

Unconstrained Minimization (I)

Lijun Zhang

zlj@nju.edu.cn

<http://cs.nju.edu.cn/zlj>





Outline

□ Unconstrained Minimization Problems

- Basic Terminology
- Examples
- Strong Convexity
- Smoothness

□ Descent Methods

- General Descent Method
- Exact Line Search
- Backtracking Line Search



Outline

□ Unconstrained Minimization Problems

- Basic Terminology
- Examples
- Strong Convexity
- Smoothness

□ Descent Methods

- General Descent Method
- Exact Line Search
- Backtracking Line Search



Basic Terminology

□ Unconstrained Optimization Problem

$$\min f(x)$$

- $f(x): \mathbf{R}^n \rightarrow \mathbf{R}$ is convex
- $f(x)$ always have a **domain** $\text{dom } f$
 - ✓ $\text{dom } f = \mathbf{R}^n, \text{dom } f \subset \mathbf{R}^n$
- $f(x)$ is twice continuously differentiable
 - ✓ $\text{dom } f$ is open, such as $(0, \infty)$
- The problem is solvable
 - ✓ There exists an optimal point x^*

$$\inf_x f(x) = f(x^*) = p^*$$



Basic Terminology

□ Unconstrained Optimization Problem

$$\min f(x)$$

- x^* is optimal **if and only if** } Equivalent

$$\nabla f(x^*) = 0$$

- Special cases: a closed-form solution
- General cases: an iterative algorithm

- ✓ A sequence of points $x^{(0)}, x^{(1)}, \dots \in \mathbf{dom} f$ with

$$f(x^{(k)}) \rightarrow p^* \text{ as } k \rightarrow \infty$$

- ✓ A **minimizing** sequence for the problem

- ✓ The algorithm is terminated when

$$f(x^{(k)}) - p^* \leq \epsilon$$

Requirements of Iterative Algorithm



□ Initial Point

- A suitable starting point

$$x^{(0)} \in \text{dom } f$$

□ Sublevel Set is Closed

$$S = \{x \in \text{dom } f \mid f(x) \leq f(x^{(0)})\}$$

- Satisfied for all $x^{(0)} \in \text{dom } f$ if the function f is closed
 - ✓ Continuous functions with $\text{dom } f = \mathbf{R}^n$
 - ✓ Continuous functions with open domains and $f(x) \rightarrow \infty$ as $x \rightarrow \text{bd dom } f$



Outline

□ Unconstrained Minimization Problems

- Basic Terminology
- Examples
- Strong Convexity
- Smoothness

□ Descent Methods

- General Descent Method
- Exact Line Search
- Backtracking Line Search



Examples

□ Convex Quadratic Minimization

Problem

$$\min \frac{1}{2} x^\top P x + q^\top x + r$$

■ $P \in \mathbf{S}_+^n, q \in \mathbf{R}^n, r \in \mathbf{R}$

■ Optimality Condition

$$P x^* + q = 0$$

1. $P \succ 0 \Rightarrow x^* = -P^{-1}q$ (unique solution)
2. If P is singular and $q \in \mathcal{R}(P)$, any solution of $P x^* + q = 0$ is optimal
3. If $q \notin \mathcal{R}(P)$, no solution, **unbound below**



Examples

□ Convex Quadratic Minimization Problem

$$\min \frac{1}{2}x^\top Px + q^\top x + r$$

■ $P \in \mathbf{S}_+^n, q \in \mathbf{R}^n, r \in \mathbf{R}$

3. If $q \notin \mathcal{R}(P)$, no solution, **unbound below**

✓ $q = a + b, a \in \mathcal{R}(P), b \perp \mathcal{R}(P), a \perp b$

✓ Let $x = tb$

$$\begin{aligned} & \frac{1}{2}x^\top Px + q^\top x + r \\ &= t(a + b)^\top b + r \\ &= t\|b\|_2^2 + r \end{aligned}$$



