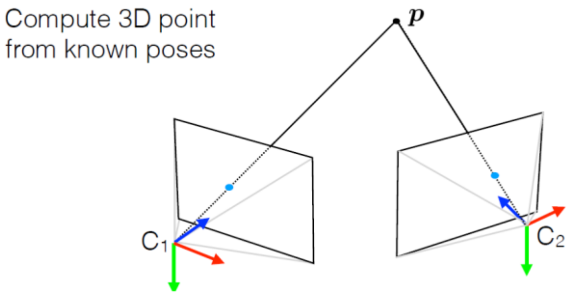
Introduction to Nonlinear Least Squares

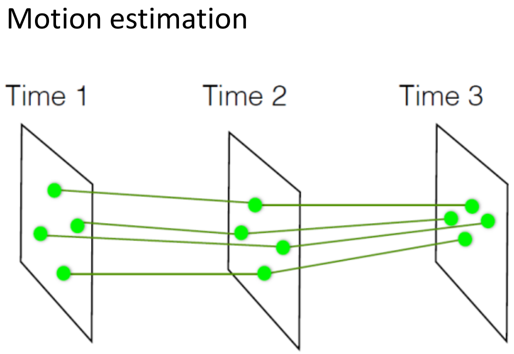
Rajat Talak

VNAV

Fall 2020

Recall ...

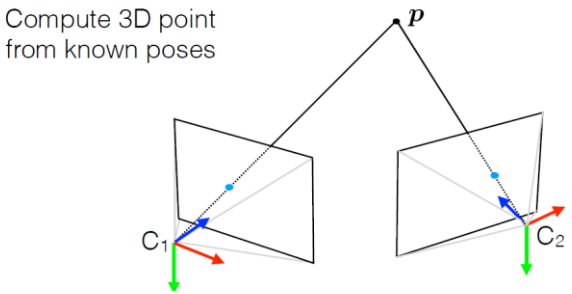


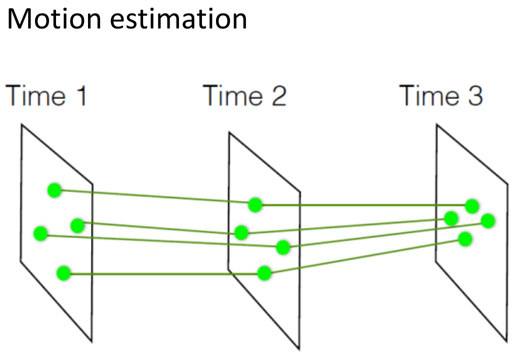


In the previous lecture:

- Perception problem can systematically formulated using estimation theory

Recall …

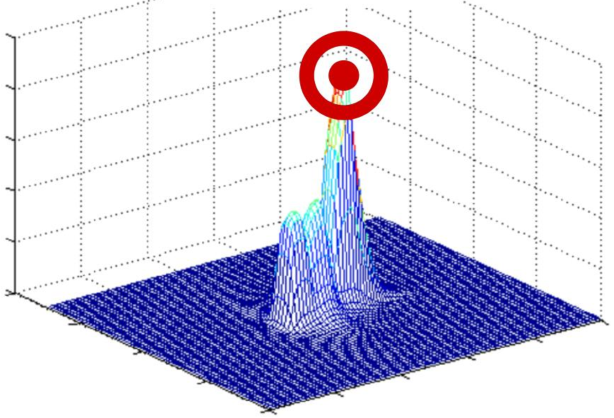




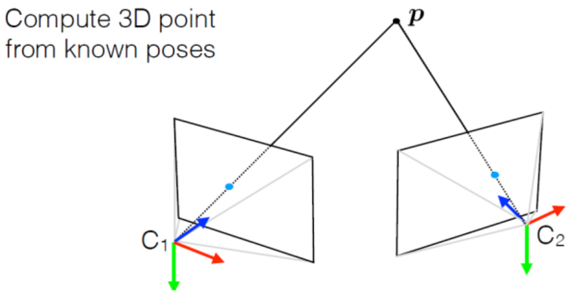
In the previous lecture:

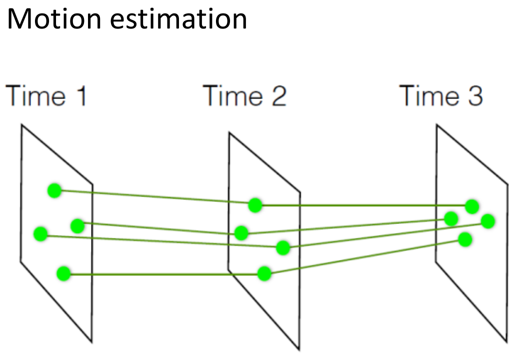
- Perception problem can systematically formulated using estimation theory
- Estimation theory:
  - (1) Maximum likelihood (ML) estimate,
  - (2) Maximum a-postiriori (MAP) estimate

$$\mathbb{P}(y_1, \dots y_N | x)$$



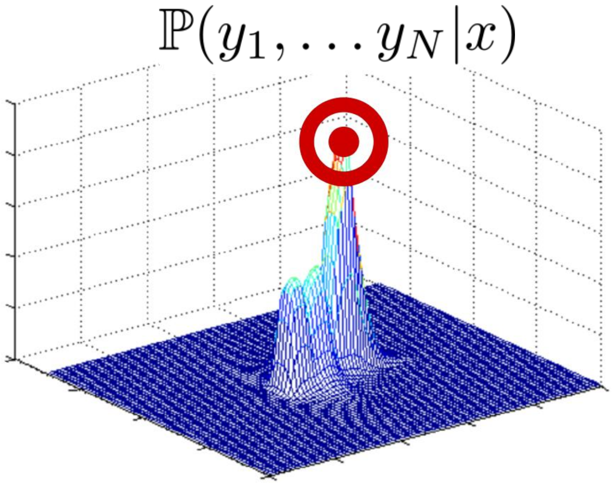
Recall …





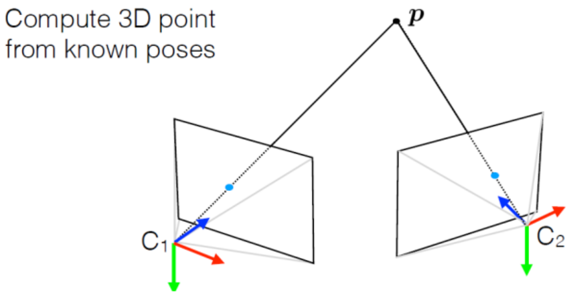
In the previous lecture:

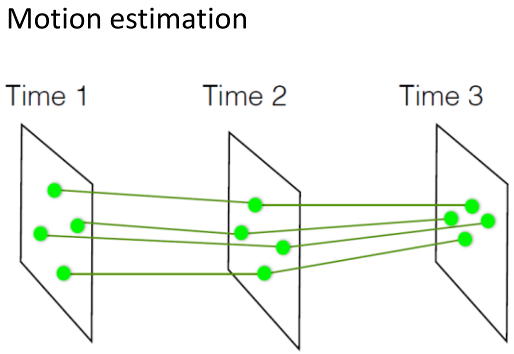
- Perception problem can systematically formulated using estimation theory
- Estimation theory:
  - (1) Maximum likelihood (ML) estimate,
  - (2) Maximum a-postiriori (MAP) estimate



Which is this? ML or MAP?

