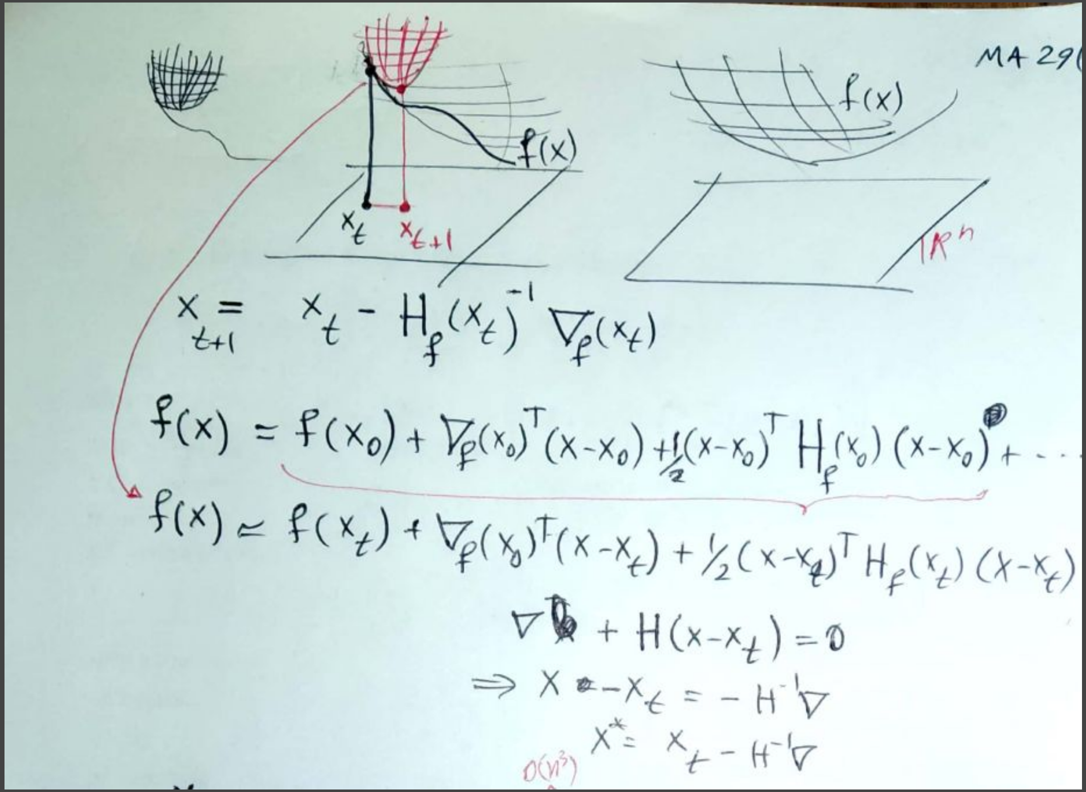
Mathematics for AI

Lecture 29

Quasi-Newton methods, Nonlinear Least Squares, Constrained Optimization, Equality Constraints, Lagrange Multipliers

Quadratic Approximation - Taylor series perspective



Newton's method - Computation Issue

$$x^* = x_t - H^{-1}\nabla$$

$O(n^3)$

Newton $\longrightarrow$ $x_{t+1} = x_t - H_f(x_t)^{-1} \nabla_f(x_t)$

$x \in \mathbb{R}^n$

$f: \mathbb{R}^n \rightarrow \mathbb{R}$

Gradient descent $\longrightarrow$ $x_{t+1} = x_t - \alpha \nabla_f(x_t)$ $O(n)$

Quasi-Newton methods

K. N. Toosi
University of Technology

Quasi-Newton methods

شبه نیوتن

- ~~Estimate~~ ~~Apporix~~ Approximate H using past gradients/steps
- Approximate H^{-1} directly.

