

Trust-Region Methods

Consider the problem

$$\min_{x \in \mathbb{R}^n} f(x) \quad (P)$$

(unconstrained problem)

Classical algorithms for solving (P):

- construct a sequence of points $\{x^k\}$ s.t. $x^k \rightarrow x^*$ for $k \rightarrow \infty$, where x^* is a stationary point of f ($\nabla f(x^*) = 0$)

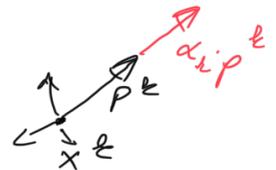
$$x^{k+1} = x^k + \underbrace{\alpha_k}_{\text{step size}} \cdot \underbrace{p^k}_{\text{search direction}}$$

- two popular strategies:

- line search and trust region method

Review line search methods

- choose a direction p^k
- choose the step size α_k



$$\alpha_k \approx \underset{\alpha > 0}{\operatorname{arg\,min}} f(x^k + \alpha p^k)$$

↓
exact line search

↓
inexact line search
- Wolfe conditions

Trust-Region- Methods

- select the "step size" first

- find a point x^{k+1} "close to x^k " which gives sufficient decrease of the objective value.
- if f is a function with a complex structure, a complete evaluation in a neighborhood of x^k is not viable. Therefore, we approximate f in a certain region by means of a linear or quadratic model.
- Recall the Taylor-series expansion of f around x^k :

$$f(x^k + p) = \underbrace{f_k}_{= f(x^k)} + \underbrace{g_k^T}_{= \nabla f(x^k)} p + \frac{1}{2} p^T \underbrace{H_k}_{= \nabla^2 f(x^k)} p \quad (k \in \{0, 1, \dots\})$$

- Idea: We approximate f with a simple objective m and solve $\tilde{x} = \arg \min_x m(x)$.

Problem: The minimizer \tilde{x} might be in a region where m poorly approximates f and therefore \tilde{x} might be far from minimal.

Solution: Restrict the search to a region where we trust that f is approximated well: Solve

$$\tilde{x}^* = \arg \min_{x \in \text{Trust Region}} m(x)$$

• Quadratic model (based on Taylor exp.)

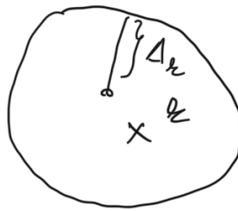
$$m_k(p) := f_k + g_k^T p + \frac{1}{2} p^T \underbrace{H_k}_{= \nabla^2 f(x^k)} p$$

symmetric matrix
uniformly bound

subproblem:

$$(1) \quad \min_{p \in \mathbb{R}^n} m_k(p) \quad \text{s.t.} \quad \|p\| \leq \Delta_k$$

where $\Delta_k > 0$ is the trust-region radius.



• Consider the quotient

$$S_k := \frac{f(x^k) - f(x^k + p^k)}{\underbrace{m_k(0) - m_k(p^k)}_{= f(x^k)}} \quad \left. \begin{array}{l} \text{real} \\ \text{decrease} \end{array} \right\} \quad \left. \begin{array}{l} \text{predicted} \\ \text{reduction} \end{array} \right\}$$

Note that $m_k(0) > m_k(p^k)$, because p^k is obtained by minimizing m_k over a region that includes $p=0$.

Hence, we can consider the following cases:

- $S_k < 0 \Rightarrow f(x^k) < f(x^k + p^k)$,
i.e., the new objective value is larger than the previous value $f(x^k)$;
 \Rightarrow step p^k is rejected and the trust region is shrunked.

- $\rho_k < \alpha$ ($0 < \alpha < 1$) \rightarrow we reject the new point; i.e., we set $x^{k+1} = x^k$ and minimize m_k for a smaller radius Δ_k
- $\rho_k \approx 1 \Rightarrow$ good agreement between the model m_k and f over this step $\Rightarrow x^{k+1}$ is accepted.
- $\rho_k \gg 1$ "very large" \Rightarrow increasing the trust region is beneficial for the following iterations.