Recall …

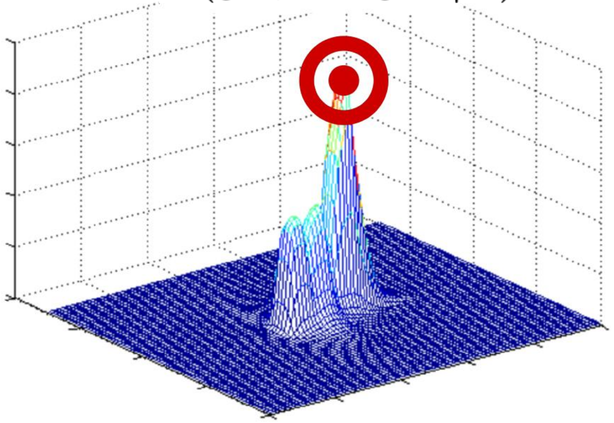




In the previous lecture:

- Perception problem can systematically formulated using estimation theory
- Estimation theory:
  - (1) Maximum likelihood (ML) estimate,
  - (2) Maximum a-postiriori (MAP) estimate

$$\mathbb{P}(y_1, \dots y_N | x)$$



- Abstract Model:

$$y_i = f_i(x) + n_i$$

If $n_i \sim \mathcal{N}(0, \Sigma_i)$
and independent across i

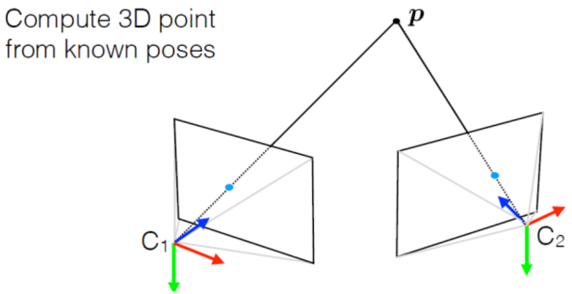
x state variable

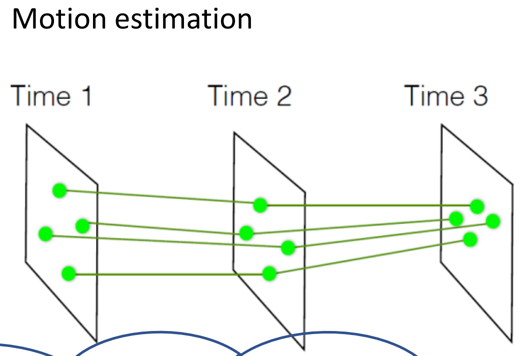
y_i measurements

n_i noise

$$\hat{x} = \arg\min_x \sum_i ||y_i - f_i(x)||^2_{\Sigma_i}$$

Recall …





In the previous lecture:

- Perception problem can systematically formulated using estimation theory
- Estimation theory:
  - (1) Maximum likelihood (ML) estimate,
  - (2) Maximum a-postiriori (MAP) estimate
- Abstract Model:

$$y_i = f_i(x) + n_i$$

If $n_i \sim \mathcal{N}(0, \Sigma_i)$
and independent across i

x state variable

y_i measurements

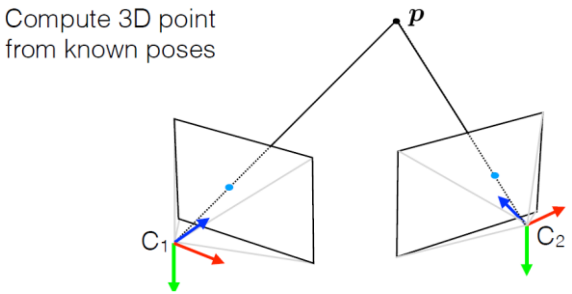
n_i noise

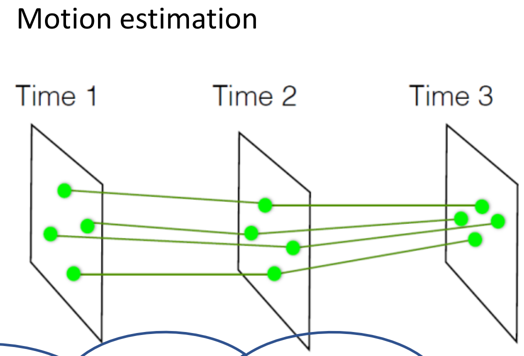
$$\hat{x} = \arg\min_x \sum_i ||y_i - f_i(x)||^2_{\Sigma_i}$$

$$||z||_S = \sqrt{z^T S^{-1} z} \qquad S \succ 0$$

is called Mahalanobis distance.

Recall …





In the previous lecture:

- Perception problem can systematically formulated using estimation theory
- Estimation theory:
  - (1) Maximum likelihood (ML) estimate,
  - (2) Maximum a-postiriori (MAP) estimate
- Linear Model:

$$y_i = A_i x + n_i$$

x state variable

y_i measurements

n_i noise

If $n_i \sim \mathcal{N}(0, \Sigma_i)$
and independent across i

$$\hat{x} = \arg\min_x \sum_i ||y_i - A_i x||^2_{\Sigma_i}$$

$$||z||_S = \sqrt{z^T S^{-1} z} \qquad S \succ 0$$

is called Mahalanobis distance.

Today

- Nonlinear least squares problem
- Gauss-Newton Method

A quick detour

- Nonlinear optimization
- Convexity
- Optimality conditions
- Gradient descent and Newton’s method

Nonlinear Least Squares Problem

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x)||^2 = \sum_{i=1}^{m} |r_i(x)|^2$$

- $r : \mathbb{R}^n \to \mathbb{R}^m$ and $r(x) = [r_1(x), r_2(x), \ldots r_m(x)]^T$
- $r_i(x)$ is the residual function

Nonlinear Least Squares Problem

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x)||^2 = \sum_{i=1}^{m} |r_i(x)|^2$$

- $r : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $r(x) = [r_1(x), r_2(x), \ldots r_m(x)]^T$
- $r_i(x)$ is the residual function
- For our abstract model $\quad y_i = f_i(x) + n_i$

$$r_i(x) = \Sigma_i^{-\frac{1}{2}}(y_i - f_i(x))$$

- Linear model $\quad y_i = A_i x + n_i$

$$r_i(x) = \Sigma_i^{-\frac{1}{2}}(y_i - A_i x)$$

Nonlinear Least Squares Problem

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x)||^2 = \sum_{i=1}^{m} |r_i(x)|^2$$

- $r : \mathbb{R}^n \to \mathbb{R}^m$ and $r(x) = [r_1(x), r_2(x), \ldots r_m(x)]^T$
- $r_i(x)$ is the residual function
- if $r(x) = Ax - b$ we call it linear least squares problem

Nonlinear Least Squares Problem

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x)||^2 = \sum_{i=1}^{m} |r_i(x)|^2$$

- $r : \mathbb{R}^n \to \mathbb{R}^m$ and $r(x) = [r_1(x), r_2(x), \ldots r_m(x)]^T$
- $r_i(x)$ is the residual function
- if $r(x) = Ax - b$ we call it linear least squares problem

Question: How do we solve this?

