Trust Region Methods

Lecturer: Pradeep Ravikumar
Co-instructor: Aarti Singh

Convex Optimization 10-725/36-725

Trust Region Methods

$$\min_p m_k(p) \approx f(x_k + p)$$

$$\text{s.t. } p \in R_k$$

- Iteratively solve approximations to objective function that are accurate only in "trust region"
- restrict step to lie in trust region R_k

A Popular Approximation for the Objective Function

- Recall Taylor's Theorem: for some scalar t in (0,1)

$$f(x_k + p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T \nabla^2 f(x_k + tp) p,$$

- So: $m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T B_k p,$
    - for some positive-definite symmetric B_k satisfies:

    $$m_k(p) - f(x_k + p) = O(\|p\|^2)$$

    - so the approx. error is small when p is small

A Popular Approximation for the Objective Function

- Recall Taylor's Theorem: for some scalar t in (0,1)

$$f(x_k + p) = f_k + \nabla f_k^T p + \tfrac{1}{2} p^T \nabla^2 f(x_k + tp)p,$$

- So: $m_k(p) = f_k + \nabla f_k^T p + \tfrac{1}{2} p^T B_k p,$
  - for some positive-definite symmetric B_k satisfies:

$$m_k(p) - f(x_k + p) = O(\|p\|^2)$$

  - so the approx. error is small when p is small

$\{p : \|p\| \leq \Delta_k\}$ is the trust-region

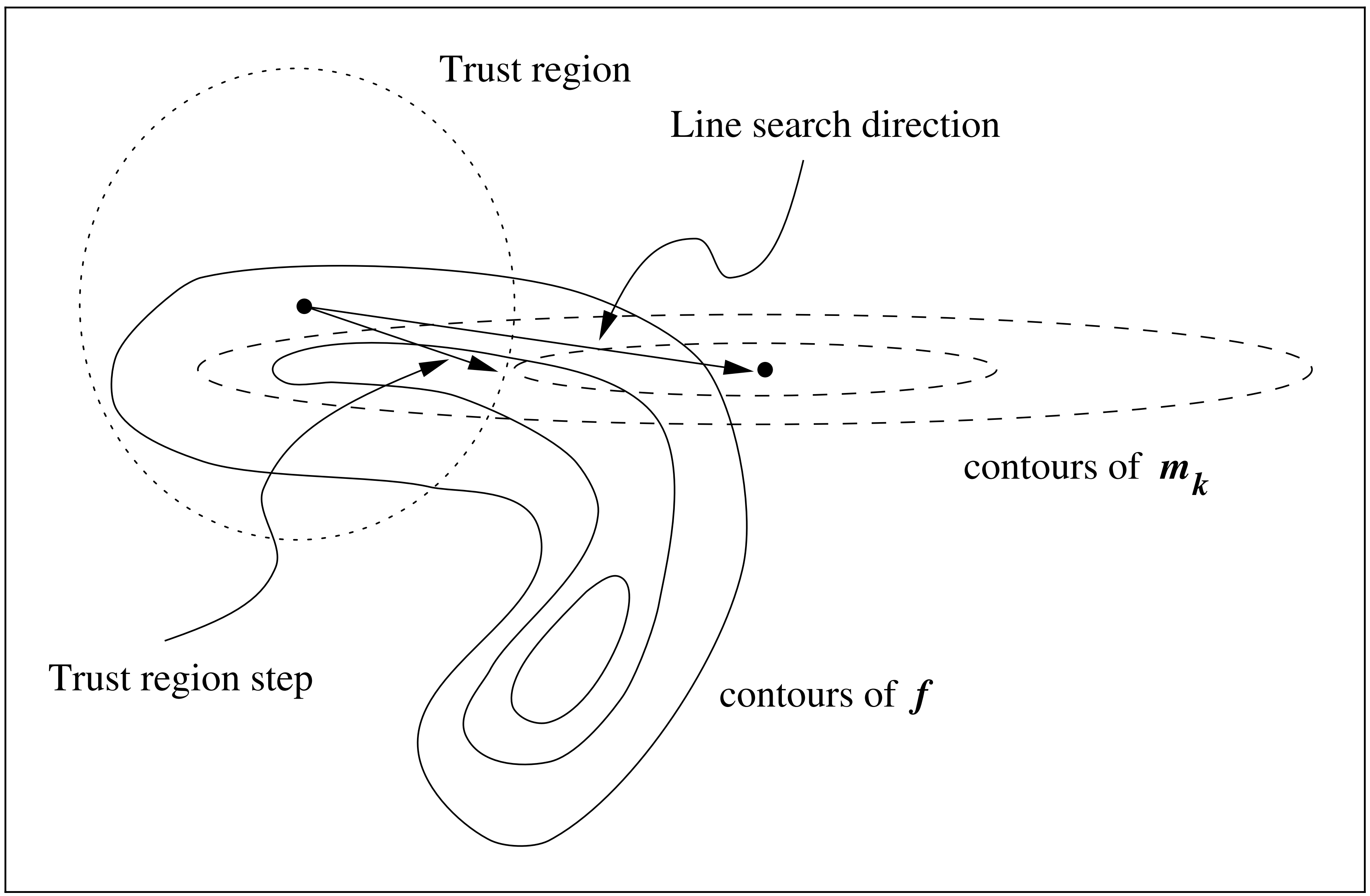
Δ_k is known as the trust-region radius

Quadratic Trust Region Method

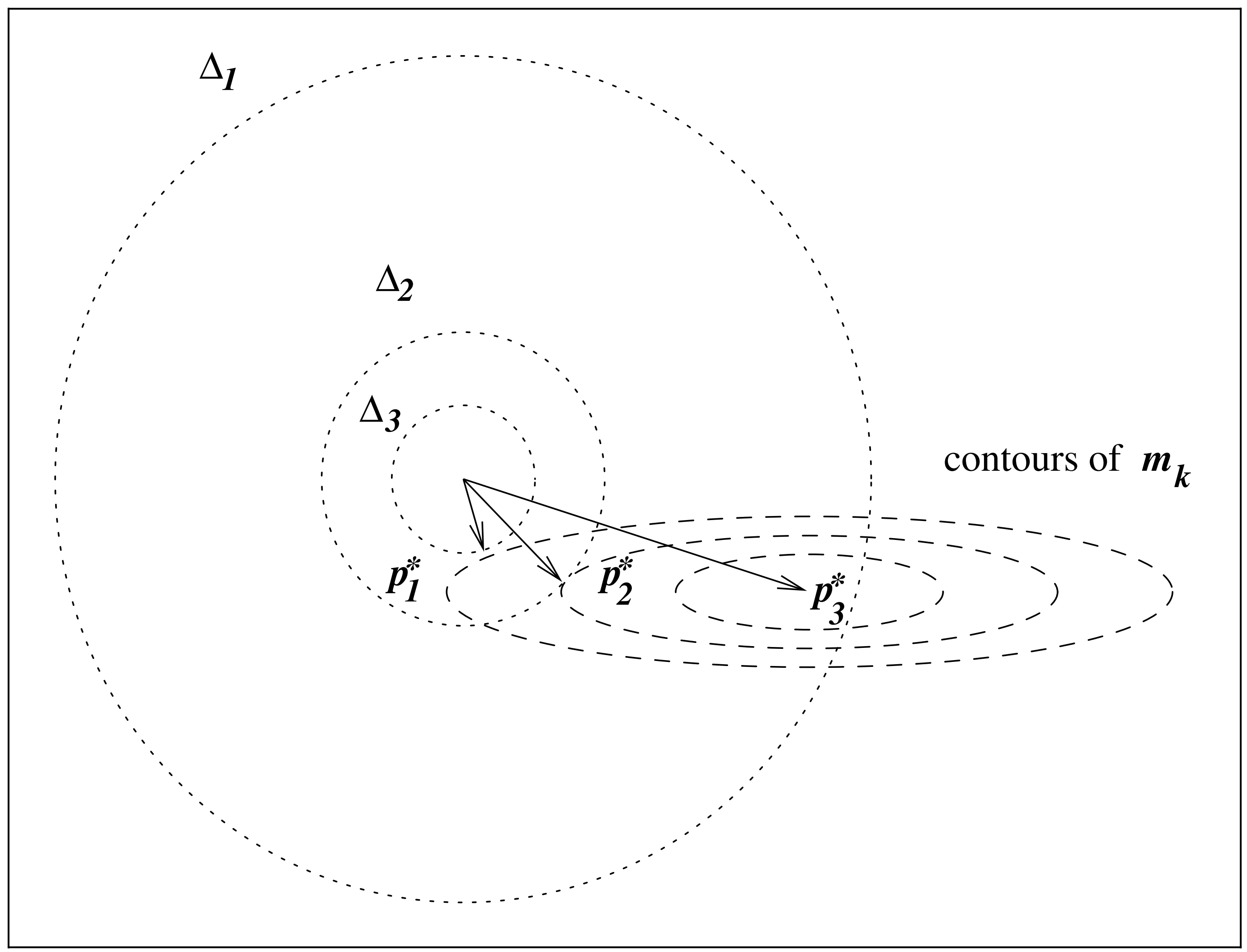
$$\min_{p \in \mathbb{R}^n} m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T B_k p \qquad \text{s.t. } \|p\| \leq \Delta_k,$$

