# Optimization I: Motivation, One-Variable Algorithms

CS 205A:
Mathematical Methods for Robotics, Vision, and Graphics

Doug James (and Justin Solomon)

# Optimization Objectives

| Problem | Objective |
|---|---|
| Least-squares | $E(\vec{x}) = \|A\vec{x} - \vec{b}\|_2^2$ |
| Project $\vec{b}$ onto $\vec{a}$ | $E(c) = \|c\vec{a} - \vec{b}\|_2$ |
| Eigenvectors of symmetric matrix | $E(\vec{x}) = \vec{x}^\top A\vec{x}$ |
| Pseudoinverse | $E(\vec{x}) = \|\vec{x}\|_2^2$ |
| Principal components analysis | $E(C) = \|X - CC^\top X\|_{\text{Fro}}$ |
| Broyden step | $E(J_k) = \|J_k - J_{k-1}\|_{\text{Fro}}^2$ |

# Optimization Constraints

| **Problem** | **Constraints** |
|---|---|
| Least-squares | None |
| Project $\vec{b}$ onto $\vec{a}$ | None |
| Eigenvectors of symmetric matrix | $\|\vec{x}\|_2 = 1$ |
| Pseudoinverse | $A^\top A\vec{x} = A^\top \vec{b}$ |
| Principal components analysis | $C^\top C = I_{d\times d}$ |
| Broyden step | $J_k \cdot \Delta\vec{x} = \Delta f(\vec{x})$ |

# Variational Problem-Solving

Define **objective function** measuring desirable properties and minimize it.

# General Motivation

**So far:**

Optimality conditions solvable in closed-form

# General Motivation

**So far:**

Optimality conditions solvable in closed-form

**What if:**

We're not so lucky?

# Today

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

*No constraints on* $\vec{x}$.

# Nonlinear Least-Squares

E.g. for fitting an exponential:

$$E(a, c) = \sum_i (y_i - ce^{ax_i})^2$$

# Maximum Likelihood Estimation

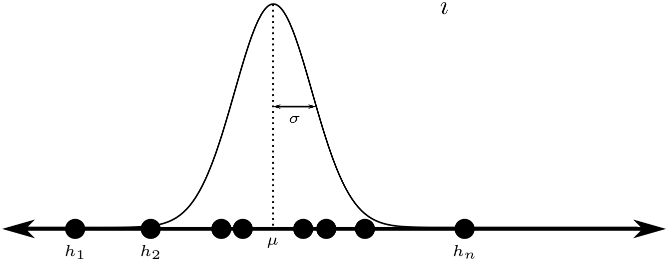
$$g(h; \mu, \sigma) \equiv \frac{1}{\sigma\sqrt{2\pi}} e^{-(h-\mu)^2/2\sigma^2}$$

# Maximum Likelihood Estimation

$$g(h;\mu,\sigma) \equiv \frac{1}{\sigma\sqrt{2\pi}} e^{-(h-\mu)^2/2\sigma^2}$$

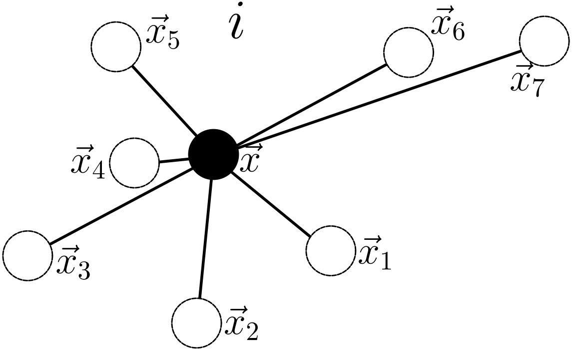
$$\Downarrow \text{ (independent sample)}$$

