3.6. The least squares problem.

\[ f(x) = \sum_{k=1}^{n} r_k(x)^2 = \| r(x) \|^2, \quad r = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} \]

Special, but often used problem.

**Linear case:** \( r(x) = Ax - b \)

where \( A = m \square \), \( b = m \square \) - given and \( x \in \mathbb{R}^n \) is unknown.

- \( \exists x \in \mathbb{R}^n : Ax = b \Rightarrow \min_x f(x) = 0 \)

- \( \nexists x \in \mathbb{R}^n \Rightarrow \exists x \in \mathbb{R}^n : Ax \approx b \)

**Solution:**

\[ f(x) = \| Ax - b \|^2 = \quad \]

\[ = x^T A^T A x - 2 b^T A x + b^T b. \]

Quadratic function!

Stationary point: \( \nabla f = 0 \)

\[ \nabla f(x) = 2 A^T A x - 2 A^T b = 0 \iff \]

\[ \iff A^T A x = A^T b \quad - \text{normal equation (\*)} \]

**Lemma:** \( x \) solves \( \min_x \| Ax - b \|^2 \Rightarrow \)

\[ \iff \bar{x} \text{ is a solution to (\*)}. \]

**Proof:** \( \implies \) see above.

\( \iff \bar{x} \) solves (\*) \( \Rightarrow \bar{x} \) is stationary point:

**Ex. 1.5:**

\[ f(x) = (x - \bar{x})^T A^T A (x - \bar{x}) + f(\bar{x}) = \]

\[ = \| A(x - \bar{x}) \|^2 + f(\bar{x}) \geq f(\bar{x}) \quad \text{and equality when } x = \bar{x} \Rightarrow \bar{x} \text{ optimal} \]
Remark: if $A^TA$ is invertible then
\[ x = (A^TA)^{-1}A^Tb \quad \text{unique solution}. \]

$A^+$: pseudoinverse.

\* if $A^TA$ not invertible then
\[ x = A^+b \quad \text{one of many solutions}. \]

Note: not $x = (A^TA)^+A^Tb$.

**Geometrical interpretation of LS:**
\[ A\cdot x = \begin{bmatrix} \vec{a}_1 & \ldots & \vec{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \sum_{k=1}^{n} \vec{a}_k \cdot x_k \]

linear combination of columns.

\[ \min \| b - A\cdot x \|_2 \Rightarrow \]
\[ \Rightarrow \text{shortest distance} = \]
\[ \Rightarrow \text{orthogonal} \Rightarrow \]
\[ A \cdot x_{\min} \perp A\cdot x, \quad \forall x \in \mathbb{R}^n. \]

(normal equation)

**Nonlinear case:**
\[ f(x) = \sum_{i=1}^{m} r_i(x)^2. \]

Try Newton: calculate $\nabla f, H$:
\[ \frac{\partial f}{\partial x_k} = \sum_{i=1}^{m} 2 r_i \frac{\partial r_i}{\partial x_k} = 2 \left[ \frac{\partial r_i}{\partial x_k} \bigg|_{x_k} \right] r_i \]

Denote $J = \{ J_{ij} \} = \{ \frac{\partial r_i}{\partial x_j} \}$ - Jacobian

Then
\[ \nabla f = 2 J^T r \]

Then
\[ \frac{\partial^2 f}{\partial x_k \partial x_j} = 2 \sum_{i=1}^{m} \left[ \frac{\partial r_i}{\partial x_j} \frac{\partial r_i}{\partial x_k} + r_i \frac{\partial^2 r_i}{\partial x_k \partial x_j} \right] \]

\[ \Rightarrow H = 2 J^T J + 2 \sum_{i=1}^{m} r_i \cdot \nabla^2 r_i \approx \]

\[ \approx 2 J^T J \]

\[ x_{k+1} = x_k - (J^T J)^{-1} J^T r(x_k) \quad \text{Gauss-Newton} \]
Ch 9. Penalty and barrier functions

The problem: \( \min_{x \in S \subset \mathbb{R}^n} f(x) \)

We assumed that \( S = \mathbb{R}^n \).

It was crucial in all methods that one can move in any direction.

Example: \( \min (x^2 + x_2^2) \mid x_1 + x_2 \geq 1 \)

Starting at \( (0, 2) \):

\(-\nabla f = [-4] \)

\( \text{Wrong!} \)

Remark: SD, Newton, CC, Conjugate Directions etc will never find the minimum when applied to \( f(x) \).

We need to pass information to the search direction about the constraints.

9.2. Penalty function method

\( \min_{x \in S \subset \mathbb{R}^n} f(x) \quad S = \{ x \in \mathbb{R}^n : g(x) \leq 0, h(x) = 0 \} \)

Here \( g(x) = \begin{bmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{bmatrix}, h(x) = \begin{bmatrix} h_1(x) \\ \vdots \\ h_k(x) \end{bmatrix} \).

\( g(x) \leq 0 \implies \text{all } g_k(x) \leq 0 \).

