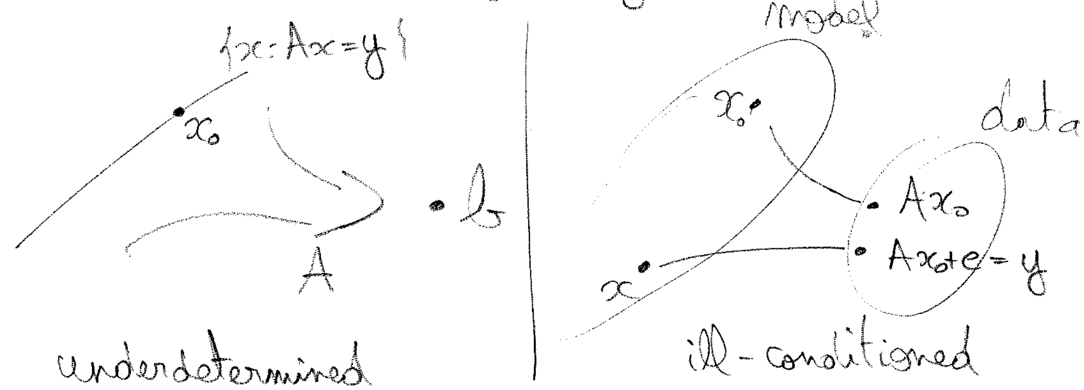
02/27/18 Sparse regularization



In both cases, it pays to impose more structure on x.
Prominent example: x is <u>k-sparse</u>
(x has k nonzero components)

<u>Def</u>. ℓ_0 quasi-norm:

$$\|x\|_{\ell_0} = |\operatorname{supp} x| = \underbrace{|\{i,\ 1 \le i \le n,\ x_i \ne 0\}|}_{\text{cardinality}}$$

<u>Def</u>. ℓ_p norm,
- if $0 < p < \infty$, $\|x\|_p = \left(\sum_i |x_i|^p\right)^{1/p}$
- if $p = \infty$, $\|x\|_\infty = \max_i |x_i|$

<u>Prop</u>. $\|x\|_{\ell_0} = \lim_{p \to 0} \|x\|_p^p$.

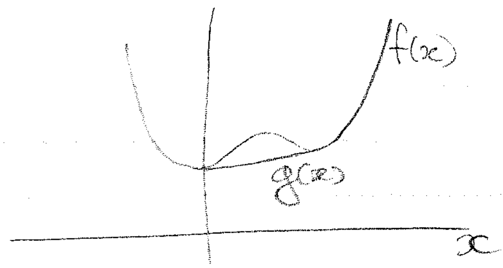
Sparse regularization (ideal, ℓ_0):

$$\min \|Ax - y\|^2 \text{ s.t. } \|x\|_{\ell_0} \le k \quad \text{for some } k$$

Computationally hard, no known polynomial-time algorithm in general ("NP hard")
$\to$ search over all $\binom{n}{k}$ possible supports for x.

Need a convex replacement / relaxation

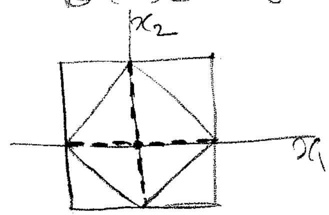
Def. Convex envelope of a function $f(x)$:
longest convex $g(x)$ such that $g(x) \leq f(x)$



Ex. $x \in \mathbb{R}^2$: $\quad \|x\|_{\ell_0} = \begin{cases} 0 & x = 0 \\ 1 & \text{on axes} \\ 2 & \text{elsewhere} \end{cases}$

envelope $(\|x\|_{\ell_0}) = 0$.
(homogeneous of order 0 : $\|\alpha x\|_{\ell_0} = \|x\|_{\ell_0}, \quad \alpha \neq 0$)

Ex. Find the envelope of $\|x\|_{\ell_0}$
for x restricted to $\|x\|_\infty \leq 1$.



$g(x)$ convex and
$g(x) \leq 1$ on axes $\cap$ square
(dihedral skeleton ----)
$\rightarrow g(x) \leq 1$ in diamond
$\rightarrow g(x) = \|x\|_1 = \sum |x_i|$
(or $\|x\|_1 / \alpha$ if $\|x\|_\infty \leq \alpha$)

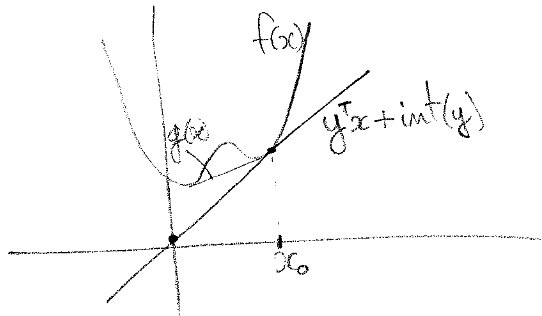
Thm. Convex envelope of $\|x\|_{\ell_0}$,
for x restricted to $\|x\|_\infty \leq 1$,
is $\|x\|_1$

Relax to
$$\begin{cases} \min \|Ax-y\|^2 \ \text{s.t.}\ \|x\|_1 \le \tau & (P_\tau) \\ \min \|Ax-y\|^2 + \lambda \|x\|_1 & (L_\lambda) \\ \min \|x\|_1 \ \text{s.t.}\ \|Ax-y\| \le \varepsilon. & (Q_\varepsilon) \end{cases}$$

(Donoho, early 1990s).