Nonlinear Least Squares Problem

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x)||^2 = \sum_{i=1}^{m} |r_i(x)|^2$$

Nonlinear optimization problem

- $r : \mathbb{R}^n \to \mathbb{R}^m$ and $r(x) = [r_1(x), r_2(x), \ldots r_m(x)]^T$
- $r_i(x)$ is the residual function
- if $r(x) = Ax - b$ we call it linear least squares problem

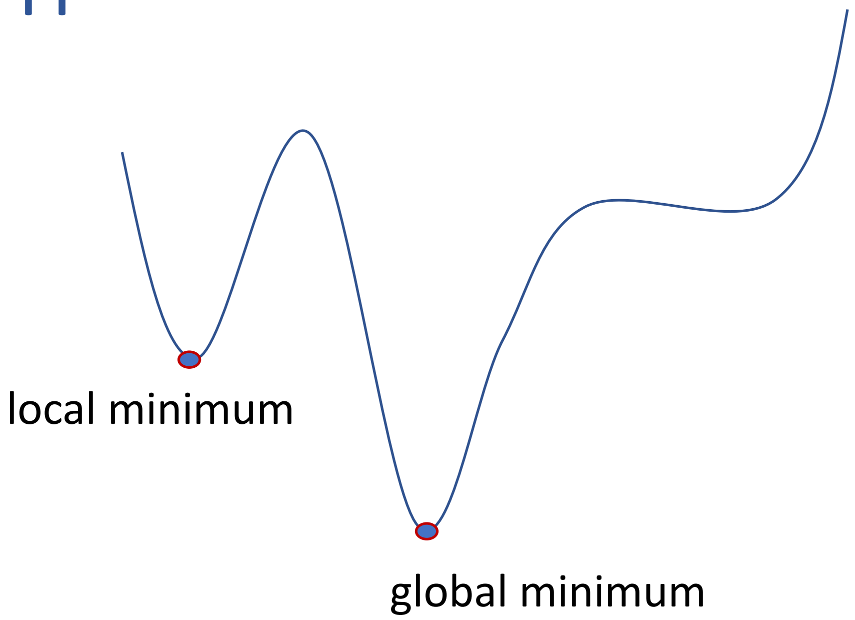
Question: How do we solve this?

Nonlinear Optimization Problem

- Unconstrained nonlinear optimization problem:

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad g(x)$$

$$g : \mathbb{R}^n \to \mathbb{R}$$



- Global minimum:

$$x^* \text{ is global minimum iff } g(x^*) \leq g(x) \text{ for all } x \in \mathbb{R}^n$$

- Local minimum:

$$x^* \text{ is a local minimum iff } \exists\ r > 0 \text{ s.t. } g(x^*) \leq g(x) \text{ for all } x \in \mathcal{B}(x^*, r)$$

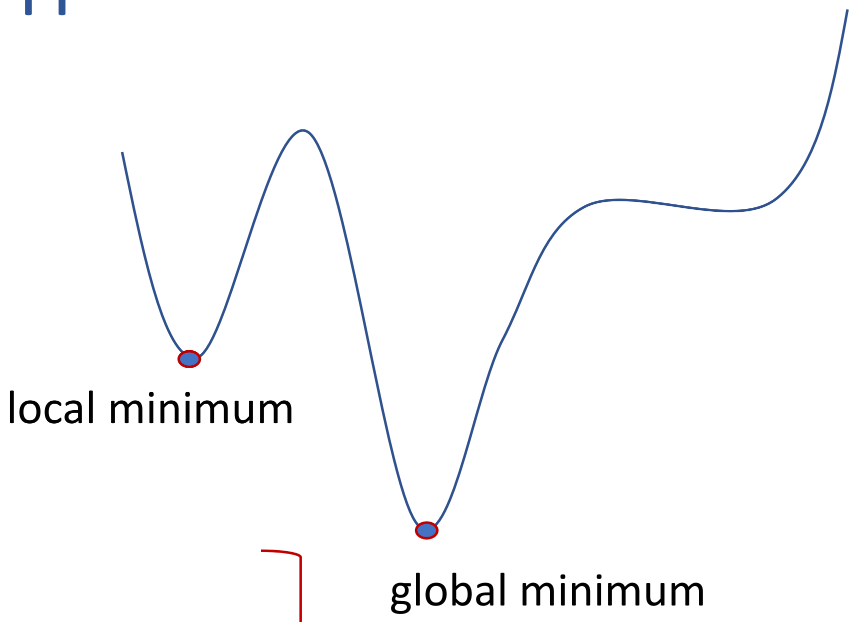
$$\mathcal{B}(x, r) = \{z \in \mathbb{R}^n \mid ||x - z|| \leq r\}$$

Nonlinear Optimization Problem

- Unconstrained nonlinear optimization problem:

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad g(x)$$

$$g : \mathbb{R}^n \to \mathbb{R}$$



- Necessary conditions for local minimum

$$x \text{ is a local minimum} \implies g'(x) = 0 \text{ and } g''(x) \geq 0$$

- Sufficient conditions for local minimum

$$g'(x) = 0 \text{ and } g''(x) > 0 \implies x \text{ is a local minimum of } g$$

$$n = 1$$

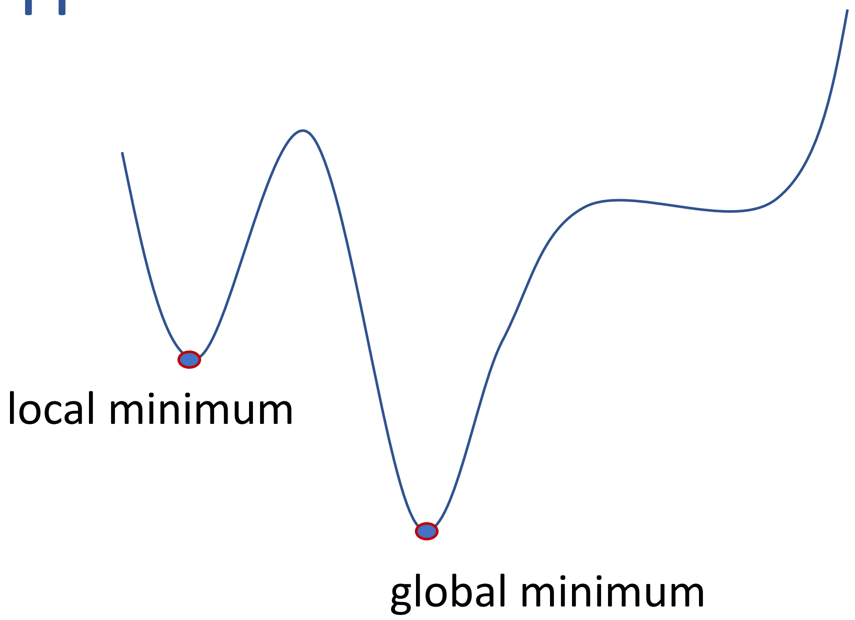
$$g : \mathbb{R} \to \mathbb{R}$$

Nonlinear Optimization Problem

- Unconstrained nonlinear optimization problem:

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad g(x)$$

$$g : \mathbb{R}^n \to \mathbb{R}$$



- Necessary conditions for local minimum

$$x \text{ is a local minimum} \implies \nabla g(x) = 0 \text{ and } \nabla^2 g(x) \succeq 0$$

- Sufficient conditions for local minimum

$$\nabla g(x) = 0 \text{ and } \nabla^2 g(x) \succ 0 \implies x \text{ is a local minimum of } g$$

Recall

$$\nabla g(x) = \begin{pmatrix} \frac{\partial g}{\partial x_1} \\ \frac{\partial g}{\partial x_2} \\ \vdots \\ \frac{\partial g}{\partial x_n} \end{pmatrix}$$

$$\nabla^2 g(x) = \begin{pmatrix} \frac{\partial^2 g}{\partial x_1^2} & \frac{\partial^2 g}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 g}{\partial x_1 \partial x_n} \\ \frac{\partial^2 g}{\partial x_2 \partial x_1} & \frac{\partial^2 g}{\partial x_2^2} & \cdots & \frac{\partial^2 g}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 g}{\partial x_n \partial x_1} & \frac{\partial^2 g}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 g}{\partial x_n^2} \end{pmatrix}$$

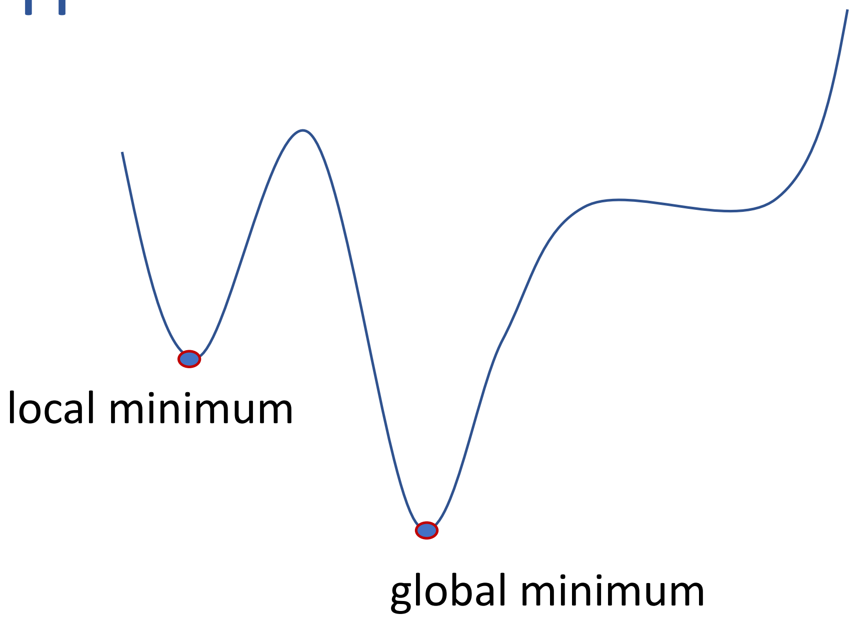
Hessian

Nonlinear Optimization Problem

- Unconstrained nonlinear optimization problem:

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad g(x)$$

$$g : \mathbb{R}^n \to \mathbb{R}$$



- Necessary conditions for local minimum

$$x \text{ is a local minimum} \implies \nabla g(x) = 0 \text{ and } \nabla^2 g(x) \succeq 0$$

- Sufficient conditions for local minimum

$$\nabla g(x) = 0 \text{ and } \nabla^2 g(x) \succ 0 \implies x \text{ is a local minimum of } g$$

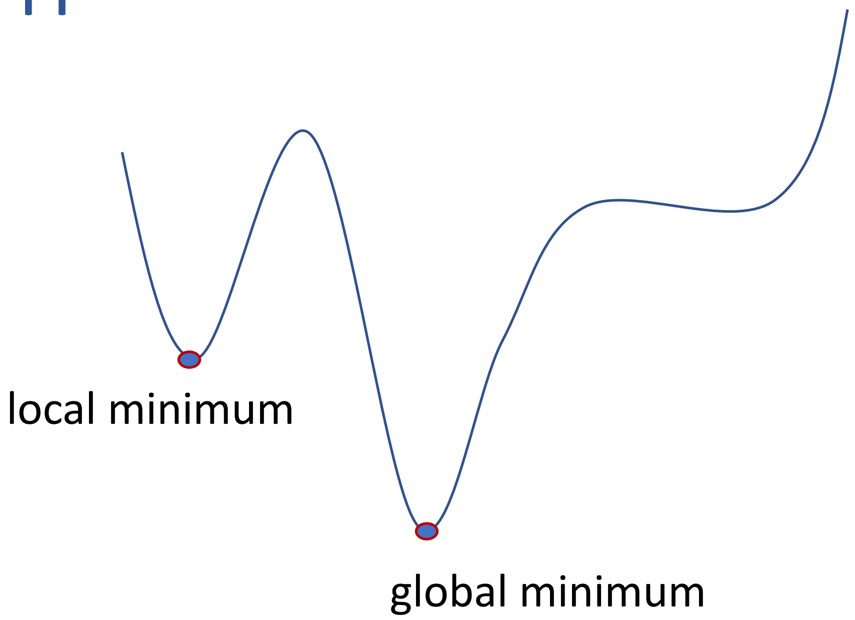
- Gradient descent converges to local minimum $\quad x_{t+1} = x_t - \alpha_t \nabla g(x_t)$

Nonlinear Optimization Problem

- Unconstrained nonlinear optimization problem:

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad g(x)$$

$$g : \mathbb{R}^n \to \mathbb{R}$$



- Necessary conditions for local minimum

$$x \text{ is a local minimum} \implies \nabla g(x) = 0 \text{ and } \nabla^2 g(x) \succeq 0$$

- Sufficient conditions for local minimum

$$\nabla g(x) = 0 \text{ and } \nabla^2 g(x) \succ 0 \implies x \text{ is a local minimum of } g$$

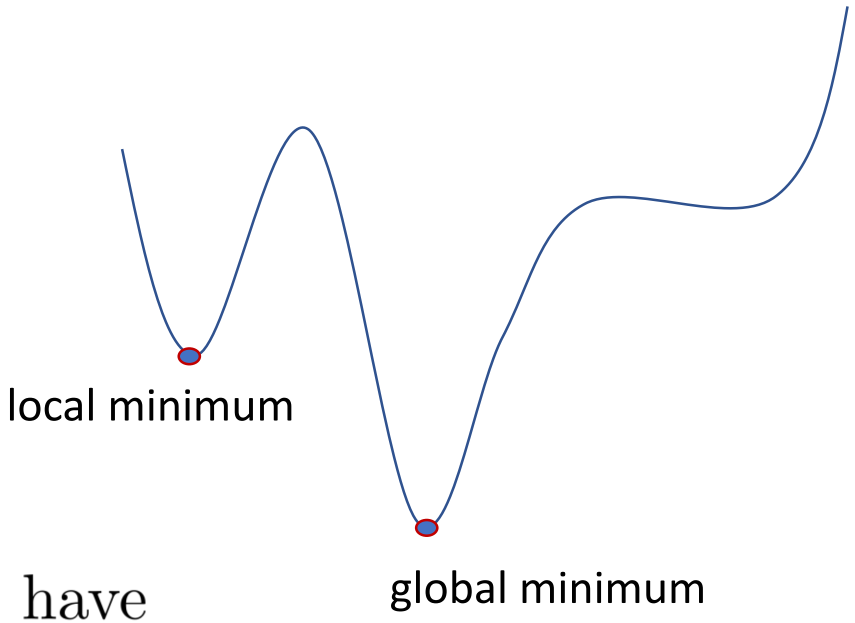
Finding global minimum is hard!! … possible with an added structure of convexity