Examples

□ Least-Squares Problem

$$\min \|Ax - b\|_2^2 = x^\top A^\top Ax - 2b^\top Ax + b^\top b$$

- $A \in \mathbf{R}^{m \times n}, b \in \mathbf{R}^m$ are problem data

- Optimality Condition

$$\nabla f(x^*) = 2A^\top Ax^* - 2A^\top b = 0$$

- Normal Equations

$$A^\top Ax^* = A^\top b$$



Examples

□ Unconstrained Geometric Programming

$$\min f(x) = \log \left(\sum_{i=1}^m \exp(a_i^\top x + b_i) \right)$$

■ Optimality Condition

$$\nabla f(x^*) = \frac{\sum_{i=1}^m \exp(a_i^\top x^* + b_i) a_i}{\sum_{i=1}^m \exp(a_i^\top x^* + b_i)} = 0$$

✓ No analytical solution

■ An Iterative Algorithm

✓ $\text{dom } f = \mathbf{R}^n$, any point can be chosen as $x^{(0)}$



Examples

□ Analytic Center of Linear Inequalities

$$\min f(x) = - \sum_{i=1}^m \log(b_i - a_i^\top x)$$

- $\text{dom } f = \{x \mid a_i^\top x < b_i, i = 1, 2, \dots, m\}$
- f is called as the logarithmic barrier for the inequalities $a_i^\top x < b_i$
- The solution of this problem is called the **analytic center** of the inequalities
- An Iterative Algorithm
 - ✓ $x^{(0)}$ must satisfy $a_i^\top x^{(0)} < b_i$



Outline

□ Unconstrained Minimization Problems

- Basic Terminology
- Examples
- Strong Convexity
- Smoothness

□ Descent Methods

- General Descent Method
- Exact Line Search
- Backtracking Line Search



Strong Convexity

□ $f(\cdot)$ is strongly convex on S , if $\exists m > 0$

$$\nabla^2 f(x) \succeq mI, \quad \forall x \in S$$

1. A Quadratic Lower Bound

■ $\forall x, y \in S, \exists z \in [x, y]$

$$\begin{aligned} f(y) &= f(x) + \nabla f(x)^\top (y - x) + \frac{1}{2} (y - x)^\top \nabla^2 f(z) (y - x) \\ &\geq f(x) + \nabla f(x)^\top (y - x) + \frac{m}{2} \|y - x\|_2^2 \end{aligned}$$



Strong Convexity

□ $f(\cdot)$ is strongly convex on S , if $\exists m > 0$

$$\nabla^2 f(x) \succeq mI, \quad \forall x \in S$$

1. A Quadratic Lower Bound

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) + \frac{m}{2} \|y - x\|_2^2, \quad \forall x, y \in S$$

- When $m = 0$, reduce to the first-order condition of convex functions



Strong Convexity

□ $f(\cdot)$ is strongly convex on S , if $\exists m > 0$

$$\nabla^2 f(x) \succeq mI, \quad \forall x \in S$$

1. A Quadratic Lower Bound

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) + \frac{m}{2} \|y - x\|_2^2, \quad \forall x, y \in S$$

2. A Condition for Suboptimality

$$\begin{aligned} f(y) &\geq \min_y f(x) + \nabla f(x)^\top (y - x) + \frac{m}{2} \|y - x\|_2^2 \\ &= f(x) + \nabla f(x)^\top (\tilde{y} - x) + \frac{m}{2} \|\tilde{y} - x\|_2^2, \quad \tilde{y} = x - \frac{1}{m} \nabla f(x) \\ &= f(x) - \frac{1}{2m} \|\nabla f(x)\|_2^2 \end{aligned}$$



Strong Convexity

□ $f(\cdot)$ is strongly convex on S , if $\exists m > 0$

$$\nabla^2 f(x) \succeq mI, \quad \forall x \in S$$

1. A Quadratic Lower Bound

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) + \frac{m}{2} \|y - x\|_2^2, \quad \forall x, y \in S$$

2. A Condition for Suboptimality

$$p_* \geq f(x) - \frac{1}{2m} \|\nabla f(x)\|_2^2 \iff f(x) - p_* \leq \frac{1}{2m} \|\nabla f(x)\|_2^2$$

- If the gradient is small at x , then it is nearly optimal

$$\|\nabla f(x)\|_2 \leq (2m\epsilon)^{\frac{1}{2}} \Rightarrow f(x) - p^* \leq \epsilon$$