Algorithm: Trust Region Method

- Choose $x^0 \in \mathbb{R}^n$, $\Delta_0 \in (0, +\infty)$, η_1 and η_2 with $0 < \eta_1 < \eta_2 < 1$, σ_1 and σ_2 with $0 < \sigma_1 < 1 < \sigma_2$, and set $k := 0$.
- If $\nabla f(x^k) = 0$, then STOP.
- Compute $f_k := f(x^k)$, $g_k := \nabla f(x^k)$, and choose a symmetric matrix $H_k \in \mathbb{R}^{(n,n)}$. Set

$$m_k(p) := f_k + g_k^T p + \frac{1}{2} p^T H_k p.$$
- Compute an (approximate) solution p^k of

$$\min_p m_k(p) \quad \text{s.t.} \quad \|p\| \leq \Delta_k,$$

and set $\hat{x}^k := x^k + p^k$.

5. Compute

$$S_k := \frac{f(x^k) - f(\hat{x}^k)}{m_k(0) - m_k(p^k)}$$

and choose

$$\Delta_{k+1} \in \begin{cases} [\Delta_k, \sigma_2 \Delta_k) & \text{if } S_k \geq \eta_2 \\ [\sigma_1 \Delta_k, \Delta_k) & \text{if } S_k \in [\eta_1, \eta_2) \\ (0, \sigma_1 \Delta_k) & \text{if } S_k < \eta_1 \end{cases}$$

6. If $S_k \geq \eta_1$, set $x^{k+1} := \hat{x}^k$ and $k := k+1$, go to step 2.
 Otherwise: Set $x^{k+1} := x^k$ and $k := k+1$, and go to step 4.

In order to solve the subproblem (1), we make use of the following statement

Theorem: Let $H \in \mathbb{R}^{(n,n)}$ be symmetric, positive definite and $c \in \mathbb{R}^n$. Then the unique minimizer p^* of the problem

$$\min c^T p + \frac{1}{2} p^T H p \quad (2)$$

$$p \quad \text{s.t.} \quad \|p\| \leq \Delta$$

satisfies

$$p^* = -H^{-1}c$$

if $\|H^{-1}c\| \leq \Delta$, and

$$\left. \begin{aligned} p^* &= -(H + \lambda^* I)^{-1} c, \\ \|p^*\| &= \Delta, \quad \lambda^* > 0 \end{aligned} \right\} (\exists)$$

otherwise.

Proof: Problem (2) is equivalent to

$$\min_p h(p) := c^T p + \frac{1}{2} p^T H p$$

$$\text{s.t.} \quad g(p) := p_1^2 + \dots + p_n^2 - \Delta^2 \leq 0$$

$$\begin{aligned} \text{We have:} \quad \nabla h &= c + H p \\ \nabla g &= 2p \end{aligned}$$

The KKT-conditions yield:

$$\frac{c + H p + \lambda \nabla g}{\lambda} = 0 \Leftrightarrow -c = H p - 2\lambda p = (H + 2\lambda I) p$$

$$\lambda (p_1^2 + \dots + p_n^2 - \Delta^2) = 0 \Leftrightarrow \begin{cases} \lambda (\|p\|^2 - \Delta^2) = 0 \\ \lambda \geq 0 \end{cases}$$

$$p_1^2 + \dots + p_n^2 - \Delta^2 \leq 0 \Leftrightarrow \|p\| \leq \Delta$$

- $\lambda = 0$: $c + H p = 0 \Rightarrow \begin{cases} p = -H^{-1}c \\ \|p\| \leq \Delta \end{cases}$
- $\lambda > 0$:

$$c + Hp + \frac{1}{2} \lambda p = 0$$

$$\Rightarrow \begin{cases} p = -(H + \frac{1}{2} \lambda I)^{-1} c \\ \|p\| = \Delta \end{cases}$$



Interpretation:

- Case 1: If the minimize $-H^{-1}c$ of the unconstrained problem $\min_p c^T p + \frac{1}{2} p^T H p$ belongs to the region $\{p \in \mathbb{R}^n \mid \|p\| \leq \Delta\}$, then $-H^{-1}c$ solves (2).
- Case 2: Otherwise, the minimize p^* lies on the boundary of that region, i.e., $\|p^*\| = \Delta$.