Convex Problems

- Convex optimization problem:

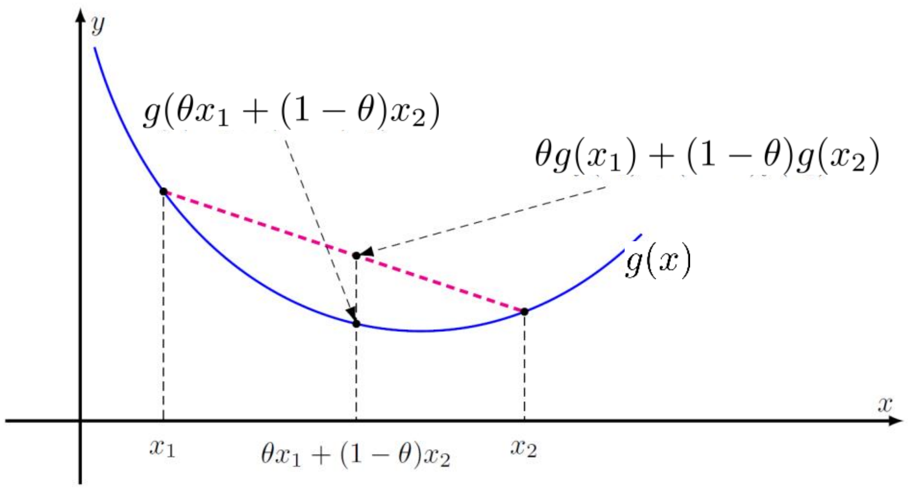
$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad g(x)$$

$g : \mathbb{R}^n \to \mathbb{R}$



- g is convex iff for all $x_1, x_2 \in \mathbb{R}^n$ and $\theta \in [0,1]$ we have

$$g(\theta x_1 + (1-\theta)x_2) \leq \theta g(x_1) + (1-\theta) g(x_2)$$

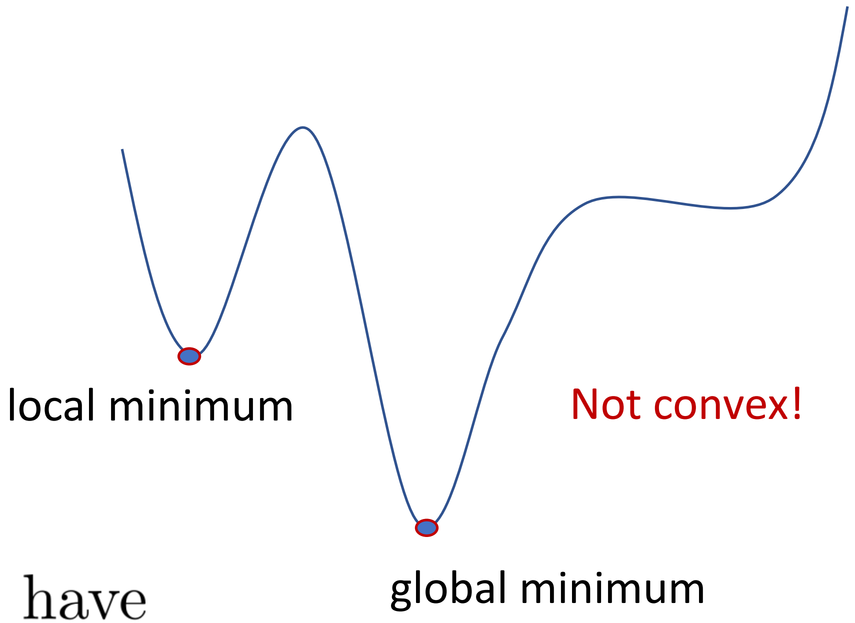


Convex Problems

- Convex optimization problem:

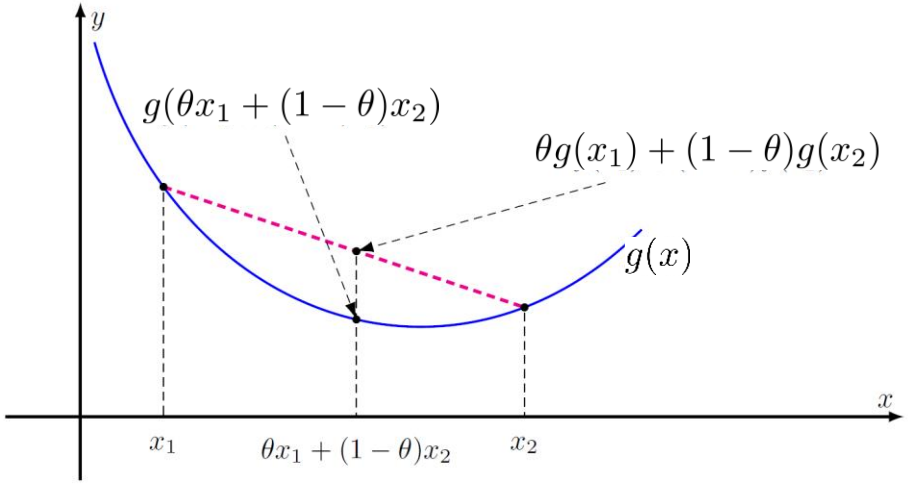
$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad g(x)$$

$$g : \mathbb{R}^n \to \mathbb{R}$$

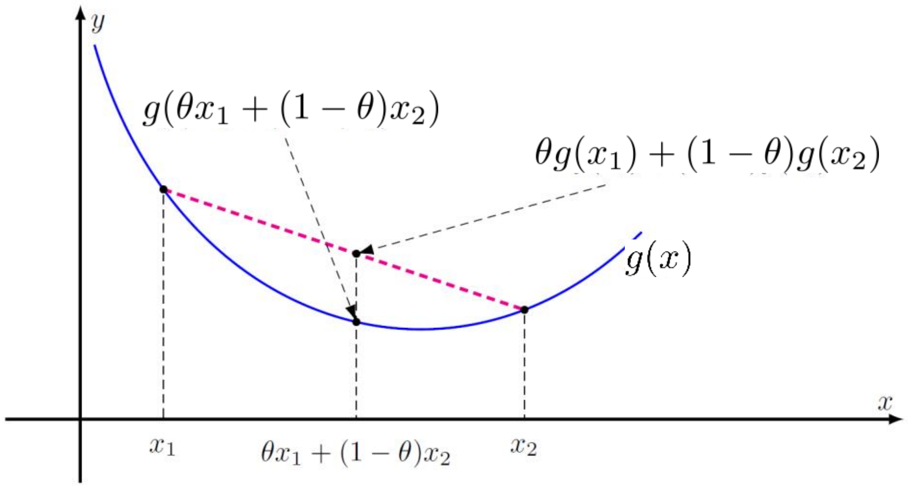


- g is convex iff for all $x_1, x_2 \in \mathbb{R}^n$ and $\theta \in [0,1]$ we have

$$g(\theta x_1 + (1-\theta)x_2) \leq \theta g(x_1) + (1-\theta) g(x_2)$$



Convex Problems



- Convex optimization problem:

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad g(x)$$

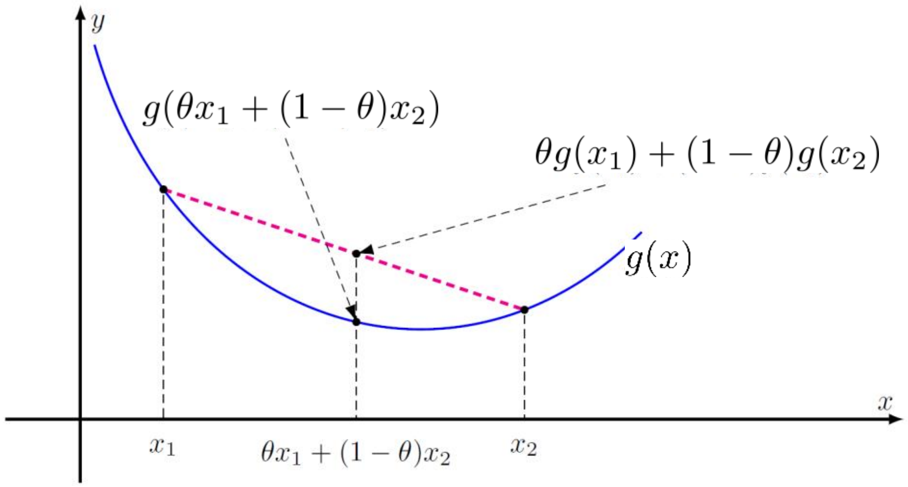
$$g : \mathbb{R}^n \to \mathbb{R}$$

- g is convex iff for all $x_1, x_2 \in \mathbb{R}^n$ and $\theta \in [0,1]$ we have

$$g(\theta x_1 + (1-\theta)x_2) \leq \theta g(x_1) + (1-\theta)g(x_2)$$

- g is convex iff for all $x, y \in \mathbb{R}^n$ $\quad g(y) \geq g(x) + \nabla g(x)^T (y - x)$

Convex Problems



- Convex optimization problem:

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad g(x)$$

$$g : \mathbb{R}^n \to \mathbb{R}$$

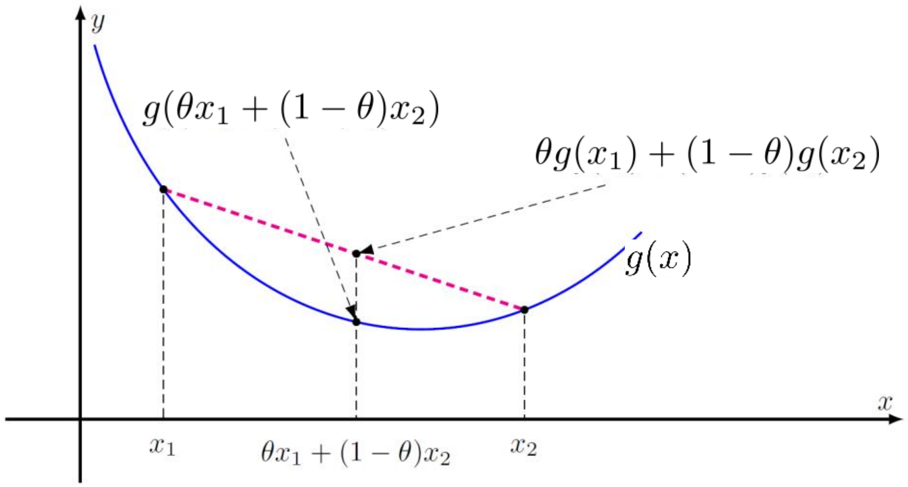
- g is convex iff for all $x_1, x_2 \in \mathbb{R}^n$ and $\theta \in [0,1]$ we have

$$g(\theta x_1 + (1-\theta)x_2) \leq \theta g(x_1) + (1-\theta)g(x_2)$$

- g is convex iff for all $x, y \in \mathbb{R}^n$ $\quad g(y) \geq g(x) + \nabla g(x)^T(y-x)$

- g is convex iff for all $x \in \mathbb{R}^n$ $\quad \nabla^2 g(x) \succeq 0$

Convex Problems



- Convex optimization problem:

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad g(x)$$

$$g : \mathbb{R}^n \to \mathbb{R}$$

- Local minima is also a global minima

- Necessary and sufficient condition

$$x \text{ is a global minima} \Leftrightarrow \nabla g(x) = 0 \text{ and } \nabla^2 g(x) \succeq 0$$

- Gradient descent converges to global minima

$$x_{t+1} = x_t - \alpha_t \nabla g(x_t)$$

Back to Nonlinear Least Squares Problem

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x)||^2 = \sum_{i=1}^{m} |r_i(x)|^2$$

- $r : \mathbb{R}^n \to \mathbb{R}^m$ and $r(x) = [r_1(x), r_2(x), \ldots r_m(x)]^T$
- $r_i(x)$ is the residual function
- if $r(x) = Ax - b$ we call it linear least squares problem

Linear Least Squares Problem

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||Ax - b||^2$$

- $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$
- The objective function is convex!

$$\nabla^2 g(x) = 2A^T A \succeq 0$$

- Gradient descent algorithm converges to the global minimum

$$x_{t+1} = x_t - 2\alpha_t A^T (Ax_t - b)$$

- But, we can do much better (computationally) by exploiting the problem structure and using the optimality conditions

Linear Least Squares Problem

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||Ax - b||^2$$

- $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$
- The objective function is convex!

$$\nabla^2 g(x) = 2A^T A \succeq 0$$

- Recall: x is a global minima $\Leftrightarrow \nabla g(x) = 0$ and $\nabla^2 g(x) \succeq 0$

Linear Least Squares Problem

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||Ax - b||^2$$

- $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$
- The objective function is convex!

  $$\nabla^2 g(x) = 2A^T A \succeq 0$$

- Recall: x is a global minima $\Leftrightarrow \nabla g(x) = 0$ and $\nabla^2 g(x) \succeq 0$
- $\nabla g(x) = A^T A x - A^T b$
- x is a global minima $\Leftrightarrow A^T A x = A^T b$

Linear Least Squares Problem

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||Ax - b||^2$$

- $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$
- The objective function is convex!

$$\nabla^2 g(x) = 2A^T A \succeq 0$$

- Recall: x is a global minima $\Leftrightarrow \nabla g(x) = 0$ and $\nabla^2 g(x) \succeq 0$
- $\nabla g(x) = A^T Ax - A^T b$
- x is a global minima $\Leftrightarrow$ $\boxed{A^T Ax = A^T b}$

suffices to solve this linear system of equations

Linear Least Squares Problem

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||Ax - b||^2$$

- $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$
- The objective function is convex!

$$\nabla^2 g(x) = 2A^T A \succeq 0$$

- Recall: x is a global minima $\Leftrightarrow \nabla g(x) = 0$ and $\nabla^2 g(x) \succeq 0$
- $\nabla g(x) = A^T Ax - A^T b$
- x is a global minima $\Leftrightarrow \boxed{A^T Ax = A^T b}$

suffices to solve this linear system of equations

Do not invert!