Strong Convexity

□ $f(\cdot)$ is strongly convex on S , if $\exists m > 0$

$$\nabla^2 f(x) \succeq mI, \quad \forall x \in S$$

3. An Upper Bound of $\|x^* - x\|_2$

$$p_* = f(x^*)$$

$$\geq f(x) + \nabla f(x)^\top (x^* - x) + \frac{m}{2} \|x^* - x\|_2^2$$

$$\geq f(x) - \|\nabla f(x)\|_2 \|x^* - x\|_2 + \frac{m}{2} \|x^* - x\|_2^2$$

$$\geq p_* - \|\nabla f(x)\|_2 \|x^* - x\|_2 + \frac{m}{2} \|x^* - x\|_2^2$$



Strong Convexity

□ $f(\cdot)$ is strongly convex on S , if $\exists m > 0$

$$\nabla^2 f(x) \succeq mI, \quad \forall x \in S$$

3. An Upper Bound of $\|x^* - x\|_2$

$$\frac{m}{2} \|x^* - x\|_2^2 \leq \|\nabla f(x)\|_2 \|x^* - x\|_2$$

$$\Rightarrow \|x^* - x\|_2 \leq \frac{2}{m} \|\nabla f(x)\|_2$$

■ $x \rightarrow x^*$, as $\nabla f(x) \rightarrow 0$

■ The optimal point x^* is unique



Outline

□ Unconstrained Minimization Problems

- Basic Terminology
- Examples
- Strong Convexity
- Smoothness

□ Descent Methods

- General Descent Method
- Exact Line Search
- Backtracking Line Search



Smoothness

□ $f(\cdot)$ is smooth on S , if $\exists M > 0$

$$\nabla^2 f(x) \preceq MI, \quad \forall x \in S$$

1. A Quadratic Upper Bound

■ $\forall x, y \in S, \exists z \in [x, y]$

$$\begin{aligned} f(y) &= f(x) + \nabla f(x)^\top (y - x) + \frac{1}{2} (y - x)^\top \nabla^2 f(z) (y - x) \\ &\leq f(x) + \nabla f(x)^\top (y - x) + \frac{M}{2} \|y - x\|_2^2 \end{aligned}$$



Smoothness

□ $f(\cdot)$ is smooth on S , if $\exists M > 0$

$$\nabla^2 f(x) \preceq MI, \quad \forall x \in S$$

1. A Quadratic Upper Bound

$$f(y) \leq f(x) + \nabla f(x)^\top (y - x) + \frac{M}{2} \|y - x\|_2^2, \quad \forall x, y \in S$$

2. An Upper Bound of Gradients

$$\begin{aligned} \min_y f(y) &\leq \min_y f(x) + \nabla f(x)^\top (y - x) + \frac{M}{2} \|y - x\|_2^2 \\ &= f(x) + \nabla f(x)^\top (\tilde{y} - x) + \frac{M}{2} \|\tilde{y} - x\|_2^2, \quad \tilde{y} = x - \frac{1}{M} \nabla f(x) \\ &= f(x) - \frac{1}{2M} \|\nabla f(x)\|_2^2 \end{aligned}$$



Smoothness

□ $f(\cdot)$ is smooth on S , if $\exists M > 0$

$$\nabla^2 f(x) \preceq MI, \quad \forall x \in S$$

1. A Quadratic Upper Bound

$$f(y) \leq f(x) + \nabla f(x)^\top (y - x) + \frac{M}{2} \|y - x\|_2^2, \quad \forall x, y \in S$$

2. An Upper Bound of Gradients

$$p^* \leq f(x) - \frac{1}{2M} \|\nabla f(x)\|_2^2$$

$$\Rightarrow \frac{1}{2M} \|\nabla f(x)\|_2^2 \leq f(x) - p_*$$



Condition Number

□ Condition Number of a Matrix A

$$\text{cond}(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}$$

□ $f(\cdot)$ is both strongly convex and smooth

$$mI \preceq \nabla^2 f(x) \preceq MI, \quad \forall x \in S$$

■ Condition number of f

$$\kappa = \frac{M}{m} \geq \text{cond}(\nabla^2 f(x))$$

■ Has a strong effect on the efficiency of optimization methods



Condition Number

□ Geometric Interpretations

- Width of a convex set $C \subseteq \mathbf{R}^n$, in the direction q where $\|q\|_2 = 1$

$$W(C, q) = \sup_{z \in C} q^T z - \inf_{z \in C} q^T z$$

- Minimum width and maximum width of C

$$W_{\min} = \inf_{\|q\|_2=1} W(C, q), \quad W_{\max} = \sup_{\|q\|_2=1} W(C, q)$$