In theory, it is easy to reduce any \( S \) to the case of the whole \( \mathbb{R}^n \).

\[ F(x) = \begin{cases} f(x) & \text{if } x \in S \\ +\infty & \text{otherwise} \end{cases} \]

\( \Rightarrow \min_{x \in S} f(x) = \min_{x \in \mathbb{R}^n} F(x) \)

In practice, \( +\infty \) is replaced by something large.
For example, take
\[ \alpha(x) = \begin{cases} 0 & \text{if } x \in S \\ >0 & \text{otherwise} \end{cases} \]
and build \[ q(x) = f(x) + \mu \cdot \alpha(x) \Rightarrow \]
\[ q(x) \approx F(x) \text{ for large } \mu > 0. \]

Typical choice of \( \alpha \):

* \( q_k(x) \leq 0 \):
  \[ \alpha_k(x) = \max \{ 0, q_k(x) \} \]
  (alt. \( \alpha_k(x) = \max \{ 0, q_k(x) \}^2 \))

* \( h_j(x) = 0 \): \( \alpha_j(x) = h_j(x)^2 \)

Overall \( \alpha(x) \) for \( q(x) \leq 0 \), \( h(x) = 0 \):

\[ \alpha(x) = \sum_{k=1}^{m} \left( \max \{ 0, q_k(x) \} \right)^2 + \sum_{j=1}^{p} h_j(x)^2 \]
Remark: to start with large $\mu$ is bad, (ill-conditioned problem). In practice, iterations start from small $\mu$ and then gradually increase it using the answers as new starting points.

Strategy: pick $\mu_1 < \mu_2 < \ldots < \mu_n \to +\infty$

$q_{\mu}(x) = f(x) + \mu \cdot \alpha(x)$

<table>
<thead>
<tr>
<th>Start</th>
<th>$x_0$</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$\ldots$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solution to min $q_{\mu}$</td>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_3$</td>
<td>$\ldots$</td>
<td></td>
</tr>
</tbody>
</table>

- Convergence analysis (Th. 1, p. 316)

$x_k \to \bar{x} \Rightarrow \bar{x}$ is the solution to $\min f(x), x \in S$

- Better to use Newton/quasi Newton/CG
- SD is sensitive to ill-conditioned problem.
- Each $x_k$ is not feasible (outside $S$).
- Exterior approximations to $\bar{x}$.

9.3. Barrier function method

Only inequalities: $\min f(x) \mid q_j(x) \leq 0$

We take another approximation of $F(x) = \begin{cases} f(x) & \text{if } x \in S \\ +\infty & \text{otherwise} \end{cases}$

Take $\beta(x) = \begin{cases} \rightarrow +\infty & \text{if some } g_k(x) \to 0 \\ "+\infty" & \text{otherwise} \end{cases}$

[$\star$] "$+\infty$" is some very large number to prevent the line search to leave $S$

and build $q_\varepsilon(x) = f(x) + \varepsilon \cdot \beta(x) \Rightarrow$

$\Rightarrow q_\varepsilon(x) \approx F(x)$ for small $\varepsilon > 0$. 
Typical choice of $\beta$:

- $\beta_k(x) = \begin{cases} \frac{-1}{g_k(x)} & \text{if } g_k(x) < 0 \\ +\infty & \text{otherwise} \end{cases}$

- $\beta_k(x) = \begin{cases} 0 & \text{if } g_k(x) \leq -1 \\ -\ln(-g_k(x)) - g_k(x) - 1 & \text{if } -1 < g_k(x) < 0 \\ +\infty & \text{otherwise} \end{cases}$

General $g(x) \leq 0$:

$\beta(x) = \sum_{k=1}^{m} \beta_k(x)$

Example (again) $\min(x_1^2 + x_2^2) \mid x_1 + x_2 \geq 1$

$q(x) = \begin{cases} x_1^2 + x_2^2 + \frac{\epsilon}{x_1 + x_2 - 1} & \text{if } x_1 + x_2 > 1 \\ 10^5 & \text{otherwise} \end{cases}$

(alt. realmax in MATLAB)

Remark: similar strategy as for penalty:

Pick not very small $\epsilon_1 > \epsilon_2 > \ldots > \epsilon_k \rightarrow 0^+$

$q_\epsilon(x) = f(x) + \epsilon \beta(x)$

<table>
<thead>
<tr>
<th>$\epsilon_1$</th>
<th>$\epsilon_2$</th>
<th>$\epsilon_3$</th>
<th>$\ldots$</th>
</tr>
</thead>
<tbody>
<tr>
<td>start $X_0$</td>
<td>$X_1$</td>
<td>$X_2$</td>
<td>$\ldots$</td>
</tr>
<tr>
<td>solution $X_1$</td>
<td>$X_2$</td>
<td>$X_3$</td>
<td>$\ldots$</td>
</tr>
</tbody>
</table>

- Similar convergence analysis (Th. 2, p. 325)
- SD is no good here either.
- Line search must be used to stay in $S$.
- Each $x_k$ is feasible.

Interior approximations to $\bar{x}$. 

Example $f(x) = 1 - x^2$, $-1 \leq x \leq 1$