\} \) is convergent. Then, by Proposition 8.3.20, we obtain

\[
\limsup_{k \to \infty \atop k \in \mathcal{K}'} \frac{\theta(x^k + t_k d^k) - \theta(x^k) - t_k \alpha(x^k, d^k)}{t_k} \leq 0
\]

Hence, Theorem 8.3.3 applies and thus, \( \alpha(x^k, d^k) \to 0 \) as \( k \to \infty \), \( k \in \mathcal{K}' \). By Eq. (2), we also have \( d^k \to 0 \) as \( k \to \infty \), \( k \in \mathcal{K}' \).
Forcing Function Convergence Properties

Recall:

**Theorem 8.3.3.** Let $X \subseteq \mathbb{R}^n$ be a convex set and let $\theta$ be locally Lipschitz function on $X$. Let $\{x^k\}$ be a sequence generated by the Line Search Algorithm. Assume that $\{x^k \mid k \in K\}$ is a subsequence with the following properties:

1. The function value sequence $\{\theta(x^k) \mid k \in K\}$ is bounded below.

2. For every sequence of positive scalars $\{t_k \mid k \in K\}$ converging to zero, there holds

$$\limsup_{\substack{k \to \infty \\cap \\ K \in K}} \frac{\theta(x^k + t_k d^k) - \theta(x^k) + t_k \sigma(x^k, d^k)}{t_k} \leq 0$$

Then, we have

$$\lim_{\substack{k \to \infty \\cap \\ K \in K}} \sigma(x^k, d^k) = 0.$$
Convergence Result

**Proposition 8.3.21:** Let $X \subseteq \mathbb{R}^n$ be closed convex, and $F, G : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable. Let $\{x^k\}$ be generated by Line Search Algorithm, and assume that $x^*$ is its accumulation point. Then, $x^*$ is a solution to the complementarity problem

$$0 \leq F(x) \perp G(x) \geq 0 \quad \text{with } x \in X$$

if and only if for every partitioning of the index set $\mathcal{I}_\equiv(x^*)$ into three mutually disjoint sets $A_F$, $A_\equiv$ and $A_G$, there exists a vector $d^* \in X - x^*$ satisfying

$$F_i(x^*) + \nabla F_i(x^*)^T d^* = 0 \quad \text{for all } i \in \mathcal{I}_F(x^*) \cup A_F$$

$$G_i(x^*) + \nabla G_i(x^*)^T d^* = 0 \quad \text{for all } i \in \mathcal{I}_G(x^*) \cup A_G$$

$$\left[\Phi_i(x^*)\right]^2 + \max\{F_i(x^*)\nabla F_i(x^*)^T d^*, G_i(x^*)\nabla G_i(x^*)^T d^*\} \leq 0 \quad \text{for all } i \in A_\equiv$$

(3)
Proof

Only If Part: Suppose that the given partition property of $\mathcal{I}_-(x^*)$ holds. We prove that $x^*$ solves the $CP$. Since $x^*$ is an accumulation point of $\{x^k\}$, we can find an appropriate subsequence $\{x^k \mid k \in \mathcal{K}\}$ such that

$$\mathcal{I}_F(x^k) = C_F, \quad \mathcal{I}_G(x^k) = C_G, \quad \mathcal{I}_-(x^k) = C_- \quad \text{for all } k \in \mathcal{K}. \tag{4}$$

Then, by continuity of $F, \nabla F, G$ and $\nabla G$, we obtain

$$\mathcal{I}_F(x^*) \subseteq C_F, \quad \mathcal{I}_G(x^*) \subseteq C_G, \quad C_- \subseteq \mathcal{I}_-(x^*). \tag{5}$$

Define a partition of the index set $\mathcal{I}_-(x^*)$, as follows:

$$A_F = C_F \setminus \mathcal{I}_F(x^*), \quad A_G = C_G \setminus \mathcal{I}_G(x^*), \quad A_- = C_- \tag{6}$$