- Condition number of C

✓ $\text{cond}(C)$ is small implies C it is nearly spherical

$$\text{cond}(C) = \frac{W_{\max}^2}{W_{\min}^2}$$



Condition Number

□ Geometric Interpretations

- α -sublevel set of f

$$C_\alpha = \{x | f(x) \leq \alpha\}, \quad p^* \leq \alpha \leq f(x_0)$$

- $f(\cdot)$ is both strongly convex and smooth

$$p_* + \frac{M}{2} \|y - x^*\|_2^2 \geq f(y) \geq p_* + \frac{m}{2} \|y - x^*\|_2^2$$

$$B_{\text{inner}} \subseteq C_\alpha \subseteq B_{\text{outer}}$$

$$B_{\text{inner}} = \left\{ y \left| \|y - x^*\| \leq \left(\frac{2(\alpha - p^*)}{M} \right)^{1/2} \right. \right\} \quad B_{\text{outer}} = \left\{ y \left| \|y - x^*\| \leq \left(\frac{2(\alpha - p^*)}{m} \right)^{1/2} \right. \right\}$$



Condition Number

□ Geometric Interpretations

- α -sublevel set of f

$$C_\alpha = \{x | f(x) \leq \alpha\}, \quad p^* \leq \alpha \leq f(x_0)$$

- $f(\cdot)$ is both strongly convex and smooth

$$p_* + \frac{M}{2} \|y - x^*\|_2^2 \geq f(y) \geq p_* + \frac{m}{2} \|y - x^*\|_2^2$$

$$y \in C_\alpha \Rightarrow \alpha \geq f(y) \Rightarrow \alpha \geq p_* + \frac{m}{2} \|y - x^*\|_2^2$$

$$\Rightarrow \frac{m}{2} \|y - x^*\|_2^2 \leq \alpha - p_* \Rightarrow \|y - x^*\| \leq \sqrt{\frac{2}{m} (\alpha - p_*)}$$

$$\Rightarrow y \in B_{\text{outer}} = \left\{ y \mid \|y - x^*\| \leq \left(\frac{2(\alpha - p_*)}{m} \right)^{1/2} \right\} \Rightarrow C_\alpha \subseteq B_{\text{outer}}$$



Condition Number

□ Geometric Interpretations

- α -sublevel set of f

$$C_\alpha = \{x | f(x) \leq \alpha\}, \quad p^* \leq \alpha \leq f(x_0)$$

- $f(\cdot)$ is both strongly convex and smooth

$$p_* + \frac{M}{2} \|y - x^*\|_2^2 \geq f(y) \geq p_* + \frac{m}{2} \|y - x^*\|_2^2$$

$$y \in B_{\text{inner}} = \left\{ y \mid \|y - x^*\| \leq \left(\frac{2(\alpha - p^*)}{M} \right)^{1/2} \right\}$$

$$\Rightarrow f(y) \leq p_* + \frac{M}{2} \|y - x^*\|_2^2 \leq \alpha \Rightarrow y \in C_\alpha \Rightarrow B_{\text{inner}} \subseteq C_\alpha$$



Condition Number

□ Geometric Interpretations

- α -sublevel set of f

$$C_\alpha = \{x | f(x) \leq \alpha\}, \quad p^* \leq \alpha \leq f(x_0)$$

- $f(\cdot)$ is both strongly convex and smooth

$$B_{\text{inner}} \subseteq C_\alpha \subseteq B_{\text{outer}}$$

$$B_{\text{inner}} = \left\{ y \left| \|y - x^*\| \leq \left(\frac{2(\alpha - p^*)}{M} \right)^{1/2} \right. \right\} \quad B_{\text{outer}} = \left\{ y \left| \|y - x^*\| \leq \left(\frac{2(\alpha - p^*)}{m} \right)^{1/2} \right. \right\}$$

- Condition number of C_α

$$\text{cond}(C_\alpha) \leq \kappa = \frac{M}{m}$$



Discussions

□ Parameters m and M

- Known only in rare cases
- Unknown in general

□ They are conceptually useful

- They establish that the algorithm converges
- The convergence behavior of optimization algorithms depends on them