$\|p\| = \sqrt{p^T p}$ is the ℓ_2 or Euclidean norm

Line Search vs Trust Region



Solution of Trust Region Problem for Different Radii



Adaptive Trust Region Radius

Given $\bar{\Delta} > 0$, $\Delta_0 \in (0, \bar{\Delta})$, and $\eta \in \left[0, \frac{1}{4}\right)$:

for $k = 0, 1, 2, \ldots$

Obtain p_k by solving trust region problem

Evaluate ρ_k (reduction ratio) $= \dfrac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)}$

Adaptive Trust Region Radius

Given $\bar{\Delta} > 0$, $\Delta_0 \in (0, \bar{\Delta})$, and $\eta \in \left[0, \frac{1}{4}\right)$:

for $k = 0, 1, 2, \ldots$

Obtain p_k by solving trust region problem

Evaluate ρ_k (reduction ratio) $= \dfrac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)}$

if $\rho_k < \frac{1}{4}$

$\Delta_{k+1} = \frac{1}{4}\|p_k\|$ … reduce trust region radius

else

if $\rho_k > \frac{3}{4}$ and $\|p_k\| = \Delta_k$

$\Delta_{k+1} = \min(2\Delta_k, \bar{\Delta})$ … increase trust region radius

else

$\Delta_{k+1} = \Delta_k$; … same trust region radius

Adaptive Trust Region Radius

Given $\bar{\Delta} > 0$, $\Delta_0 \in (0, \bar{\Delta})$, and $\eta \in \left[0, \frac{1}{4}\right)$:

for $k = 0, 1, 2, \ldots$

Obtain p_k by solving trust region problem

Evaluate ρ_k (reduction ratio) $= \dfrac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)}$

if $\rho_k < \frac{1}{4}$

$\Delta_{k+1} = \frac{1}{4}\|p_k\|$ … reduce trust region radius

else

if $\rho_k > \frac{3}{4}$ and $\|p_k\| = \Delta_k$

$\Delta_{k+1} = \min(2\Delta_k, \bar{\Delta})$ … increase trust region radius

else

$\Delta_{k+1} = \Delta_k$; … same trust region radius

if $\rho_k > \eta$

$x_{k+1} = x_k + p_k$

else … take step only if relative reduction is large

$x_{k+1} = x_k$;

end (for).

How to solve trust region problem?

Unconstrained Optimum

- Trust Region Problem:

$$\min_{p \in \mathbb{R}^n} m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T B_k p \qquad \text{s.t. } \|p\| \leq \Delta_k,$$

- So the unconstrained optimum can be written as:

$$p_k^B = -B_k^{-1} \nabla f_k$$

- So if unconstrained optimum lies within trust region, it is also the constrained optimum:

$$p_k^B \text{ is the solution to the trust region problem when } \|p_k^B\| \leq \Delta_k$$