Nonlinear least squares

$$\cancel{f(\theta)}\quad C(\theta) = \sum_{i=1}^{n} r_i(\theta)^2 \qquad \text{nonlinear least squares}$$

$$C(\theta) = \begin{bmatrix} r_1(\theta) & r_2(\theta) & \cdots & r_n(\theta) \end{bmatrix} \begin{bmatrix} r_1(\theta) \\ r_2(\theta) \\ \vdots \\ r_n(\theta) \end{bmatrix} = r(\theta)^T r(\theta)$$

Gauss-Newton method

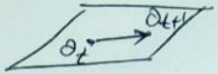
K. N. Toosi
University of Technology

$$C(\theta) = \sum_{r=1}^{n} r_i(\theta)^2 = \vec{r}(\theta)^T \vec{r}(\theta) = \| \vec{r}(\theta) \|^2$$

MA 29

$\theta \in \mathbb{R}^p$ $r(\theta) \in \mathbb{R}^n$ $r : \mathbb{R}^p \to \mathbb{R}^n$

$$\theta_t \longrightarrow \theta_{t+1}$$



Gauss-~~Neth~~ Newton method:

$$\vec{r}(\theta) \simeq \vec{r}(\theta_t) + J_r(\theta_t)(\theta - \theta_t)$$

$$\theta_{t+1} = \underset{\theta}{\operatorname{argmin}} \| r(\theta_t) + J(\theta - \theta_t) \|^2$$

$$= \underset{\theta}{\operatorname{argmin}} \| J\theta - (\underbrace{J\theta_t - r(\theta_t)}_{b}) \|^2$$

$$= (J^T J)^{-1} J^T (J\theta_t - r(\theta_t))$$

$$(J^T J)\theta = J^T (J\theta_t - r(\theta_t))$$

Levenberg-Marquardt

Levenberg-Marquardt

$$\left(J^TJ + \lambda I\right)\theta = J^T\left(J\theta_t - r(\theta_t)\right)$$

Constrained Optimization

Constrained optimization

$$\min_{x \in \mathbb{R}^n} (Ax-b)^T(Ax-b)$$

$$\min_{x \in C} f(x)$$

Constrained Optimization

$$\min_{x \in C} f(x)$$

$$\min_{x} f(x) \quad \text{subject to} \quad x \in C$$

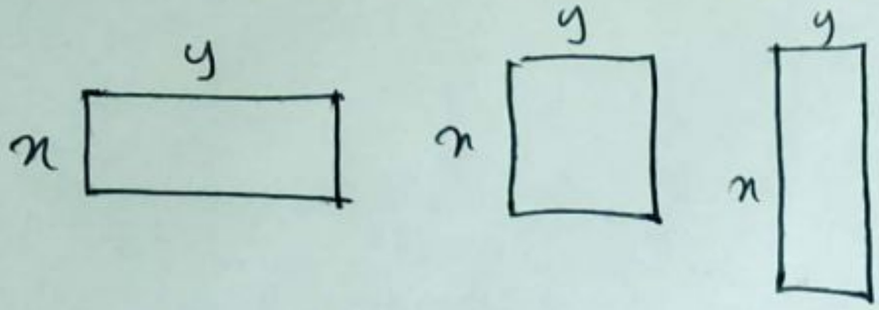
$f : \mathbb{R}^2 \to \mathbb{R}$

$$\min f(x) \quad \text{subject to} \quad \underbrace{\|x\| \leq 1}_{\text{قید}}$$

$$\min f(x) \quad \text{subject to} \quad \|x\| = 1$$

Example

Among the rectangle with circumference MA29 = 2 which one has the largest area?