In order to calculate p^* , we must first determine the value of λ^* . This can be done with the aid of the Spectral Theorem:

Spectral Theorem: Let $H \in \mathbb{R}^{(n,n)}$ be a symmetric matrix. Then H has n real eigenvalues $\lambda_1, \dots, \lambda_n$, and there exist eigenvectors y^1, \dots, y^n s.t. the matrix $Y := (y^1, \dots, y^n)$ is orthogonal. The matrix H is

$$Y^T Y = Y Y^T = I$$

diagonalizable, i.e.,

$$H = Y D Y^T,$$

where D is the diagonal matrix

$$D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}.$$

Let $H = Y D Y^T$ be the diagonalization of H . Since Y is orthogonal, we obtain:

$$\begin{aligned} H + \alpha^* I &= Y D Y^T + Y \alpha^* I Y^T \\ &= Y (D + \alpha^* I) Y^T. \end{aligned}$$

$$\Rightarrow (H + \alpha^* I)^{-1} = Y (D + \alpha^* I)^{-1} Y^T.$$

Combining this with (2), we get

$$\begin{aligned} \Delta^2 &= \|p^*\|^2 = \|(H + \alpha^* I)^{-1} c\|^2 \\ &= \|Y (D + \alpha^* I)^{-1} Y^T c\|^2 \end{aligned}$$

$\|Yx\| = \|x\| \quad \forall x$ (orthogonal transformation is invariant under Euclid. norm)

$$= \|(D + \alpha^* I)^{-1} Y^T c\|^2$$

$$= \left\| \begin{pmatrix} \frac{1}{\lambda_1 + \alpha^*} & & \\ & \ddots & \\ & & \frac{1}{\lambda_n + \alpha^*} \end{pmatrix} Y^T c \right\|^2$$

$$\| \underbrace{O^T Y c}_{n \times 1} \|_{n \times 1}^2$$

$$= \left\| \left(\frac{c}{z_1 + z^*}, \dots, \frac{c}{z_n + z^*} \right) \right\|$$

($e^i \dots$ unit vector)

$$= \sum_{i=1}^n \frac{\alpha_i^2}{(z_i + z^*)^2} \quad (4)$$

$$\left(\alpha_i = e^{iT} Y^T c = c^T Y e^i = c^T y^i \right)$$

\Rightarrow Now, z^* is uniquely determined

$\Rightarrow p^*$ can be calculated using z^* .

Example: We solve (2) for $H = \begin{pmatrix} 4 & 4 \\ 4 & 10 \end{pmatrix}$,

$$c = \begin{pmatrix} -16 \\ -34 \end{pmatrix}, \quad \Delta = 2.$$

We have $H^{-1} = \begin{pmatrix} 5/12 & -1/6 \\ -1/6 & 1/6 \end{pmatrix}$, and

$$H^{-1} \cdot c = \begin{pmatrix} -5/12 \cdot 16 + 1/6 \cdot 34 \\ 1/6 \cdot 16 - 1/6 \cdot 34 \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$$

$$\Rightarrow \| H^{-1} \cdot c \| = \sqrt{10} > \Delta$$

\Rightarrow solution p^* can be calculated by (4).

