03/01/2018

$$\min \|Ax-y\|_2^2 + \lambda \|x\|_1$$

Hard questions, yet to be answered:

- is the argmin any good, i.e., close to x_0 if $y = Ax_0 + e$, and x_0 is sparse?
- how to min. in practice

Today: duality, variants (wavelets, TV).

$$\min \underbrace{\|x\|}_{\text{some norm}} \text{ s.t. } Ax = b$$

$$\mathcal{L}(x, v) = \|x\| + v^T(Ax - b)$$
$$= -b^T v + \underbrace{(A^T v)^T x + \|x\|}_{\text{inf over } x.}$$

<u>Def.</u> (dual norm)

$$\|y\|_* = \sup_x \{ y^T x : \|x\| \le 1 \}$$

note $|y^T x| \le \|x\| \; \|y\|_*$

<u>Ex</u> $\sup_x \{ y^T x : \|x\|_2 \le 1 \} = \|y\|_2$

because $\underbrace{|y^T x| \le \|y\|_2}_{\text{Cauchy-Schwarz}}$ when $\|x\|_2 = 1$

equality when $x = y / \|y\|_2$.

<u>Ex.</u> $\sup_x \{ y^T x : \|x\|_\infty \le 1 \} = \|y\|_1$

because $\sum y_i x_i$ is max when $x_i = y_i / |y_i|$

<u>Ex.</u> $\sup_x \{ y^T x : \|x\|_1 \le 1 \} = \|y\|_\infty$

because $\sum y_i x_i$ is max when $x_i = \begin{cases} \mathrm{sgn}(y_i) : i = \operatorname{argmax} |y_i| \\ 0 \text{ otherwise} \end{cases}$

<u>Def</u> (convex conjugate) $f^*(y) = \sup_x \{ y^T x - f(x) \}$

<u>Prop.</u> The conjugate function of $\|x\|$ is

$$f^*(y) = \begin{cases} 0 & \text{if } \|y\|_* \le 1 \\ \infty & \text{otherwise} \end{cases}$$
$$= \chi_{\|y\|_* \le 1}(y) \qquad \text{(indicator)}$$

<u>Pf.</u> $f^*(y) = \sup_x \{ y^T x - \|x\| \}$.

- if $\|y\|_* \le 1$, then $y^T x \le \|x\| \, \|y\|_* \le \|x\|$

  $y^T x - \|x\| \le 0$, choose $x = 0$

  $\Rightarrow f^*(y) = 0$

- if $\|y\|_* > 1$, then $\exists z$, $\|z\| = 1$, $y^T z > 1$

  Let $x = tz$, $t > 1$. Then

  $y^T x - \|x\| = t(y^T z - \|z\|) \to \infty$ as $t \to \infty$

  $\Rightarrow f^*(y) = +\infty$ $\square$.

<u>Back to ex.</u> $\mathcal{L}(x, \nu) = -b^T \nu + (A^T \nu)^T x + \|x\|$

$= -b^T \nu - ((-A^T \nu)^T x - \|x\|)$

$g(\nu) = \inf_x \mathcal{L}(x, \nu) = -b^T \nu - f^*(-A^T \nu)$

$$= \begin{cases} -b^T \nu & \text{if } \|A^T \nu\|_* \le 1 \\ \infty & \text{otherwise} \end{cases}$$

Dual: $$\boxed{\begin{aligned} &\max \ -b^T \nu \\ &\text{s.t. } \|A^T \nu\|_* \le 1. \end{aligned}}$$

Wavelet - sparse regularization (Mallat, Meyer 1986) (Donoho 1992)

(a) Orthonormal basis $\lambda = (j,k)$ integers

$$\Psi_{i\lambda} = 2^{j/2}\psi(2^j t_i - k) \qquad t_i = \frac{i}{N}$$

⚠ watch edge effects

$$x = \Psi c \quad , \quad c = \Psi^T x$$

$$\min_c \| A \Psi c - y \|_2^2 + \lambda \| c \|_1$$

(b) Frame ("redundant basis")

$$x = \Psi c, \quad c = \Psi^+ x \quad \text{with} \quad \Psi\Psi^+ = I$$

pseudo-inverse $\Psi^T(\Psi\Psi^T)^{-1}$ $\qquad \Psi^+\Psi = P_{\mathrm{Ran}\,\Psi^+}$