Let $d^* \in X - x^*$ be a vector satisfying Eq. (3). For any $t \in [0, 1]$, the vector $d = t(d^* + x^* - x^k)$ is feasible in direction-search problem at iteration $k$, implying that

$$a(x^k, d) + \frac{1}{2} d^T H^k d \geq a(x^k, d^k) + \frac{1}{2} (d^k)^T H^k d^k$$
We have \( a(x^k, d^k) \rightarrow 0 \) and \( d^k \rightarrow 0 \) as \( k \rightarrow \infty \), \( k \in \mathcal{K} \). Furthermore, by definition of \( a \), we obtain \( a(x^k, d) = t a(x^k, d^* + x^* - x^k) \). Thus,

\[
    t \lim_{k \rightarrow \infty} a(x^k, d^* + x^* - x^k) + t^2 \lim_{k \rightarrow \infty} \sup_{k \in \mathcal{K}} (d^*)^T H^k d^* \geq 0
\]

By the uniform max and min eigenvalue boundedness property for \( H^k \), we have \( \lim_{k \rightarrow \infty} \sup_{k \in \mathcal{K}} (d^*)^T H^k d^* \leq t^2 c_2 \|d^*\|^2 \). Hence

\[
    t \lim_{k \rightarrow \infty} a(x^k, d^* + x^* - x^k) + t^2 c_2 \|d^*\|^2 \geq 0.
\]

Dividing by \( t \), and letting \( t \downarrow 0 \), we obtain \( \lim_{k \rightarrow \infty} a(x^k, d^* + x^* - x^k) \geq 0 \).

By the definition of \( a \) and equation (4), due to taking limits in \( a(x^k, d^* + x^* - x^k) \), we have

\[
    \sum_{i \in \mathcal{C}_F} F_i(x^*) \nabla F_i(x^*)^T d^* + \sum_{i \in \mathcal{C}_G} G_i(x^*) \nabla G_i(x^*)^T d^* \\
    + \sum_{i \in \mathcal{C}_=} \max \left\{ F_i(x^*) \nabla F_i(x^*)^T d^*, G_i(x^*) \nabla G_i(x^*)^T d^* \right\} \geq 0
\]
From the definition of the partition in Eq. (6) and given relation (3), we see that

\[ F_i(x^*) \nabla F_i(x^*)^T d^* = -[F_i(x^*)]^2 \quad \text{for all } i \in C_F \]

\[ G_i(x^*) \nabla G_i(x^*)^T d^* = -[G_i(x^*)]^2 \quad \text{for all } i \in C_G \]

while the last term is less than \(-[\Phi_i(x^*)]^2\) for \(i \in C_-\). Using the preceding relations, Eq. (5), and ..., we can see that

\[-2\theta(x^*) \geq 0 \quad \text{Homework 4}\]

implying that \(\theta(x^*) = 0\).
Trust Region Algorithms

- Similar to Line Search Algorithms

- Differ mainly in the direction-search problem
  - Do not require positive definiteness of $H^k$
  - Use another parameter $\Delta$ to bound the length of directions $d_k$, ensuring the existence of the solution $d^k$

- One interpretation of the direction-search problem is that it approximates locally the objective function $\theta(x)$
  $$\theta(x) \approx \theta(x^k) + \theta'(x^k; x - x^k) + \frac{1}{2} (x - x^k)^T H^k (x - x^k)$$
  or in terms of a majorant/approximation $a(x, \cdot)$ for the derivative $\theta'(x^k; \cdot)$
  $$\theta(x) \approx \theta(x^k) + a(x^k, x - x^k) + \frac{1}{2} (x - x^k)^T H^k (x - x^k)$$
In Trust Region Methods, the right-hand side is taken as a local model of $\theta(x)$ that we trust within a ball $B(x^k, \Delta)$ for some $\Delta > 0$.

