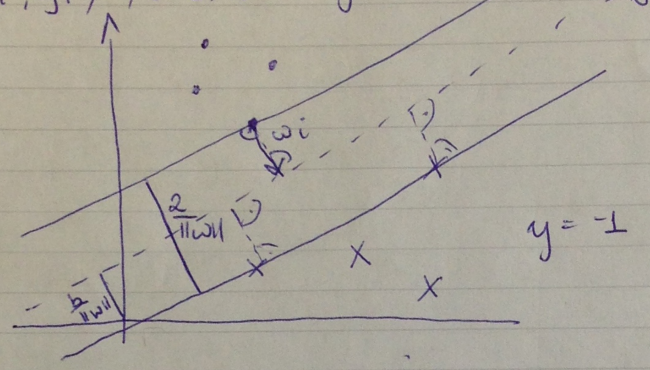
Lecture 1 — Numerical Optimization

Example:

Support vector machine:

Given training data (x_i, y_i), $i = 1..n$: $y = 1$



Looking for maximum margin hyperplanes (*), (**)

Hyperplanes: $wx_i - b = \pm 1$

$$\max_w \frac{2}{\|w\|}$$

s.t. $wx_i - b \geq 1$ if $y_i = 1$

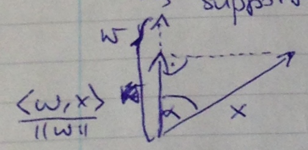
$wx_i - b \leq -1$ if $y_i = -1$

$$\min_w \|w\| \quad s.t. \quad y_i(wx_i - b) \geq 1 \quad i = 1, .., n$$

Classifier: $x \to \operatorname{sgn}(wx - b)$

The max-margin hyperplane is fully determined by the points closest to it (on both sides)

→ support vectors.

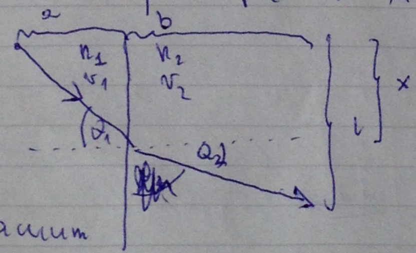


$$\langle w, x\rangle = \|w\| \cdot \|x\| \cdot \cos\alpha$$

Example:

Fermat's principle: light travels along the fastest path.

Snell's law



$v_1 = \frac{c}{n_1}$ $\quad v_2 = \frac{c}{n_2}$

$$\min T = \frac{\sqrt{x^2 + a^2}}{v_1} + \frac{\sqrt{b^2 + (l-x)^2}}{v_2}$$

$$\frac{dT}{dx} = 0$$

c – speed of light in vacuum

Lecture 1 Numerical Optimisation

$$\frac{dT}{dx} = \frac{\overbrace{x}^{\sin\theta_1}}{v_1\sqrt{x^2+a^2}} + \frac{\overbrace{-(l-x)}^{-\sin\theta_2}}{v_2\sqrt{(l-x)^2+b^2}} = 0$$

$$\Rightarrow \frac{\sin\theta_1}{v_1} = \frac{\sin\theta_2}{v_2}$$

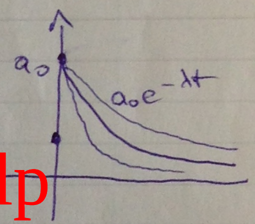
$$n_1 \sin\theta_1 = n_2 \sin\theta_2$$

Example Data fitting

$f(\lambda) = \| a_0 e^{-\lambda t_i} - y_i \|_2^2$ exponential decay

$\frac{da}{dt} = -\lambda a \rightsquigarrow a(t) = a_0 e^{-\lambda t}$

$a_0 = a(t=0)$



Example: Calculus of variations

$$J[y] = \int_{x_1}^{x_2} L(x, y(x), y'(x))\, dx$$

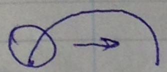
Euler-Lagrange equation

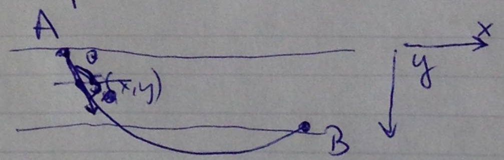
$$\frac{\partial L}{\partial y} - \frac{d}{dx}\frac{\partial L}{\partial y'} = 0$$

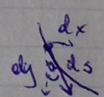
Example Brachistochrone curve

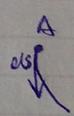
curve of fastest descent between A & B

Solution: inverse cycloid







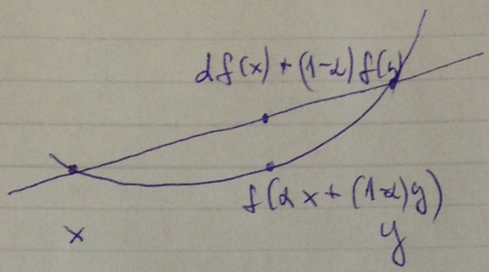


En. conserv.: $m v^2 = m \cdot g \cdot y \Rightarrow v = \sqrt{2gy}$ [Johann Bernoulli]

$\frac{\sin\theta}{v} = \frac{1}{v_m} = \text{const} = \frac{dx}{v\,ds}$ [shortest path]

$v_m^2 dx^2 = v^2 ds^2 = v^2(dx^2 + dy^2) \Rightarrow dx = \frac{v\,dy}{\sqrt{v_m^2 - v^2}} = \sqrt{\frac{y}{y_m - y}}\, dy$

Convex function:



Taylor theorem 1D

$f : \mathbb{R} \to \mathbb{R}$, k times differentiable at a

then $\exists h_k : \mathbb{R} \to \mathbb{R}$:

$$f(x) = \overbrace{f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \dots + \frac{f^{(k)}(a)}{k!}(x-a)^k}^{P_k(x)}$$

$$+ h_k(x)(x-a)^k$$

and $\lim_{x \to a} h_k(x) = 0$ [Peano reminder]

The error (reminder)

$$R_k(x) = f(x) - P_k(x) = o(|x-a|^k), \quad x \to a$$

If $f : \mathbb{R} \to \mathbb{R}$, $k+1$ differentiable on (x, a) and $f^{(k)}$ continuous on $[x, a]$

Lagrange: $R_k(x) = \frac{f^{(k+1)}(\xi_L)}{(k+1)!}(x-a)^{k+1}$ for some $\xi_L \in (a, x)$

weaker than

Cauchy: $R_k(x) = \frac{f^{(k+1)}(\xi_C)}{k!}(x-\xi_C)^k (x-a)$ for some $\xi_C \in (a, x)$

Integral form: $f^{(k)}$ absolutely continuous on $[a, x]$ $[\Rightarrow f^{(k+1)} \in L^1]$

$$R_k(x) = \int_a^x \frac{f^{(k+1)}(t)}{k!}(x-t)^k \, dt$$

stronger assumption: $f^{(k+1)}$ continuous on $[a, x]$

Taylor theorem: ND

1D Taylor thm. in θ

$$f(x + \theta p) = f(x) + \theta \nabla f(x)^T p + \frac{1}{2}\theta^2 p^T \nabla^2 f(x + \bar{\theta} p) p$$

$$\bar{\theta} \in (0, \theta)$$

Lagrange

Setting $\theta = 1$ we recover the ND theorem.

$$R_1(x) = \int_0^\theta p^T \frac{\nabla^2 f(x + tp)}{1} p \cdot (\theta - t)\, dt$$