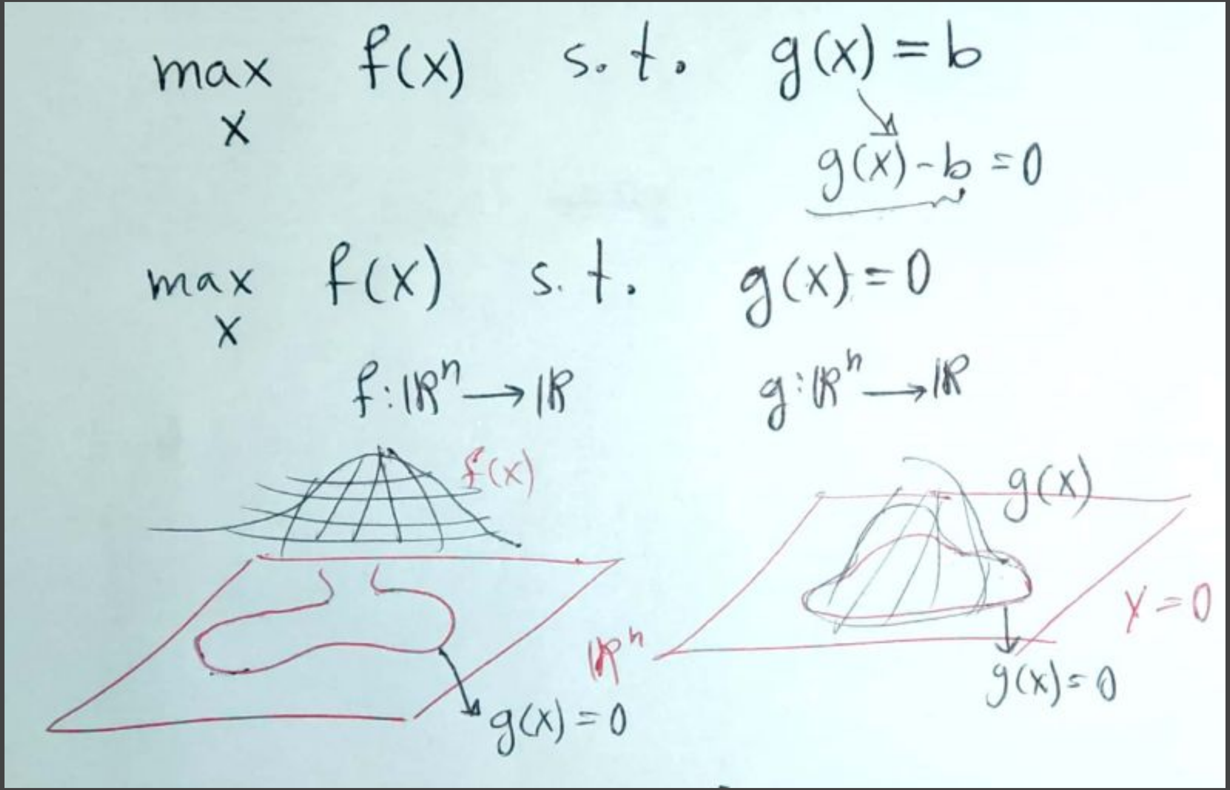


$$\boxed{x+y=2}$$

$$x^*, y^* = \underset{x,y}{\operatorname{argmax}} \; xy \quad \text{subject to} \quad x+y=2$$