We then minimize (approximately) the model function within the specified ball, and decide on $x^{k+1}$ upon sufficient “descent”

- We change the parameter $\Delta$ from iteration to iteration
- Based on information gathered, we may decide to shrink or enlarge $\Delta$
- Depends on our belief of the quality of the local model within the $\Delta$-ball
The idea of Trust Region

Let \( \theta : \mathbb{R}^n \rightarrow \mathbb{R} \) be twice continuously differentiable, and suppose we want to minimize \( \theta \) over \( \mathbb{R}^n \). At a given iterate \( x^k \), suppose we approximate the function with a second-order model

\[
\theta(x^k + d) \approx \theta(x^k) + \nabla \theta(x^k)^T d + \frac{1}{2} d^T \nabla^2 \theta(x^k) d
\]

- A natural idea is to minimize the right-hand side over \( d \in \mathbb{R}^n \)

- Potential sources of trouble
  - The quadratic function in \( d \) may be unbounded below, so minimizing the right-hand side is meaningless
  - Even if it is bounded below, and therefore, the minimizer \( d^k \) exists, the quality of \( d^k \) becomes an issue
  - Direction \( d^k \) may lie far out of the region where the quadratic model provides good local approximation for \( \theta(x^k + d) \)
A possible remedy is to bound the region over which we minimize,

$$\min_{\|d\| \leq \Delta} \left\{ \nabla \theta(x^k)^T d + \frac{1}{2} d^T \nabla^2 \theta(x^k) d \right\}$$
Generic Trust Region Problem

Let $X \subseteq \mathbb{R}^n$ be closed convex, and let $\theta : \mathbb{R}^n \rightarrow \mathbb{R}$. For a given $\bar{x} \in X$, a scalar $\Delta > 0$, and a symmetric matrix $H$, the direction $\bar{d}$ is computed by solving (possibly approximately) the following problem

$$\begin{align*}
\text{minimize} & \quad a(\bar{x}, d) + \frac{1}{2} d^T H d \\
\text{subject to} & \quad \|d\| \leq \Delta, \quad d \in X - \bar{x}
\end{align*}$$

- In theory, any $\| \cdot \|$ can be used to define the trust region
- In practice, mostly used are Euclidean norm and max-norm
- Trust Region problem plays a role similar to that of the direction-search problem in Line Search Methods
- Uses the forcing function $\sigma(x, d) = -[a(x, d) + d^T H d/2]$
- Has a major operational difference from the Line Search Methods
  - After a sufficient objective decrease, shrinks or enlarges the trust region parameter $\Delta$ and generates a new iterate $x^{k+1}$
  - After an insufficient decrease, the parameter $\Delta$ shrinks and the algorithm does not move, $x^{k+1} = x^k$
Generic Trust Region Algorithm

Step 0  Choose \( x^0 \in X \) and the parameters \( 0 < \gamma_1 < \gamma_2 < 1, \Delta_0 > 0, \\Delta_{\text{min}} > 0, \rho \in (0, 1] \). Set \( k = 0 \).

Step 1  Find a vector \( d^k \) such that \( \|d^k\| \leq \Delta_k, d^k \in X - x^k \), and

\[
a(x^k, d^k) + \frac{1}{2} (d^k)^T H d^k \leq \rho \min_{\|d\| \leq \Delta, \ d \in X - x^k} \left\{ a(x^k, d) + \frac{1}{2} d^T H d \right\}
\]

Step 2  If \( \sigma(x^k, d^k) = 0 \) stop.

Step 3  In case of insufficient decrease, i.e., \( \theta(x^k + d^k) - \theta(x^k) > -\gamma_1 \sigma(x^k, d^k) \)

set \( x^{k+1} = x^k, \Delta_{k+1} = \Delta_k / 2, k := k + 1 \) and go to Step 1. Otherwise, set \( x^{k+1} = x^k + d^k, \) and

\[
\Delta_{k+1} = \begin{cases} 
\max\{2\Delta_k, \Delta_{\text{min}}\} & \text{if } \theta(x^k + d^k) - \theta(x^k) \leq -\gamma_2 \sigma(x^k, d^k) \\
\max\{\Delta_k, \Delta_{\text{min}}\} & \text{otherwise}
\end{cases}
\]

set \( k := k + 1 \) and go to Step 1.