□ In Practice

- Estimate their values
- Design parameter-free algorithms



Outline

□ Unconstrained Minimization Problems

- Basic Terminology
- Examples
- Strong Convexity
- Smoothness

□ Descent Methods

- General Descent Method
- Exact Line Search
- Backtracking Line Search



Iterative Methods

□ A Minimizing Sequence

$$x^{(k+1)} = x^{(k)} + t^{(k)} \Delta x^{(k)}, \quad k = 1, \dots$$

- k is the the iteration number
- $x^{(k)}$ is the output of iterative methods
- $\Delta x^{(k)}$ is the step or search direction
- $t^{(k)} \geq 0$ is the step size or step length

□ Shorthand

$$x := x + t\Delta x$$



Descent Methods

□ Descent Methods

$$f(x^{k+1}) < f(x^k)$$

- Except when $x^{(k)}$ is optimal
- $\forall k, x^{(k)} \in S = \{x \in \text{dom } f \mid f(x) \leq f(x^{(0)})\}$
- The search direction makes an acute angle with the negative gradient

$$\nabla f(x^{(k)})^\top \Delta x^{(k)} < 0$$

$$\left. \begin{array}{l} f(x^{k+1}) \geq f(x^k) + \nabla f(x^{(k)})^\top (x^{k+1} - x^k) \\ \nabla f(x^{(k)})^\top \Delta x^{(k)} \geq 0 \Rightarrow \nabla f(x^{(k)})^\top (x^{k+1} - x^k) \geq 0 \end{array} \right\} \Rightarrow f(x^{k+1}) \geq f(x^k)$$



Descent Methods

□ Descent Methods

$$f(x^{k+1}) < f(x^k)$$

- Except when $x^{(k)}$ is optimal
- $\forall k, x^{(k)} \in S = \{x \in \text{dom } f \mid f(x) \leq f(x^{(0)})\}$
- The search direction makes an acute angle with the negative gradient

$$\nabla f(x^{(k)})^\top \Delta x^{(k)} < 0$$

- $\Delta x^{(k)}$ is called as descent direction



General Descent Method

□ The Algorithm

Given a starting point $x \in \text{dom } f$

Repeat

1. Determine a descent direction Δx .
2. Line search: Choose a step size $t \geq 0$.
3. Update: $x := x + t\Delta x$.

until stopping criterion is satisfied.

□ Line Search

- Determine the next iterate along the line

$$\{x + t\Delta x \mid t \in \mathbf{R}_+\}$$



General Descent Method

□ The Algorithm

Given a starting point $x \in \text{dom } f$

Repeat

1. Determine a descent direction Δx .
2. Line search: Choose a step size $t \geq 0$.
3. Update: $x := x + t\Delta x$.

until stopping criterion is satisfied.

□ Stopping Criterion

$$\|\nabla f(x)\|_2 \leq \eta$$



Outline

□ Unconstrained Minimization Problems

- Basic Terminology
- Examples
- Strong Convexity
- Smoothness

□ Descent Methods

- General Descent Method
- Exact Line Search
- Backtracking Line Search



Exact Line Search

□ Minimize f along the Ray

$$t = \operatorname{argmin}_{s \geq 0} f(x + s\Delta x)$$

- The cost of the minimization problem with one variable is low

$$\min_{s \geq 0} f(x + s\Delta x)$$

- The minimizer along the ray can be found analytically



Outline

- Unconstrained Minimization Problems
 - Basic Terminology
 - Examples
 - Strong Convexity
 - Smoothness
- Descent Methods
 - General Descent Method
 - Exact Line Search
 - Backtracking Line Search



Backtracking Line Search

- Most line searches used in practice are inexact
 - Approximately minimize f along the ray
 - Just reduce f 'enough'

□ Backtracking Line Search

given a descent direction Δx for f at $x \in \text{dom } f$, $\alpha \in (0, 0.5)$, $\beta \in (0, 1)$

$t := 1$

while $f(x + t\Delta x) > f(x) + \alpha t \nabla f(x)^\top \Delta x$, $t := \beta t$



Backtracking Line Search

□ The line search is called backtracking

- It starts with unit step size and then reduces it by the factor β

$$t := 1, \quad t := \beta t$$

□ It eventually terminates

- Δx is a descent direction, i.e., $\nabla f(x)^\top \Delta x < 0$
- For small enough t