Equality Constraints

K. N. Toosi
University of Technology

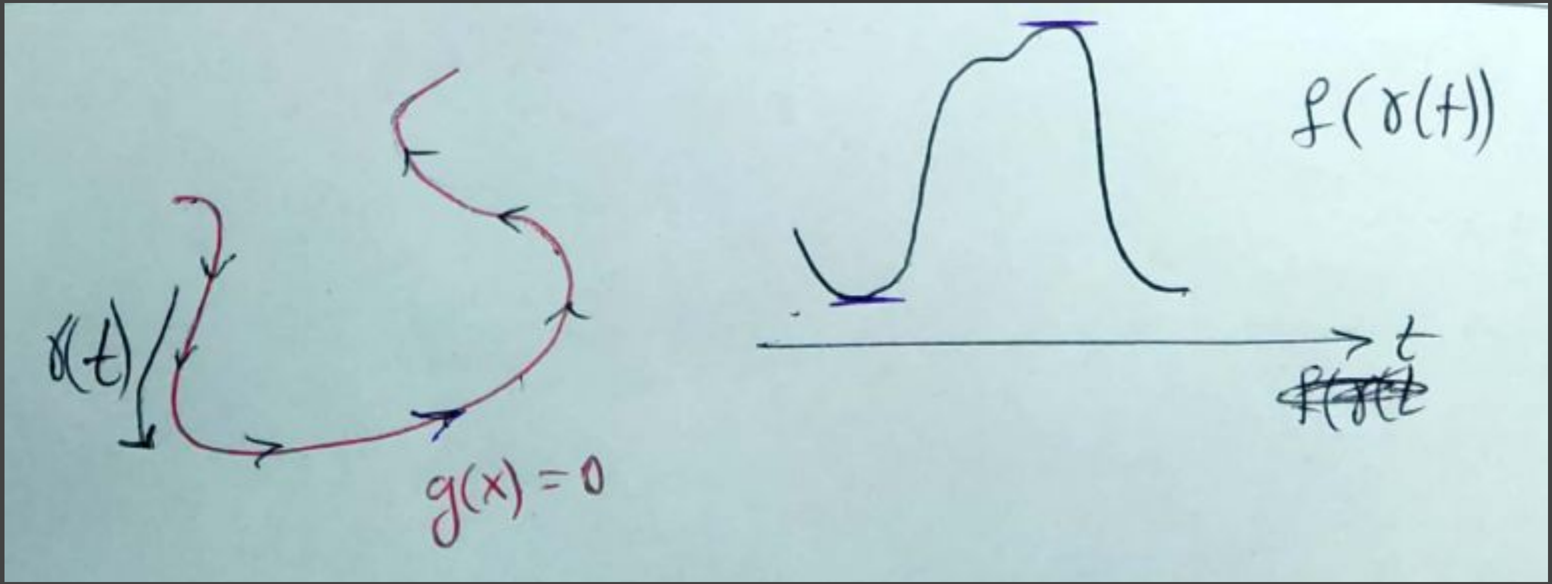


Bad way to do it!

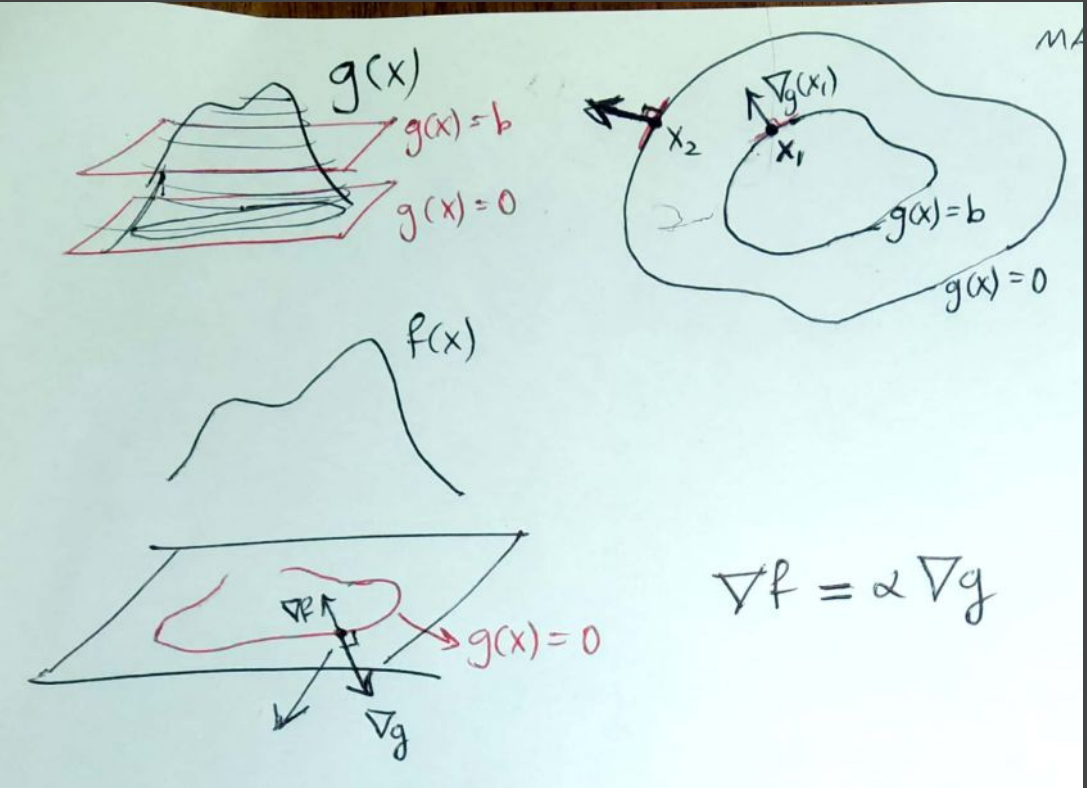
$$\max_{x} \; f(x) - \gamma \, g(x)^2 \qquad \gamma \to \infty$$