Cholesky Solver

$$(A^T A)x = A^T b$$

- Assuming $A^T A \succ 0$

Cholesky Solver

$$L = \begin{pmatrix} \ell_{11} & 0 & 0 \\ \ell_{21} & \ell_{22} & 0 \\ \ell_{31} & \ell_{32} & \ell_{33} \end{pmatrix}$$

Illustrative example

$$(A^T A)x = A^T b$$

- Assuming $A^T A \succ 0$
- Cholesky decomposition of $A^T A$

$$A^T A = LL^T$$

where L is a lower triangular and thus L^T is an upper triangular matrix

Cholesky Solver

$$(A^T A)x = A^T b$$

$$L = \begin{pmatrix} \ell_{11} & 0 & 0 \\ \ell_{21} & \ell_{22} & 0 \\ \ell_{31} & \ell_{32} & \ell_{33} \end{pmatrix}$$

Illustrative example

- Assuming $A^T A \succ 0$
- Cholesky decomposition of $A^T A$

$$A^T A = LL^T$$

where L is a lower triangular and thus L^T is an upper triangular matrix

- We now have to solve $LL^T x = A^T b$. We solve it in two steps.

Cholesky Solver

$$L = \begin{pmatrix} \ell_{11} & 0 & 0 \\ \ell_{21} & \ell_{22} & 0 \\ \ell_{31} & \ell_{32} & \ell_{33} \end{pmatrix}$$

Illustrative example

$$(A^T A)x = A^T b$$

- Assuming $A^T A \succ 0$

- Cholesky decomposition of $A^T A$

$$A^T A = LL^T$$

where L is a lower triangular and thus L^T is an upper triangular matrix

- We now have to solve $LL^T x = A^T b$. We solve it in two steps.

- Forward substitution: $Ly = A^T b$ and obtain y
- Backward substitution: $L^T x = y$ and obtain x

QR Solver

$$(A^T A)x = A^T b$$

QR Solver

$$(A^T A)x = A^T b$$

- Perform QR factorization of $A^T A$

$$A^T A = QR$$

where $Q \in \mathbb{R}^{n \times n}$ s.t. $Q^T Q = I$ and $R \in \mathbb{R}^{n \times n}$ is upper triangular

QR Solver

$$(A^T A)x = A^T b$$

- Perform QR factorization of $A^T A$

$$A^T A = QR$$

where $Q \in \mathbb{R}^{n \times n}$ s.t. $Q^T Q = I$ and $R \in \mathbb{R}^{n \times n}$ is upper triangular

- Have to now solve $QRx = A^T b$

QR Solver

$$(A^T A)x = A^T b$$

- Perform QR factorization of $A^T A$

$$A^T A = QR$$

where $Q \in \mathbb{R}^{n \times n}$ s.t. $Q^T Q = I$ and $R \in \mathbb{R}^{n \times n}$ is upper triangular

- Have to now solve $QRx = A^T b$

multiply both sides by Q^T

QR Solver

$$(A^T A)x = A^T b$$

- Perform QR factorization of $A^T A$

$$A^T A = QR$$

where $Q \in \mathbb{R}^{n \times n}$ s.t. $Q^T Q = I$ and $R \in \mathbb{R}^{n \times n}$ is upper triangular

- Have to now solve $QRx = A^T b$ — multiply both sides by Q^T
- Equivalent to solving $Rx = Q^T A^T b$

QR Solver

$$(A^T A)x = A^T b$$

- Perform QR factorization of $A^T A$

$$A^T A = QR$$

where $Q \in \mathbb{R}^{n \times n}$ s.t. $Q^T Q = I$ and $R \in \mathbb{R}^{n \times n}$ is upper triangular

- Have to now solve $QRx = A^T b$ — multiply both sides by Q^T
- Equivalent to solving $Rx = Q^T A^T b$

can be solved by backward substitution

Cholesky vs QR Solver

$$(A^T A)x = A^T b$$

- QR is slower than Cholesky
- QR gives better numerical stability than Cholesky

Linear Least Squares Problem

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||Ax - b||^2$$

- $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$
- The objective function is convex!

$$\nabla^2 g(x) = 2A^T A \succeq 0$$

- Recall: x is a global minima $\Leftrightarrow \nabla g(x) = 0$ and $\nabla^2 g(x) \succeq 0$
- $\nabla g(x) = A^T Ax - A^T b$
- x is a global minima $\Leftrightarrow$ $\boxed{A^T Ax = A^T b}$

Done!!

Back to Nonlinear Least Squares Problem

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x)||^2 = \sum_{i=1}^{m} |r_i(x)|^2$$

- $r : \mathbb{R}^n \to \mathbb{R}^m$ and $r(x) = [r_1(x), r_2(x), \ldots r_m(x)]^T$
- $r_i(x)$ is the residual function
- Linear least square if $r(x) = Ax - b$. Solved!!

Back to Nonlinear Least Squares Problem

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x)||^2 = \sum_{i=1}^{m} |r_i(x)|^2$$

- $r : \mathbb{R}^n \to \mathbb{R}^m$ and $r(x) = [r_1(x), r_2(x), \ldots r_m(x)]^T$
- $r_i(x)$ is the residual function
- Linear least square if $r(x) = Ax - b$. Solved!!

What if we linearize $r(x)$ and solve it as a linear least square?

Linear Approximations

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x)||^2$$

- $r : \mathbb{R}^n \to \mathbb{R}^m$ and $r(x) = [r_1(x), r_2(x), \ldots r_m(x)]^T$

- First-order Taylor approximation

$$r_i(x) \approx r_i(x_0) + \nabla r_i(x_0)^T (x - x_0) \quad \text{for every } i = 1, 2, \ldots m$$

compile them to get

$$r(x) \approx r(x_0) + J(x_0)(x - x_0) \qquad \text{where} \qquad J(x_0) = \begin{pmatrix} \nabla r_1(x_0)^T \\ \nabla r_2(x_0)^T \\ \vdots \\ \nabla r_m(x_0)^T \end{pmatrix}$$

Holds for any $x_0 \in \mathbb{R}^n$

Nonlinear Least Squares Problem

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x)||^2$$

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x_0) + J(x_0)(x - x_0)||^2$$

for any $x_0 \in \mathbb{R}^n$

- $r : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $r(x) = [r_1(x), r_2(x), \ldots r_m(x)]^T$
- $r_i(x)$ is the residual function