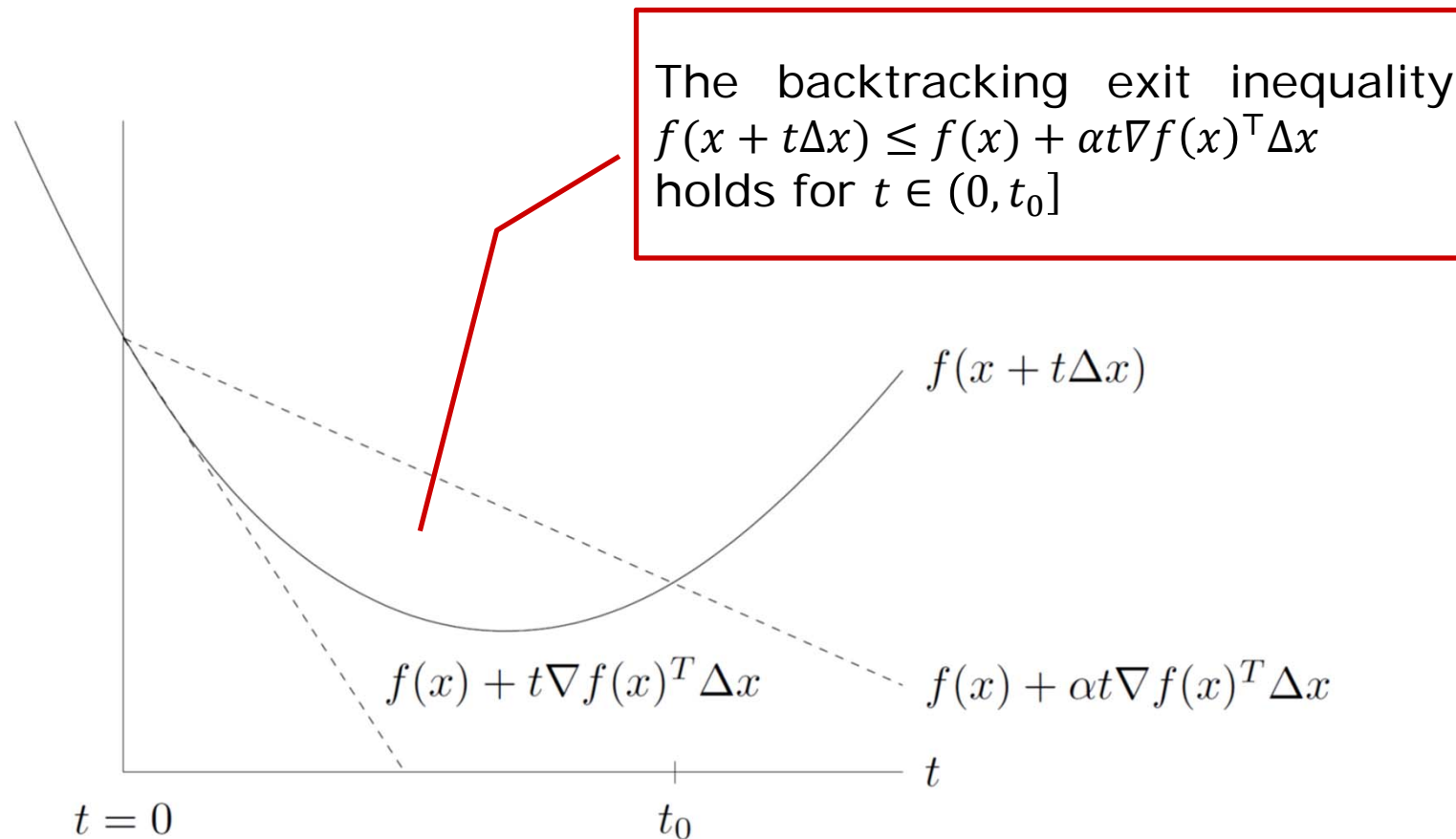
$$f(x + t\Delta x) \approx f(x) + t\nabla f(x)^\top \Delta x < f(x) + \alpha t\nabla f(x)^\top \Delta x$$

- ✓ α is the fraction of the decrease in f predicted by linear extrapolation that we will accept