Optimizing on a level curve

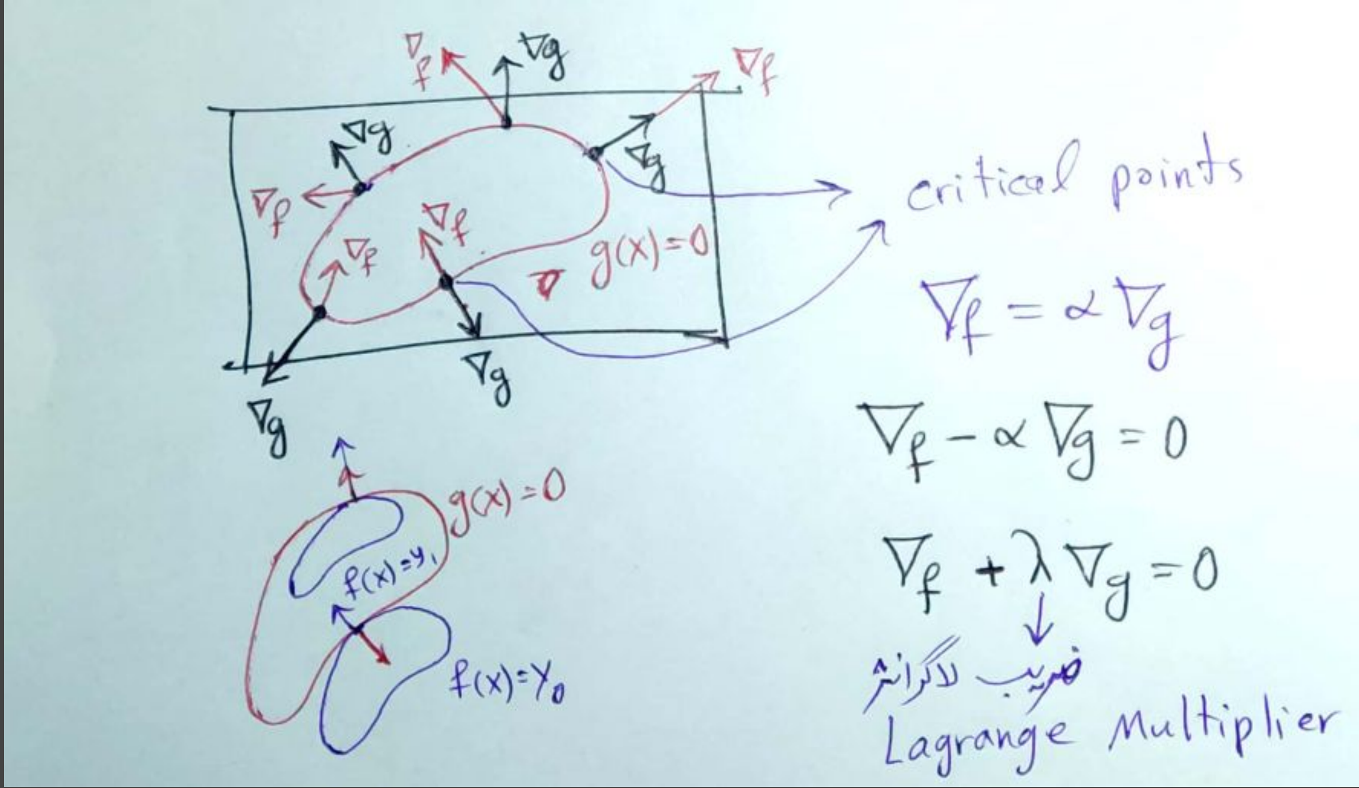
K. N. Toosi
University of Technology



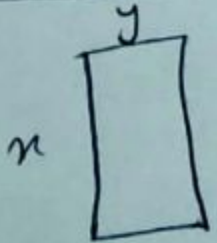
Optimizing on a level curve



Lagrange Multiplier



Example



$$\max_{x,y} \ \cancel{x+y}\ xy \quad \text{s.t.} \quad \underline{x+y=1}$$

$$f(x) = xy \qquad g(x) = x+y-1$$

$$\nabla f + \lambda \nabla g = 0 \Rightarrow \begin{cases} \begin{bmatrix} y \\ x \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0 \Rightarrow x = y = -\lambda \\ x+y=1 \xrightarrow{\text{ادامه}} x = y = 0.5 \end{cases}$$

→ دو معادله، سه مجهول

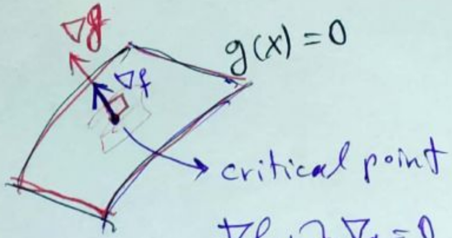
سه معادله - سه مجهول

ادامه ⇒

ND case: Optimize on level hyper-surfaces

$\cancel{\nabla f + \lambda \nabla g = 0}$

$g, f : \mathbb{R}^n \longrightarrow \mathbb{R}$



$\nabla f + \lambda \nabla g = 0$

find x, λ

$$\nabla f(x) + \lambda \nabla g(x) = 0$$

$$g(x) = 0$$

find $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$ such that

$$\left.\begin{array}{l} \nabla f(x) + \lambda \nabla g(x) = 0 \; \} \; n \text{ equations} \\ g(x) = 0 \; \} \; 1 \text{ equation} \end{array}\right\} n+1 \text{ equations}$$

$n+1$ unknowns $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$

$x_1, x_2 - x_n, \lambda$

Example

$$\max_{x} \; x^T A x \quad \text{s.t.} \quad \|x\| = 1 \qquad x \in \mathbb{R}^n$$