$$P(\{h_1,\ldots,h_n\};\mu,\sigma) = \prod_i g(h_i;\mu,\sigma)$$

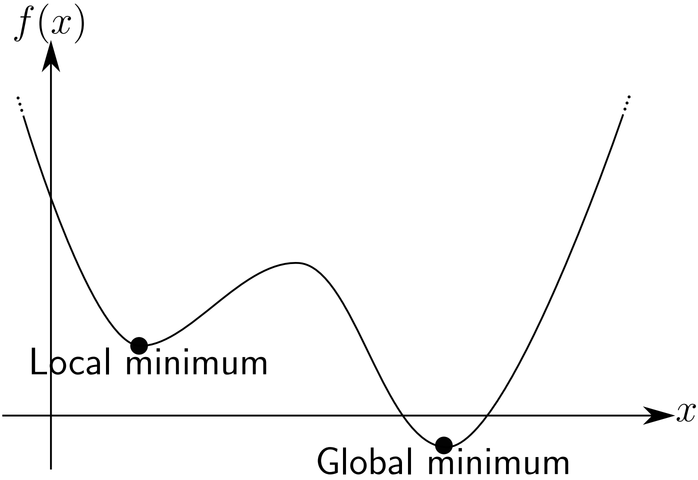


# Geometric Median Problem

$$E(\vec{x}) \equiv \sum_i \|\vec{x} - \vec{x}_i\|_2$$



# What are we looking for?



# What are we looking for?

## Global minimum

$\vec{x}^* \in \mathbb{R}^n$ is a *global minimum* of $f : \mathbb{R}^n \to \mathbb{R}$ if $f(\vec{x}^*) \leq f(\vec{x})$ for all $\vec{x} \in \mathbb{R}^n$.

# What are we looking for?

## Global minimum

$\vec{x}^* \in \mathbb{R}^n$ is a *global minimum* of $f : \mathbb{R}^n \to \mathbb{R}$ if $f(\vec{x}^*) \leq f(\vec{x})$ for all $\vec{x} \in \mathbb{R}^n$.

## Local minimum

$\vec{x}^* \in \mathbb{R}^n$ is a *local minimum* of $f : \mathbb{R}^n \to \mathbb{R}$ if $f(\vec{x}^*) \leq f(\vec{x})$ for all $\vec{x} \in \mathbb{R}^n$ satisfying $\|\vec{x} - \vec{x}^*\|_2 < \varepsilon$ for some $\varepsilon > 0$.

# Differential Optimality

$$f(\vec{x}) \approx f(\vec{x}_0) + \nabla f(\vec{x}_0) \cdot (\vec{x} - \vec{x}_0)$$

# Differential Optimality

$$f(\vec{x}) \approx f(\vec{x}_0) + \nabla f(\vec{x}_0) \cdot (\vec{x} - \vec{x}_0)$$

Take: $\vec{x} - \vec{x}_0 = \alpha \nabla f(\vec{x}_0)$:

$$f(\vec{x}_0 + \alpha \nabla f(\vec{x}_0)) \approx f(\vec{x}_0) + \alpha \|\nabla f(\vec{x}_0)\|_2$$

# Differential Optimality

$$f(\vec{x}) \approx f(\vec{x}_0) + \nabla f(\vec{x}_0) \cdot (\vec{x} - \vec{x}_0)$$

Take: $\vec{x} - \vec{x}_0 = \alpha \nabla f(\vec{x}_0)$:

$$f(\vec{x}_0 + \alpha \nabla f(\vec{x}_0)) \approx f(\vec{x}_0) + \alpha \|\nabla f(\vec{x}_0)\|_2$$

When $\|\nabla f(\vec{x}_0)\|_2 \neq 0$, the sign of $\alpha$ determines whether $f$ increases or decreases.

# Stationary Point

$$\nabla f(\vec{x}_0) = \vec{0}$$

*Doesn't change to first order*

# Typical Strategy

1. Find critical point
2. Check if it is a local minimum
3. Repeat [optional]

# Hessian

$$H_f(\vec{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial^2 x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial^2 x_n} \end{pmatrix}$$

# Hessian-Based Optimality

$$f(\vec{x}) \approx f(\vec{x}_0) + \nabla f(\vec{x}_0) \cdot (\vec{x} - \vec{x}_0) + \frac{1}{2}(\vec{x} - \vec{x}_0)^\top H_f(\vec{x} - \vec{x}_0)$$

# Hessian-Based Optimality

$$f(\vec{x}) \approx f(\vec{x}_0) + \nabla f(\vec{x}_0) \cdot (\vec{x} - \vec{x}_0) + \frac{1}{2}(\vec{x} - \vec{x}_0)^\top H_f(\vec{x} - \vec{x}_0)$$