Backtracking Line Search

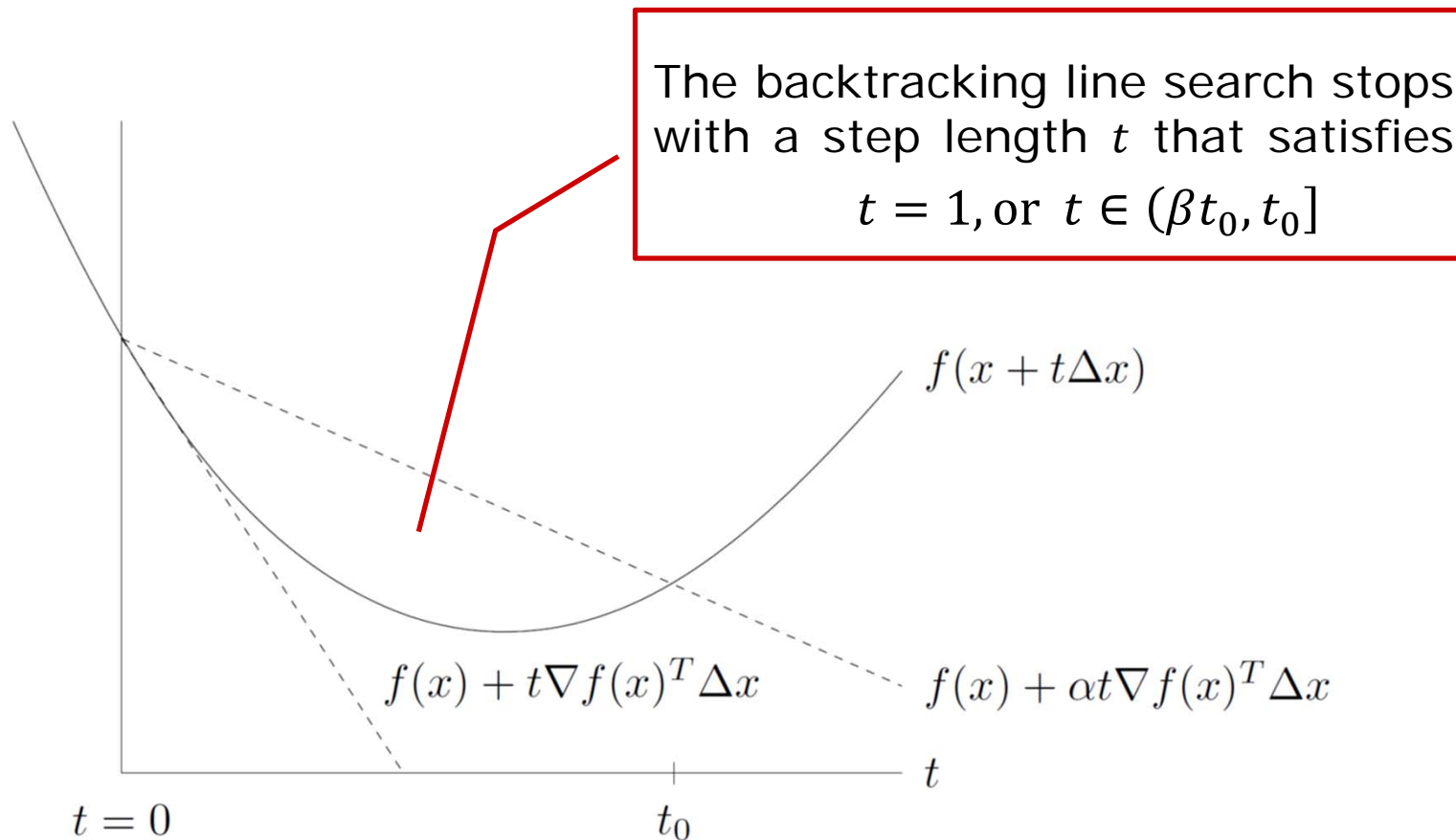
□ Graph Interpretation





Backtracking Line Search

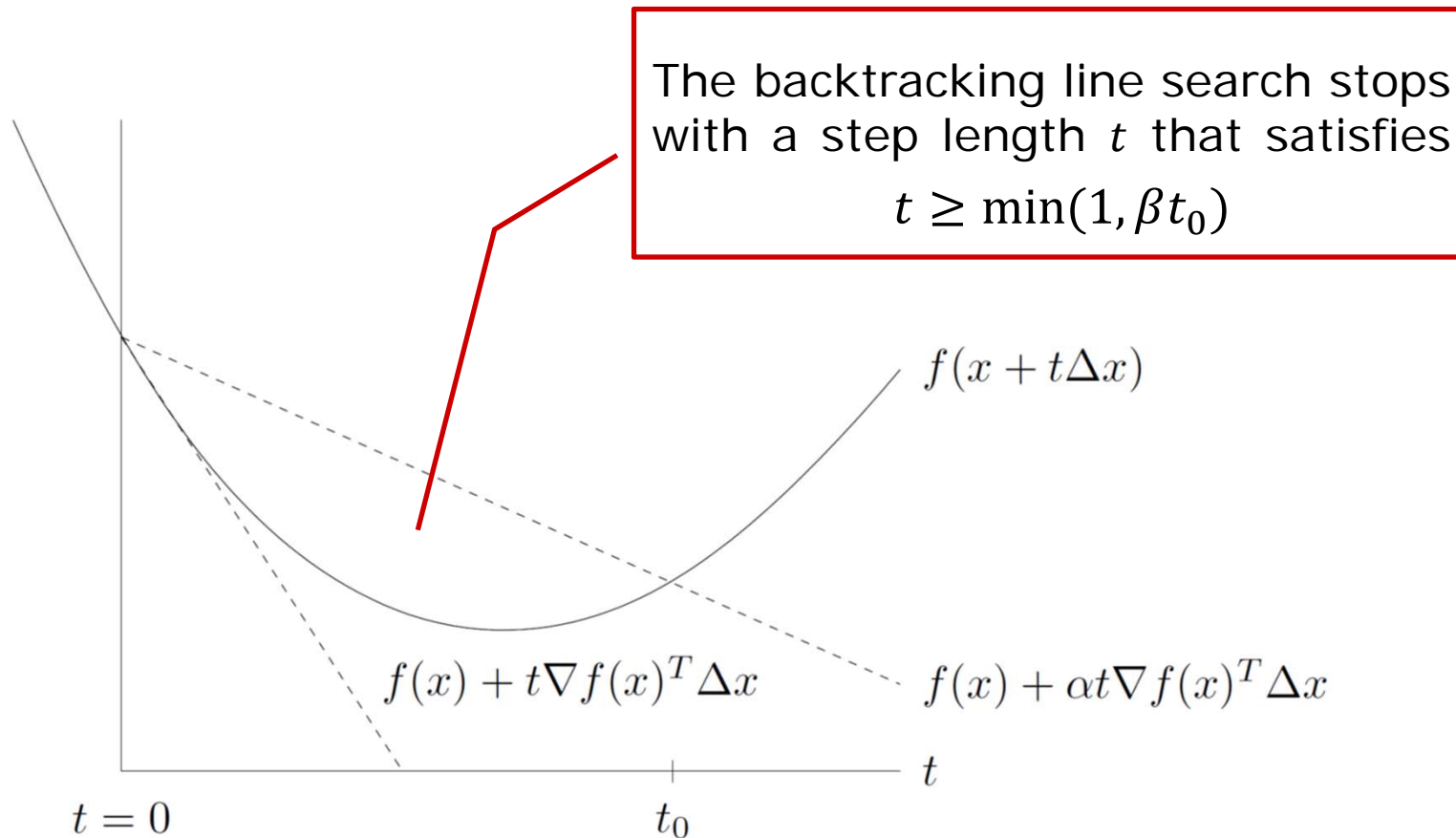
□ Graph Interpretation





Backtracking Line Search

□ Graph Interpretation





Backtracking Line Search

□ $\text{dom } f \neq \mathbf{R}^n$

$$f(x + t\Delta x) \leq f(x) + \alpha \nabla f(x)^\top \Delta x$$

■ Require $x + t\Delta x \in \text{dom } f$

□ A Practical Implementation

1. Multiply t by β until $x + t\Delta x \in \text{dom } f$
2. Check whether the above inequality holds

■ α is typically chosen between 0.01 and 0.3

■ β is often chosen between 0.1 and 0.8



Summary

□ Unconstrained Minimization Problems

- First-order Optimality Condition
- Strong Convexity and Implications
- Smoothness and Implications

□ Descent Methods

- General Descent Method
- Exact Line Search
- Backtracking Line Search