Called
- LASSO (least absolute shrinkage and selection operator), Tibshirani 1996
- basis pursuit denoising (basis pursuit when $Ax=y$ i.e. $\varepsilon=0$)

How to find the convex envelope?



$$g(x) = \sup_y \left(y^T x + \overset{\text{intercept}}{\mathrm{int}(y)}\right)$$

where

$\mathrm{int}(y)$ = largest y such that $y^T x + \mathrm{int}(y) \le f(x), \forall x$

$\iff \mathrm{int}(y) \le f(x) - y^T x$

$\rightarrow \mathrm{int}(y) = \inf_x \left(f(x) - y^T x\right).$

Make symmetric: $\mathrm{int}(y) = -f^*(y)$

<u>Def.</u> (convex conjugaté, Legendre-Fenchel trf.)

$$f^*(y) = \sup_x \left(y^T x - f(x)\right)$$

($) <u>Prop</u>. $g(x) = \sup_y \left(y^T x - f^*(y)\right) = f^{**}(x)$

is the convex envelope of $f(x)$.

<u>Pf.</u> of theorem: take two conjugates in succession

$$f^*(y) = \sup_{\|x\|_\infty \le 1} \left(y^T x - \|x\|_0\right)$$

$$\left(\text{here } f(x) = \|x\|_0 + \chi_{\|x\|_\infty \le 1}(x)\right) \qquad \chi \rightarrow \begin{cases} 0 & \text{if } \|x\|_\infty \le 1 \\ \infty & \text{otherwise} \end{cases}$$

Denote $y_{(i)}$ the components of y sorted in decreasing order of absolute value. $|y_{(i)}| \geq |y_{(i+1)}|$

For the time being, restrict $\|x\|_{\ell_0} = k$.
Then pick $x_i = \begin{cases} \mathrm{sgn}(y_i) & \text{if } |y_i| \text{ is one of the } k \text{ largest comp. of } y. \\ 0 & \text{otherwise} \end{cases}$

$$\Rightarrow y^T x - \|x\|_{\ell_0} = \sum_{i=1}^{k} |y_{(i)}| - k = \sum_{i=1}^{k} (|y_{(i)}| - 1)$$

Maximum over all values $0 \leq k \leq n$:

$$f^*(y) = \max\left\{ \sum_{i=1}^{k} (|y_{(i)}| - 1) : 0 \leq k \leq n \right\}$$

$$= \sum_{i=1}^{n} (|y_{(i)}| - 1)_+ \quad \text{(sum all the positive ones)}$$

$+ \to$ positive part.

$$= \sum_{i=1}^{n} (|y_i| - 1)_+$$

Then $f^{**}(x) = \sup_y \; y^T x - f^*(y)$

(a) if $\exists i : |x_i| > 1$, then pick $y = \alpha \delta_i$ for that value of i, and $\alpha \to \infty$

$$y^T x - f^*(y) = \alpha |x_i| + (\alpha - 1)_+ \to \infty.$$

(b) now assume $\|x\|_\infty \leq 1$.
Let k such that $|y_{(i)}| > 1$ for $1 \leq i \leq k$ (could be $k = n$).

$$y^T x - \sum_{i=1}^{k} (|y_{(i)}| - 1)$$

$$\leq \sum^{n} |y_i||x_i| - \sum |x_i| - \sum^{k} (|y_{(i)}| - 1) + \|x\|_1$$

$$= \sum^{n} (|y_{(i)}| - 1)|x_{(i)}| - \sum^{k} (|y_{(i)}| - 1) + \|x\|_1$$

$\hookrightarrow$ same component of x as $y_{(i)}$ is of y:

$$= \sum_{i=1}^{k} \underbrace{(|y_{(i)}| - 1)}_{>0} \underbrace{(|x_{(i)}| - 1)}_{\leq 1} + \sum_{i=k+1}^{n} \underbrace{(|y_{(i)}| - 1)}_{\leq 0} \underbrace{|x_{(i)}|}_{\geq 0} + \|x\|_1$$

$\leq \|x\|_1$

and equality is achieved when $k=0$ and $y_i = \mathrm{sgn}(x_i)$

$$\Rightarrow f^{**}(x) = \begin{cases} \|x\|_1 & \text{if } \|x\|_\infty \leq 1 \\ \infty & \text{otherwise} \end{cases}$$

$$= \|x\|_1 + \chi_{\|x\|_\infty \leq 1}(x).$$

$\square$.

ex. Image compression / restoration with wavelets

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

(wavelet orthobasis

$$\min_c \|A\Psi c - y\|^2 + \lambda \|c\|_1 \qquad (A = I, \text{ or a blur})$$

$\Longleftrightarrow$

$$\min_x \|Ax - y\|^2 + \lambda \|\Psi^T x\|_1$$

ex. Total variation regularization (image proc., tomography)

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

where B is a gradient operator

(Rudin, Osher, Fatemi 1990)