- $H_f$ is *positive definite* $\implies$ local minimum
- $H_f$ is *negative definite* $\implies$ local maximum
- $H_f$ is *indefinite* $\implies$ saddle point
- $H_f$ is *not invertible* $\implies$

# Hessian-Based Optimality

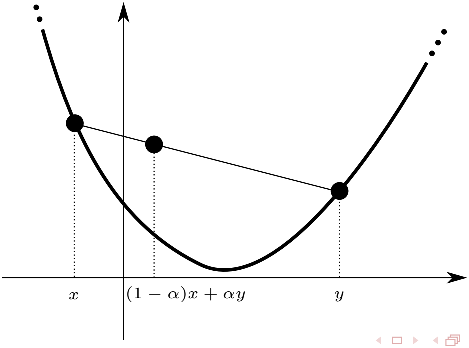
$$f(\vec{x}) \approx f(\vec{x}_0) + \nabla f(\vec{x}_0) \cdot (\vec{x} - \vec{x}_0) + \frac{1}{2}(\vec{x} - \vec{x}_0)^\top H_f(\vec{x} - \vec{x}_0)$$

- $H_f$ is *positive definite* $\Longrightarrow$ local minimum
- $H_f$ is *negative definite* $\Longrightarrow$ local maximum
- $H_f$ is *indefinite* $\Longrightarrow$ saddle point
- $H_f$ is *not invertible* $\Longrightarrow$ **nothing**

# Alternative Optimality

## Convex

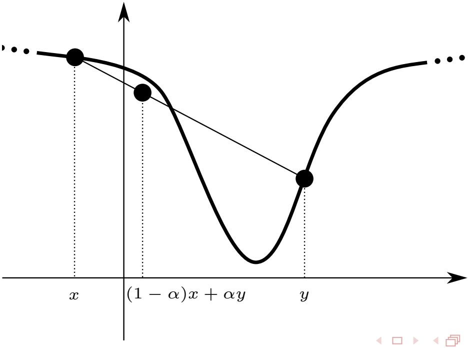
$f : \mathbb{R}^m \to \mathbb{R}$ is *convex* when for all $\vec{x}, \vec{y} \in \mathbb{R}^m$ and $\alpha \in (0,1)$, $f((1-\alpha)\vec{x} + \alpha\vec{y}) \leq (1-\alpha)f(\vec{x}) + \alpha f(\vec{y})$.



# Alternative Optimality

## Quasi-Convex

$f : \mathbb{R}^m \to \mathbb{R}$ is *convex* when for all $\vec{x}, \vec{y} \in \mathbb{R}^m$ and $\alpha \in (0,1)$, $f((1-\alpha)\vec{x} + \alpha\vec{y}) \leq \max(f(\vec{x}), f(\vec{y}))$.



# Newton's Method

Minimize $f \leftrightarrow$ find roots of $f'$

# Newton's Method

Minimize $f \leftrightarrow$ find roots of $f'$

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

# Newton's Method

Minimize $f \leftrightarrow$ find roots of $f'$

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

**Question:**

What do you need for secant?

# Newton's Method

Minimize $f \leftrightarrow$ find roots of $f'$

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

**Question:**

What do you need for secant?

**Alternative:** Successive parabolic interpolation

# Imitate Bisection?

**Question:**

Analog of Intermediate Value Theorem?

# Imitate Bisection?

**Question:**

Analog of Intermediate Value Theorem?

## Unimodular

$f : [a, b] \to \mathbb{R}$ is *unimodular* if there exists $x^* \in [a, b]$ such that $f$ is decreasing for $x \in [a, x^*]$ and increasing for $x \in [x^*, b]$.