Diagonalization: $H = Y D Y^T$ with

$$Y = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}, \quad U = \begin{pmatrix} 2 & 1 \\ 0 & 12 \end{pmatrix}$$

Thus, $z_1 = 2$, $z_2 = 12$, $y^1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ -1 \end{pmatrix}$,

$$y^2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \alpha_1 = c^T y^1 = \frac{2}{\sqrt{5}},$$

$$\alpha_2 = c^T y^2 = -\frac{84}{\sqrt{5}}.$$

By (4), we have

$$\Delta^2 = 4 = \sum_{i=1}^2 \frac{\alpha_i^2}{(z_i + z^*)^2}$$

$$= \frac{0.8}{(2 + z^*)^2} + \frac{1411.2}{(12 + z^*)^2}$$

$$\Rightarrow z^* \approx 6.807.$$

From (3), we obtain

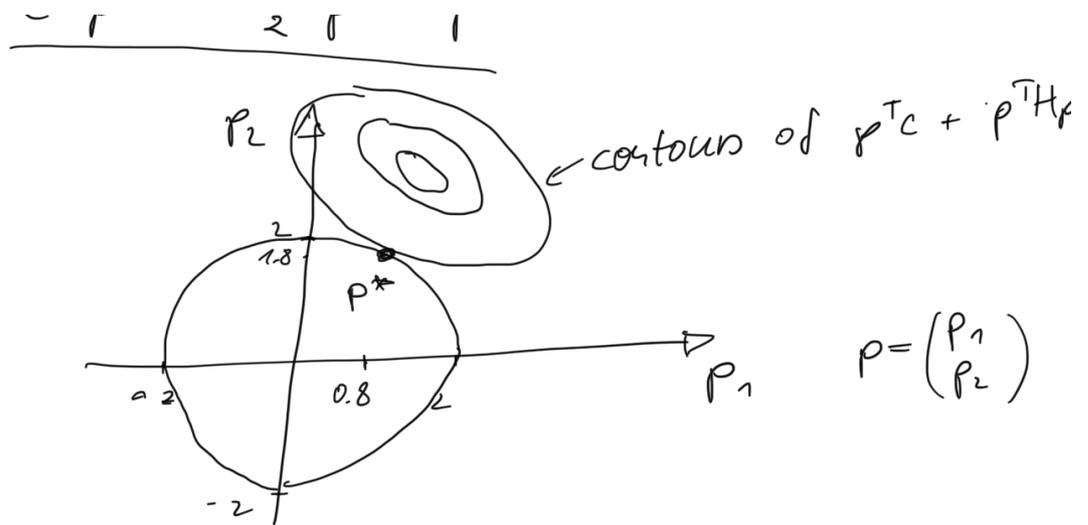
$$p^* = - (H + z^* I)^{-1} c$$

$$= \begin{pmatrix} 4 + z^* & 4 \\ 4 & 10 + z^* \end{pmatrix}^{-1} \begin{pmatrix} 16 \\ 34 \end{pmatrix}$$

$$= \frac{1}{(4 + z^*)(10 + z^*) - 16} \begin{pmatrix} 10 + z^* & -4 \\ -4 & 4 + z^* \end{pmatrix} \begin{pmatrix} 16 \\ 34 \end{pmatrix}$$

$$\approx \begin{pmatrix} 0.8024 \\ 1.832 \end{pmatrix} \text{ with objective value}$$

$$c^T p^* + \frac{1}{2} p^{*T} H p^* \approx -51.18.$$



The above theorem can be generalized for a positive semidefinite or indefinite matrix H . In these cases, the solution need not be unique.

Theorem: Let $H \in \mathbb{R}^{(n,n)}$ be symmetric and $c \in \mathbb{R}^n$. Any global minimize p^* of (2) satisfies $(H + \lambda^* I)p^* = -c$, where λ^* is s.t. $(H + \lambda^* H)$ is positive semidefinite, $\lambda^* \geq 0$, and $\lambda^* (\|p^*\| - 1) = 0$.

Proof: See D.C. Sorensen: Newton's Method with a model trust-region modification, SIAM J. Numerical Analysis 19, pp. 404-426, 1982. 