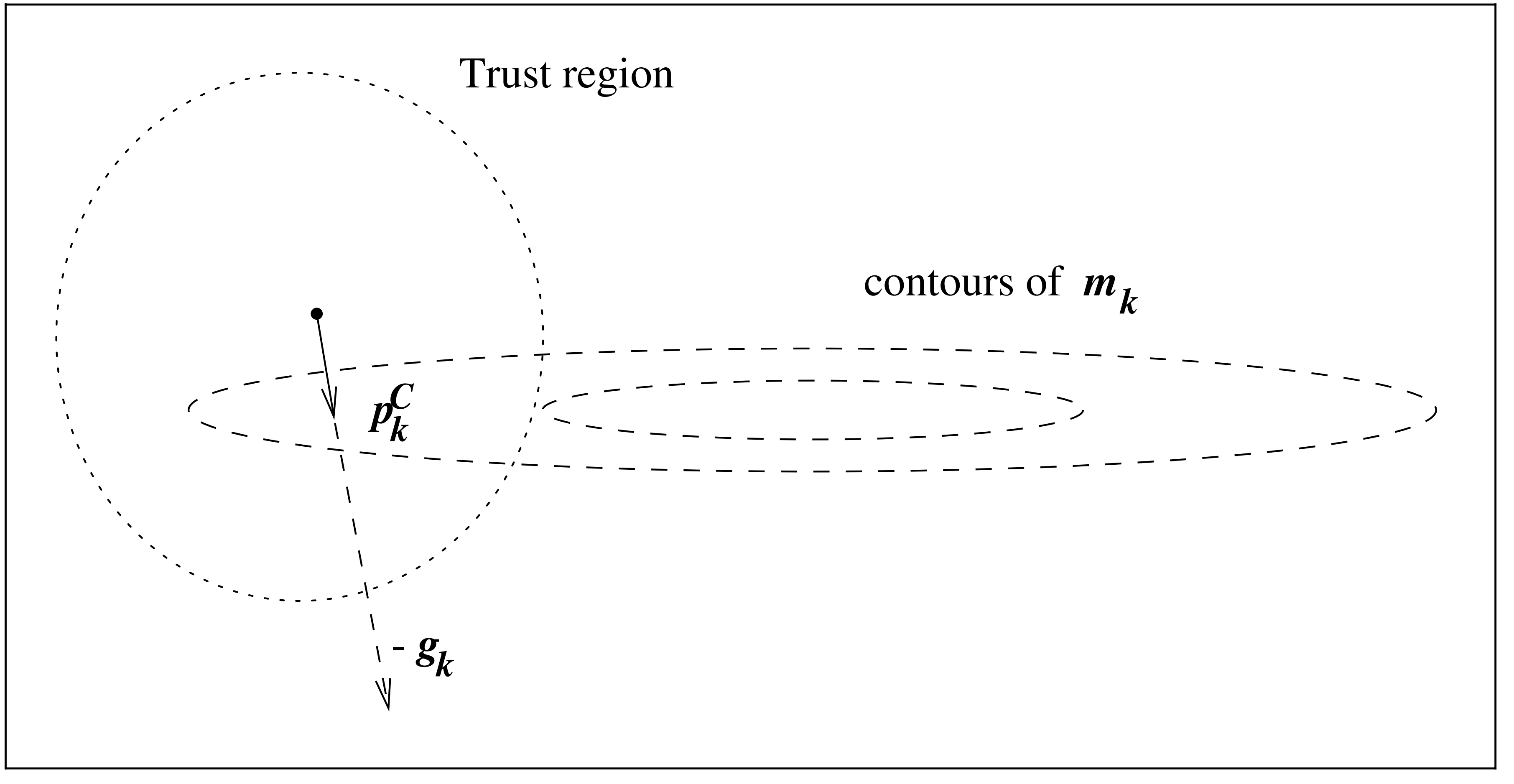
Unconstrained vs Constrained Optimum

- But the unconstrained optimum will typically not be the solution to trust region problem
- Solving exactly might be too expensive
  - recall that in “large scale” iterative methods, we do not want to spend too much computation per iteration
  - Solve trust region problem *approximately*

Approximate Solutions to Trust Region Problem

- Cauchy
- Dogleg
- Two-Dim Subspace Minimization
- One-dimensional root finding

Cauchy Point



Cauchy Point

- Solve just the linear approximation:

$$p_k^S = \arg\min_{p \in \mathbb{R}^n} f_k + \nabla f_k^T p \qquad \text{s.t. } \|p\| \leq \Delta_k;$$

Cauchy Point

- Solve just the linear approximation:

$$p_k^{\mathrm{S}} = \arg\min_{p \in \mathbb{R}^n} f_k + \nabla f_k^T p \qquad \text{s.t. } \|p\| \le \Delta_k;$$

Calculate the scalar $\tau_k > 0$ that minimizes $m_k(\tau p_k^{\mathrm{S}})$ subject to satisfying the trust-region bound, that is,

$$\tau_k = \arg\min_{\tau > 0} m_k(\tau p_k^{\mathrm{S}}) \qquad \text{s.t. } \|\tau p_k^{\mathrm{S}}\| \le \Delta_k;$$

Cauchy Point

- Solve just the linear approximation:

$$p_k^{\mathrm{S}} = \arg\min_{p \in \mathbb{R}^n} f_k + \nabla f_k^T p \qquad \text{s.t. } \|p\| \le \Delta_k;$$

Calculate the scalar $\tau_k > 0$ that minimizes $m_k(\tau p_k^{\mathrm{S}})$ subject to satisfying the trust-region bound, that is,

$$\tau_k = \arg\min_{\tau > 0} m_k(\tau p_k^{\mathrm{S}}) \qquad \text{s.t. } \|\tau p_k^{\mathrm{S}}\| \le \Delta_k;$$

Set $p_k^{\mathrm{C}} = \tau_k p_k^{\mathrm{S}}$.