Synthesis prior / formulation:

$$\min_c \| A \Psi c - y \|_2^2 + \lambda \| c \|_1$$

Analysis prior / formulation:

$$\min_x \| A x - y \|_2^2 + \lambda \| \Psi^+ x \|_1$$

(they are different because the solution c of "synthesis" is not in general in $\mathrm{Ran}\,\Psi^+$)

Tight frame: $\Psi^+ = c\,\Psi^T$ for some c

$$\iff \Psi\Psi^T = \frac{1}{c} I$$

ex.  in $\mathbb{R}^2$

(TV) Total variation regularization. (Rudin-Osher-Fatemi 1990)

1D, discrete:

$$\min_x \|Ax - y\|_2^2 + \lambda \|Bx\|_1$$

$$\text{with } B = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & & \ddots & \\ & & & -1 & 1 \end{bmatrix}$$

nD, continuous:

$$\min_f \left\| \int K(x,y) f(y)\, dy - g(x) \right\|_2^2 + \lambda \|\nabla f\|_1$$

Distributional notion: "mass" of ∇f as a measure,

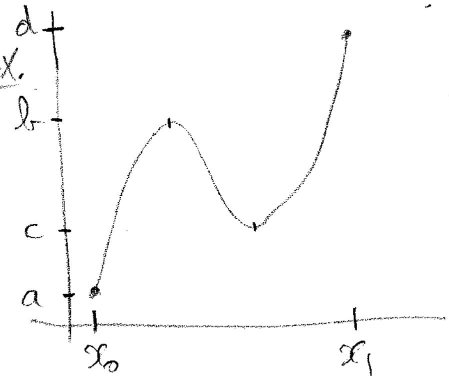
or $\mathrm{Var}(f, \Omega) = \|f\|_V =$

$$\sup\left\{ \int_\Omega f(x)\, \nabla \cdot \phi(x)\, dx : \|\phi\|_\infty \leq 1 \right\}$$

$$BV(\mathbb{R}^n) = \{ f \in L^1(\mathbb{R}^n) : \mathrm{Var}(f, \mathbb{R}^n) < \infty \}$$

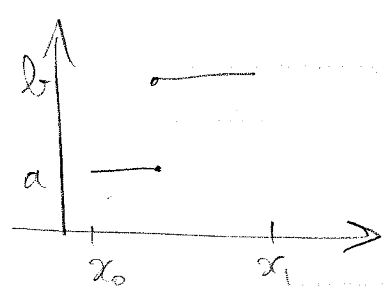
↳ bounded variation ↳ total variation

ex.



$$\mathrm{Var}(f, [x_0, x_1]) = \int_{x_0}^{x_1} |f'(x)|\, dx$$
$$= |b-a| + |b-c| + |d-c|$$
$$> |d-a|$$

ex.



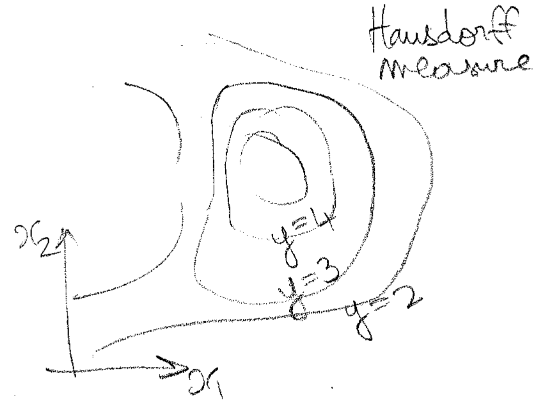
$$\mathrm{Var}(f, [x_0, x_1]) = |b-a|$$

ex. (Co-area formula)

$$\mathrm{Var}(f,\Omega) = \int_{-\infty}^{\infty} H^1(\Gamma_y)\, dy$$

Hausdorff measure

level set at level y.



$$\Rightarrow \mathrm{Var}(\chi_S, \mathbb{R}^2) = \mathrm{length}(\partial S)$$

$$\mathrm{Var}(\chi_S, \mathbb{R}^3) = \mathrm{area}(\partial S).$$

Point of TV: preserves edges, unlike $\int |\nabla f|^2$.