Nonlinear Least Squares Problem

substitute $d = (x - x_0)$

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x)||^2$$

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x_0) + J(x_0)(x - x_0)||^2$$

for any $x_0 \in \mathbb{R}^n$

- $r : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $r(x) = [r_1(x), r_2(x), \ldots r_m(x)]^T$
- $r_i(x)$ is the residual function

Nonlinear Least Squares Problem

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x)||^2$$

$$\underset{d \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x_0) + J(x_0)d||^2$$

for any $x_0 \in \mathbb{R}^n$

Get solution d^*

- $r : \mathbb{R}^n \to \mathbb{R}^m$ and $r(x) = [r_1(x), r_2(x), \ldots r_m(x)]^T$
- $r_i(x)$ is the residual function

Solution will be $x = x_0 + d^*$

Nonlinear Least Squares Problem

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x)||^2$$

$$\underset{d \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x_0) + J(x_0)d||^2$$

for any $x_0 \in \mathbb{R}^n$

Get solution d^*

- $r : \mathbb{R}^n \to \mathbb{R}^m$ and $r(x) = [r_1(x), r_2(x), \ldots r_m(x)]^T$
- $r_i(x)$ is the residual function

Solution will be $x = x_0 + d^*$ Will it? Yes or No?

Nonlinear Least Squares Problem

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x)||^2$$

$$\underset{d \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x_0) + J(x_0)d||^2$$

for any $x_0 \in \mathbb{R}^n$

Get solution d^*

- $r : \mathbb{R}^n \to \mathbb{R}^m$ and $r(x) = [r_1(x), r_2(x), \ldots r_m(x)]^T$
- $r_i(x)$ is the residual function

~~Solution will be $x = x_0 + d^*$~~ No!!

Nonlinear Least Squares Problem

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x)||^2$$

$$\underset{d \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x_t) + J(x_t)d||^2$$

Get solution d_t^*

- $r : \mathbb{R}^n \to \mathbb{R}^m$ and $r(x) = [r_1(x), r_2(x), \ldots r_m(x)]^T$
- $r_i(x)$ is the residual function
- Iterate over $x_{t+1} = x_t + d_t^*$

Nonlinear Least Squares Problem

Linear least square

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x)||^2$$

$$\underset{d \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x_t) + J(x_t)d||^2$$

Get solution d_t^*

- $r : \mathbb{R}^n \to \mathbb{R}^m$ and $r(x) = [r_1(x), r_2(x), \ldots r_m(x)]^T$
- $r_i(x)$ is the residual function
- Iterate over $x_{t+1} = x_t + d_t^*$

Nonlinear Least Squares Problem

Linear least square

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x)||^2$$

$$\underset{d \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x_t) + J(x_t)d||^2$$

Get solution d_t^*

- $r : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $r(x) = [r_1(x), r_2(x), \ldots r_m(x)]^T$
- $r_i(x)$ is the residual function
- Iterate over $x_{t+1} = x_t + d_t^*$

This is called the Gauss-Newton Method

Gauss-Newton Method

Minimize $||r(x)||^2$
$x \in \mathbb{R}^n$

1. Start with an innitial guess x_0

For t $= 0, 1, 2, \ldots$ until convergence

Gauss-Newton Method

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x)||^2$$

1. Start with an innitial guess x_0

For t = 0, 1, 2, ... until convergence

2. Linearize the residual function $r(x)$ at x_t

$$r(x_t + d) \approx r(x_t) + J(x_t)d$$

Gauss-Newton Method

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x)||^2$$

1. Start with an innitial guess x_0

For t = 0, 1, 2, . . . until convergence

2. Linearize the residual function $r(x)$ at x_t

$$r(x_t + d) \approx r(x_t) + J(x_t)d$$

3. Solve the linear least squares problem to obtain the minimum d_t

$$\underset{d \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x_t) + J(x_t)d||^2 \quad \xrightarrow[b = -r(x_t)]{A = J(x_t)} \quad J(x_t)^T J(x_t)d = -J(x_t)^T r(x_t)$$

Gauss-Newton Method

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x)||^2$$

1. Start with an innitial guess x_0

For t = 0, 1, 2, . . . until convergence

2. Linearize the residual function $r(x)$ at x_t

$$r(x_t + d) \approx r(x_t) + J(x_t)d$$

3. Solve the linear least squares problem to obtain the minimum d_t

$$\underset{d \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x_t) + J(x_t)d||^2 \quad \xrightarrow[b = -r(x_t)]{A = J(x_t)} \quad J(x_t)^T J(x_t)d = -J(x_t)^T r(x_t)$$

4. Update $x_{t+1} = x_t + d_t$

Gauss-Newton Method

Minimize $||r(x)||^2$
$x \in \mathbb{R}^n$

1. Start with an innitial guess x_0

For t = 0, 1, 2, . . . until convergence

2. Linearize the residual function $r(x)$ at x_t

$$r(x_t + d) \approx r(x_t) + J(x_t)d$$

3. Solve the linear least squares problem to obtain the minimum d_t

Minimize $||r(x_t) + J(x_t)d||^2$
$d \in \mathbb{R}^n$

$$\xrightarrow[b = -r(x_t)]{A = J(x_t)} \quad J(x_t)^T J(x_t)d = -J(x_t)^T r(x_t)$$

4. Update $x_{t+1} = x_t + \alpha_t d_t$

Nonlinear Least Squares Problem

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad ||r(x)||^2 = \sum_{i=1}^{m} |r_i(x)|^2$$

- $r : \mathbb{R}^n \to \mathbb{R}^m$ and $r(x) = [r_1(x), r_2(x), \ldots r_m(x)]^T$
- $r_i(x)$ is the residual function
- Gauss-Newton Method
- Local convergence. Cannot ensure global convergence.

Summary

- Nonlinear least squares problem

$$\text{Minimize} \quad ||r(x)||^2$$

- Linear least squares problem
  - Gradient descent
  - Cholesky solver
  - QR solver

$$\text{Minimize} \quad ||Ax - b||^2 \qquad (A^T A)x = A^T b$$

$$A^T A = LL^T$$

$$A^T A = QR$$

- Gauss-Newton Method

$$J(x_t)^T J(x_t)d = -J(x_t)^T r(x_t) \qquad x_{t+1} = x_t + \alpha_t d_t$$

A quick detour

- Nonlinear optimization
- Convexity
- Optimality conditions
- Gradient descent

Summary

- Nonlinear least squares problem
- Linear least squares problem
  - Gradient descent
  - Cholesky solver
  - QR solver
- Gauss-Newton Method

$$\text{Minimize} \quad ||r(x)||^2$$

$$\text{Minimize} \quad ||Ax - b||^2 \qquad (A^T A)x = A^T b$$

$$A^T A = LL^T$$

$$A^T A = QR$$

$$J(x_t)^T J(x_t) d = -J(x_t)^T r(x_t) \qquad x_{t+1} = x_t + \alpha_t d_t$$

A quick detour

- Nonlinear optimization
- Convexity
- Optimality conditions
- Gradient descent

Next

- Issues with Gauss-Newton Method
- Levenberg-Marquardt Method
- Nonlinear least squares on Riemannian Manifolds