- These steps have a closed form

Cauchy Point

- Cauchy Direction:

$$p_k^{\mathrm{S}} = \arg\min_{p \in \mathbb{R}^n} f_k + \nabla f_k^T p \qquad \text{s.t. } \|p\| \le \Delta_k;$$

$$\Rightarrow \quad p_k^{\mathrm{S}} = -\frac{\Delta_k}{\|\nabla f_k\|} \nabla f_k.$$

Cauchy Point

- Cauchy Direction:

$$p_k^{\mathrm{s}} = -\frac{\Delta_k}{\|\nabla f_k\|} \nabla f_k.$$

- Cauchy Point:

$$\tau_k = \arg\min_{\tau > 0} m_k(\tau p_k^{\mathrm{s}}) \qquad \text{s.t. } \|\tau p_k^{\mathrm{s}}\| \leq \Delta_k;$$

Cauchy Point

- Cauchy Direction:

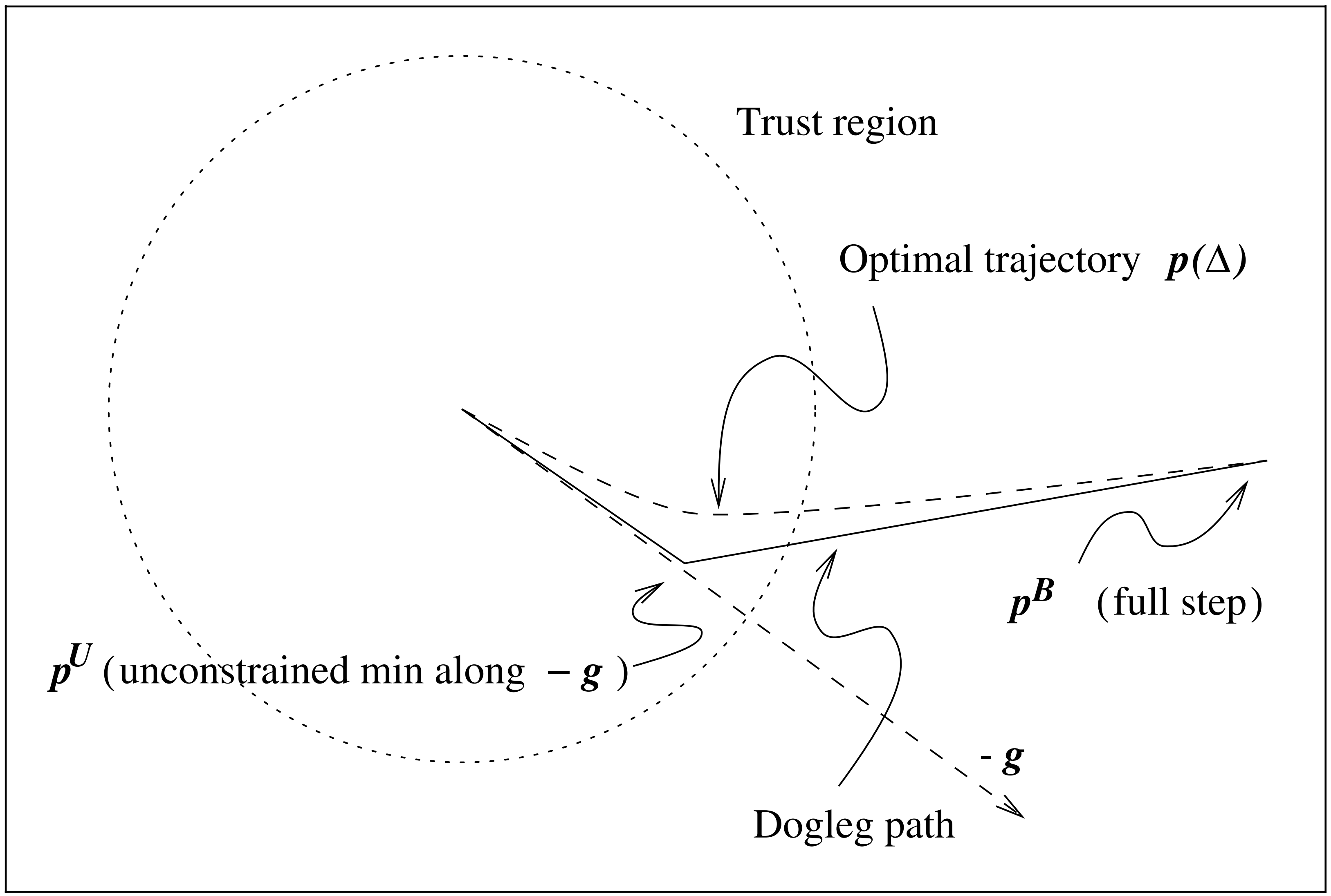
$$p_k^{\mathrm{S}} = -\frac{\Delta_k}{\|\nabla f_k\|} \nabla f_k.$$