# Observations about Unimodular Functions

- $f(x_0) \geq f(x_1) \implies f(x) \geq f(x_1)$ for all $x \in [a, x_0] \implies [a, x_0]$ can be discarded
- $f(x_1) \geq f(x_0) \implies f(x) \geq f(x_0)$ for all $x \in [x_1, b] \implies [x_1, b]$ can be discarded

# Unimodular Optimization v1.0

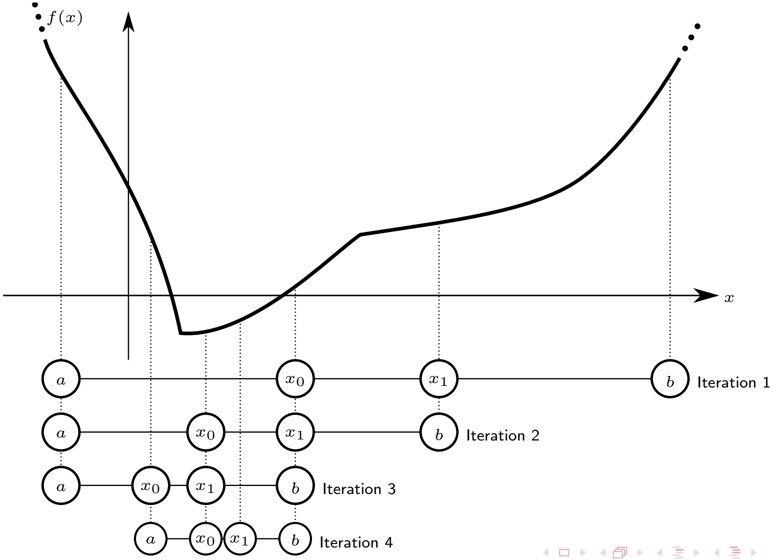
Iteratively remove $1/3$ of interval in each iteration.

# Unimodular Optimization v1.0

Iteratively remove $1/3$ of interval in each iteration.

Requires two evaluations per iteration.

# New Idea



f(x)
x
a
x0
x1
b
Iteration 1
a
x0
x1
b
Iteration 2
a
x0
x1
b
Iteration 3
a
x0
x1
b
Iteration 4

# Reuse Evaluations?

$$x_0 = \alpha \quad x_1 = 1 - \alpha$$

$$\alpha \in (0, 1/2)$$

# Reuse Evaluations?

$$x_0 = \alpha \quad x_1 = 1 - \alpha$$

$$\alpha \in (0, 1/2)$$

Remove right interval $[x_1, b]$
$\longrightarrow$new interval is $[0, 1 - \alpha]$.

# Reuse Evaluations?

$$x_0 = \alpha \quad x_1 = 1 - \alpha$$

$$\alpha \in (0, 1/2)$$

Remove right interval $[x_1, b]$
$\longrightarrow$new interval is $[0, 1 - \alpha]$.

**New bounds:**

$$\tilde{x}_0 = \alpha(1 - \alpha)$$

$$\tilde{x}_1 = (1 - \alpha)^2$$

# Reuse Evaluations?

**To reuse:** $(1-\alpha)^2 = \alpha$

$$\implies \quad \alpha = \frac{1}{2}(3-\sqrt{5})$$

$$1-\alpha = \frac{1}{2}(\sqrt{5}-1) \equiv \tau$$

# Reuse Evaluations?

**To reuse:** $(1-\alpha)^2 = \alpha$

$$\implies \quad \alpha = \frac{1}{2}(3-\sqrt{5})$$

$$1-\alpha = \frac{1}{2}(\sqrt{5}-1) \equiv \tau$$

**Golden ratio...**

# Golden Section Search

1. Initialize $a$ and $b$ so that $f$ is unimodular on $[a, b]$.
2. Take $x_0 = a + (1 - \tau)(b - a), x_1 = a + \tau(b - a)$; initialize $f_0 = f(x_0), f_1 = f(x_1)$.
3. Iterate until $b - a$ is sufficiently small:

   3.1 If $f_0 \geq f_1$, then remove the interval $[a, x_0]$:
   - Move left side: $a \leftarrow x_0$
   - Reuse previous iteration: $x_0 \leftarrow x_1$, $f_0 \leftarrow f_1$
   - Generate new sample: $x_1 \leftarrow a + \tau(b - a)$, $f_1 \leftarrow f(x_1)$

   3.2 If $f_1 > f_0$, then remove the interval $[x_1, b]$:
   - Move right side: $b \leftarrow x_1$
   - Reuse previous iteration: $x_1 \leftarrow x_0$, $f_1 \leftarrow f_0$
   - Generate new sample: $x_0 \leftarrow a + (1 - \tau)(b - a)$, $f_0 \leftarrow f(x_0)$

Next