- Interpret L_1 norm heuristic as convex relaxation

$$\min.\ \mathrm{card}(x)$$
subject to. $x \in C \quad \|x\|_\infty \le R$

$\Downarrow$ equivalent

$$\left.\begin{aligned} &\min.\ \mathbb{1}^T z \\ &\text{s.t.}\quad |x_i| \le R z_i \\ &\qquad x \in C \quad z_i \in \{0,1\} \end{aligned}\right\}$$

boolean convex

convex except the boolean constraint

$\Downarrow$ convex relaxation

$$\begin{aligned} &\min.\ \mathbb{1}^T z \\ &\text{s.t.}\quad |x_i| \le R z_i \\ &\qquad x \in C \quad 0 \le z_i \le 1 \end{aligned}$$

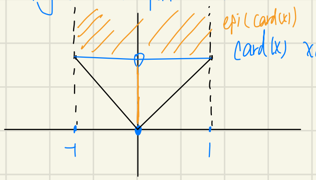
$\Downarrow$?

$$\begin{aligned} &\min.\ (1/R)\,\|x\|_1 \\ &\text{s.t.}\quad x \in C \\ &\qquad \|x\|_\infty \le R \end{aligned}$$

optimal value of this proplem is a lower bound of the original problem

- Interpretation via convex envelop

Convex envelop f^{env} of a function of a function f on a convex set C is the largest convex function that is an underestimator of f on C



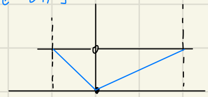
$\mathrm{epi}(f^{env}) = \mathrm{ConvexHull}(\mathrm{epi}(f))$

$f^{env} = (f^*)^*$ (with some technical conditions)

for scalar x, $|x|$ is the convex envelope of $\mathrm{card}(x)$ on $[-1,1]$

for $x \in \mathbb{R}^n$, $1/R\ \|x\|_1$ is the convex envelope of $\mathrm{card}(x)$ on $\{z \mid \|z\|_\infty \le R\}$

eg. $x \in [-1,2]$



convex envelop of $\mathrm{card}(x)$ for $x \in [-1,2]$

A much better relaxation than $\|x\|_1$.

works much better in theory and practice

- Weighted and asymmetric l_1 heuristic

minimize card(x) over convex set C

suppose we know a bounding box (min. / max. x_i over C $\Rightarrow$ solve 2n convex problems)

$x \in C \Rightarrow l_i \le x_i \le r_i$

if $u_i < 0$ or $l_i > 0$, then card(x) = 1 for all $x \in C$

assuming $l_i < 0$ and $r_i > 0$:

$$\sum_{i}^{n} \left(\frac{(x_i)_+}{u_i} + \frac{(x_i)_-}{-l_i} \right)$$

- Regression selector

min. $\|Ax-b\|_2$

subject to card(x) $\le k$

heuristic:

minimize $\|Ax-b\|_2 + \gamma\|x\|_1$

find smallest γ that gives card(x) $\le k$

fixed the sparsity pattern (i.e. subset of features) and find x that min. $\|Ax-b\|_2$

- Sparse signal reconstruction

convex cardinality problem:

min. $\|Ax-y\|_2$

s.t. card(x) $\le k$

l_1 heuristic:

min. $\|Ax-y\|_2$

s.t. $\|x\|_1 \le \beta$

lasso

another form:

min. $\|Ax-y\|_2 + \gamma\|x\|_1$

(basis pursuit denoising)

- Total variation reconstruction

fit x_{cor} with piecewise constant $\bar{x}$, no more than k jumps

convex cardinality problem:

min. $\|x_{cor} - x\|_2$

s.t. card(Dx) $\le k$ D is the first-order difference matrix

heuristic: min. $\|\bar{x} - x_{cor}\|_2 + \gamma \underbrace{\|D\bar{x}\|_1}_{\text{total variation}}$ vary γ to adjust number of jumps

if γ is too large, $\bar{x}$ will be a constant $\bar{x}_i = \text{mean}(x_{cor})$

unlike l_2-based reconstruction, total-variation reconstruction filters high frequency noise out while preserving sharp jumps

· Total variation image reconstruction

$x \in \mathbb{R}^n$ are values of pixels on $N \times N$ grid (e.g. $N=31$, $n=961$)

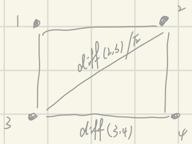
assumption: x has relatively few changes in value (like cartoon with boundary )

have $m = 120$ linear measure measurements $y = Fx$, $F_{ij} \sim \mathcal{N}(0,1)$

$$\min. \sum_{i,j} |x_{i,j} - x_{i+1,j}| + \sum |x_{i,j} - x_{i,j+1}|$$

$$\text{s.t. } y = Fx$$

in image:



⇒ multipoint approximation of gradient

minimize the sum of 2-norm of those of gradients

(approximately rotation invariant)

· Iterated weight L_1 heuristic

$\min.\ \mathbf{card}(x)$ s.t. $x \in C$

$w := 1$

repeat {

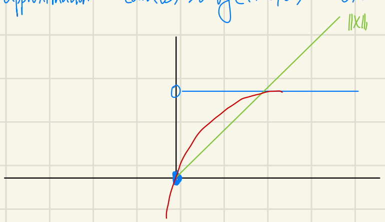
$\min.\ \|\mathrm{diag}(w)\, x\|$ s.t. $x \in C$

$w_i := 1/(\epsilon + |x_i|)$ — if $x_i = 0$, give x_i the biggest weight to make it stay 0

}

can take $x \geq 0$ by writing $x = x_+ - x_-$, $x_+ \geq 0$, $x_- \geq 0$ (pos and neg part)

use approximation $\mathbf{card}(z) \approx \log(1 + z/\epsilon)$, $\epsilon > 0$, $z \in \mathbb{R}_+$



the nonconvex problem:

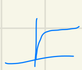
$$\min. \sum_i^n \log(1 + x_i/\epsilon)$$

$$\text{s.t. } x_i \in C, \quad x \geq 0$$

sequential convex programming applied to this problem gives iterated weighted l_1 heuristic

find a local solution by linearizing objective at current point (sequential programming)

$$\sum_i^n \log(1 + x_i/\epsilon) \approx \sum_i^n \log(1 + x_i^{(k)}/\epsilon) + \frac{x_i - x_i^{(k)}}{\epsilon + x_i^{(k)}}$$

other methods: $\min x^p$ $(p<1)$ 

card(x) are sometimes called l-0 norm

eg. $\min \sum_i |x_i|^{0.1}$ which gives very sparse solution

eg. Detecting changes in time series model

scalar time-series model

$$y(t+2) = a(t)\, y(t+1) + b(t)\, y(t) + v(t) \qquad v(t) \sim N(0, \sigma^2) \text{ iid.}$$

assumption: $a(t)$ $b(t)$ piece constant, change infrequently

given $y(t)$ $t=1, \cdots T$ estimate $a(t)$ $b(t)$ $t=1, \cdots T-2$

$$\min \ \sum_{t=1}^{T-2} \| y(t+2) - a(t)\, y(t) - b(t)\, y(t) \|_2^2$$
$$+ \gamma \sum_{t=1}^{T-2} \left(|a(t+1) - a(t)| + |b(t+1) - b(t)| \right)$$

· Extension to matracies

rank is the natural analog for matracies

convex-rank problem: convex, except for Rank in objective or constraints

analog of l_1 heuristic: nuclear norm: $\|X\|_* = \sum_i \sigma_i(X)$

(sum of singular values; dual of spectral norm (max singular value))

for $X \succeq 0$, $\sum_i \sigma_i(X) = \mathrm{Tr}(X)$

· Factor modeling

given matrix $\Sigma \in S_+^n$, find approximation $\hat{\Sigma} = FF^T + D$ (diagonal-plus-low-rank) $F \in \mathbb{R}^{n \times r}$, D is nonnegative diagonal

eg. Σ is a covariance matrix

if $D = 0 \Rightarrow$ eigenvalue decomposition

underlying model:

$X = Fz + v \quad v \sim N(0, D) \quad z \sim N(0, I)$

model with fewest factor.

$$\min_{\Sigma, D} \ \mathrm{Rank}(\Sigma)$$
$$\text{s.t.} \ \Sigma \succeq 0 \qquad D \succeq 0 \text{ diagonal}$$
$$\Sigma + D \in C$$

$\Downarrow$ trace heuristic

$$\min_{\Sigma, D} \ \mathrm{tr}(\Sigma)$$
$$\text{s.t.} \ \Sigma \succeq 0 \qquad D \succeq 0 \text{ diagonal}$$
$$\Sigma + D \in C$$

eg. $x = Fz + v$ $\quad z \sim N(0, I)$ $\quad v \sim N(0, D)$ $\quad D$ diagnal $\quad F \in \mathbb{R}^{20 \times 3}$

Σ is emperical covariance matrix from $N = 3000$ samples

set of acceptable approximations

$$C = \{ \hat{\Sigma} \mid \| \Sigma^{-1/2} (\hat{\Sigma} - \Sigma) \Sigma^{-1/2} \| \le \beta \}$$

trace heuristic

$$\min \quad \mathrm{Tr}(X)$$

$$\text{s.t.} \quad X \succeq 0 \quad d \ge 0$$

$$\| \Sigma^{-1/2} (X + \mathrm{diag}(d) - \Sigma) \Sigma^{-1/2} \| \le \beta$$