- Cauchy Point:

$$p_k^{\mathrm{C}} = -\tau_k \frac{\Delta_k}{\|\nabla f_k\|} \nabla f_k,$$

$$\tau_k = \begin{cases} 1 & \text{if } \nabla f_k^T B_k \nabla f_k \le 0; \\ \min\left(\|\nabla f_k\|^3 / (\Delta_k \nabla f_k^T B_k \nabla f_k),\, 1\right) & \text{otherwise.} \end{cases}$$

Dogleg Method



Dogleg

$$p^B = -B_k^{-1}\nabla f_k \quad \ldots \text{ unconstrained minimum}$$

$$p^U = -\frac{(\nabla f_k)^T(\nabla f_k)}{(\nabla f_k)^T B_k (\nabla f_k)}\nabla f_k \quad \ldots \text{ steepest descent}$$

Dogleg path:

$$\tilde{p}(\tau) = \begin{cases} \tau p^{\mathrm{U}}, & 0 \le \tau \le 1, \\ p^{\mathrm{U}} + (\tau - 1)(p^{\mathrm{B}} - p^{\mathrm{U}}), & 1 \le \tau \le 2. \end{cases}$$

Dogleg

$$p^B = -B_k^{-1} \nabla f_k \quad \ldots \text{ unconstrained minimum}$$

$$p^U = -\frac{(\nabla f_k)^T (\nabla f_k)}{(\nabla f_k)^T B_k (\nabla f_k)} \nabla f_k \quad \ldots \text{ steepest descent}$$

Dogleg path:

$$\tilde{p}(\tau) = \begin{cases} \tau \, p^{\mathrm{U}}, & 0 \leq \tau \leq 1, \\ p^{\mathrm{U}} + (\tau - 1)(p^{\mathrm{B}} - p^{\mathrm{U}}), & 1 \leq \tau \leq 2. \end{cases}$$

Dogleg Step:

$$\tilde{\tau} = \arg \inf_{\tau \in [0,2]} m_k(\tilde{p}(\tau))$$

$$p^D = \tilde{p}(\tilde{\tau})$$

Two-dimensional Subspace Minimization

$$\min_p m(p) = f + g^T p + \tfrac{1}{2} p^T B p \quad \text{s.t. } \|p\| \le \Delta, \; p \in \text{span}[g, B^{-1} g].$$

- Note that entire dogleg path lies in span[g, B^-1 g]
- Note also that Cauchy point is feasible

Characterization of Solution

The vector p^* is a global solution of the trust-region problem

$$\min_{p \in \mathbb{R}^n} m(p) = f + g^T p + \tfrac{1}{2} p^T B p, \quad \text{s.t. } \|p\| \leq \Delta,$$

if and only if p^* is feasible and there is a scalar $\lambda \geq 0$ such that the following conditions are satisfied:

$$(B + \lambda I) p^* = -g,$$

$$\lambda(\Delta - ||p^*||) = 0,$$

$$(B + \lambda I) \quad \text{is positive semidefinite.}$$

One-dim. root finding

- Define:

$$p(\lambda) = -(B + \lambda I)^{-1} g$$

  - \lambda large enough s.t. B + \lambda I is positive definite

- Solve:

$$\|p(\lambda)\| = \Delta.$$

  - one-dimensional root finding problem
  - Approaches include Newton Raphson

Convergence Analyses

- Loosely: the gradients converge to zero under mild regularity conditions
- Requires adaptive adjusting of trust region radius as discussed earlier