A: symmetric

$$\max_{x} \; x^T A x \quad \text{s.t.} \quad x^T x = 1$$

$$A^T = A$$

$$f(x) = x^T A x \qquad g(x) = x^T x - 1$$

Example

$$\max_{x} \; x^T A x \quad \text{s.t.} \quad x^T x = 1$$

A: symmetric

$A^T = A$

$$f(x) = x^T A x \qquad g(x) = x^T x - 1$$

$$\nabla f(x) + \lambda \nabla g(x) = 0 \Rightarrow 2Ax + \lambda(2x) = 0$$

$$Ax + \lambda x = 0 \Rightarrow Ax = (-\lambda)x \Rightarrow$$ x is an eigenvector of A

corr. eigenvalues $-\lambda$

* all Eigenvectors of A are critical points.

(v_i, λ'_i) is an eigenpair $\quad f(v_i) = v_i^T A v_i = v_i^T(\lambda'_i v_i)$

$\|v_i\| = 1 \qquad\qquad = \lambda'_i v_i^T v_i = \lambda'_i$

The Lagrangian

K. N. Toosi
University of Technology

$$\min_{x} \; f(x) \qquad \text{s.t. } \; g(x) = 0$$

$$\max_{x} \; f(x) \qquad \text{s.t. } \; g(x) = 0$$

$$\nabla f(x) + \lambda \nabla g(x) = 0$$

$$\nabla_{x} \Big(\underbrace{f(x) + \lambda g(x)}_{\text{Lagragian}} \Big) = 0$$

w.r.t. x

$$\mathcal{L}(x, \lambda) = f(x) + \lambda g(x) \qquad \text{Lagrangian}$$

$$\begin{aligned} \frac{\partial}{\partial x} \mathcal{L}(x, \lambda) = 0 \;&\Rightarrow \\ \frac{\partial}{\partial \lambda} \mathcal{L}(x, \lambda) = 0 \;&\Rightarrow \end{aligned} \left\{ \begin{aligned} &\nabla f(x) + \lambda \nabla g(x) = 0 \\ &g(x) = 0 \end{aligned} \right.$$