ℓ_1-Norm Methods for Convex-Cardinality Problems

EE364b, Stanford University

Outline

- problems involving cardinality
- the ℓ_1-norm heuristic
- convex relaxation and convex envelope interpretations
- examples
- recent results
- total variation
- iterated weighted ℓ_1 heuristic
- matrix rank constraints

ℓ_1-norm heuristics for cardinality problems

- cardinality problems arise often, but are hard to solve exactly
- a simple heuristic, that relies on ℓ_1-norm, seems to work well
- used for many years, in many fields
  - sparse design
  - LASSO, robust estimation in statistics
  - support vector machine (SVM) in machine learning
  - total variation reconstruction in signal processing, geophysics
  - compressed sensing
- recent theoretical results guarantee the method works, at least for a few problems

Cardinality

- the **cardinality** of $x \in \mathbf{R}^n$, denoted $\mathbf{card}(x)$, is the number of nonzero components of x
- $\mathbf{card}$ is separable; for scalar x, $\mathbf{card}(x) = \begin{cases} 0 & x = 0 \\ 1 & x \neq 0 \end{cases}$
- $\mathbf{card}$ is quasiconcave on $\mathbf{R}^n_+$ (but not $\mathbf{R}^n$) since

$$\mathbf{card}(x + y) \geq \min\{\mathbf{card}(x), \mathbf{card}(y)\}$$

  holds for $x, y \succeq 0$
- but otherwise has no convexity properties
- arises in many problems

General convex-cardinality problems

a **convex-cardinality problem** is one that would be convex, except for appearance of $\mathbf{card}$ in objective or constraints

examples (with $\mathcal{C}$, f convex):

- convex minimum cardinality problem:

$$\begin{array}{ll} \text{minimize} & \mathbf{card}(x) \\ \text{subject to} & x \in \mathcal{C} \end{array}$$

- convex problem with cardinality constraint:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathcal{C}, \quad \mathbf{card}(x) \leq k \end{array}$$

Solving convex-cardinality problems

convex-cardinality problem with $x \in \mathbf{R}^n$

- if we fix the sparsity pattern of x ($i.e.$, which entries are zero/nonzero) we get a convex problem
- by solving 2^n convex problems associated with all possible sparsity patterns, we can solve convex-cardinality problem
(possibly practical for $n \leq 10$; not practical for $n > 15$ or so . . .)
- general convex-cardinality problem is (NP-) hard
- can solve globally by branch-and-bound
  - can work for particular problem instances (with some luck)
  - in worst case reduces to checking all (or many of) 2^n sparsity patterns

Boolean LP as convex-cardinality problem

- Boolean LP:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b, \quad x_i \in \{0,1\} \end{array}$$

  includes many famous (hard) problems, *e.g.*, 3-SAT, traveling salesman

- can be expressed as

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b, \quad \mathbf{card}(x) + \mathbf{card}(1-x) \leq n \end{array}$$

  since $\mathbf{card}(x) + \mathbf{card}(1-x) \leq n \iff x_i \in \{0,1\}$

- conclusion: general convex-cardinality problem is hard

Sparse design

$$\begin{array}{ll} \text{minimize} & \mathbf{card}(x) \\ \text{subject to} & x \in \mathcal{C} \end{array}$$

- find sparsest design vector x that satisfies a set of specifications
- zero values of x simplify design, or correspond to components that aren't even needed
- examples:
  - FIR filter design (zero coefficients reduce required hardware)
  - antenna array beamforming (zero coefficients correspond to unneeded antenna elements)
  - truss design (zero coefficients correspond to bars that are not needed)
  - wire sizing (zero coefficients correspond to wires that are not needed)

Sparse modeling / regressor selection

fit vector $b \in \mathbf{R}^m$ as a linear combination of k regressors (chosen from n possible regressors)

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|_2 \\ \text{subject to} & \mathbf{card}(x) \leq k \end{array}$$

- gives k-term model
- chooses subset of k regressors that (together) best fit or explain b
- can solve (in principle) by trying all $\binom{n}{k}$ choices
- variations:
  - minimize $\mathbf{card}(x)$ subject to $\|Ax - b\|_2 \leq \epsilon$
  - minimize $\|Ax - b\|_2 + \lambda\, \mathbf{card}(x)$

Sparse signal reconstruction

- estimate signal x, given
  - noisy measurement $y = Ax + v$, $v \sim \mathcal{N}(0, \sigma^2 I)$ (A is known; v is not)
  - prior information $\mathbf{card}(x) \leq k$

- maximum likelihood estimate $\hat{x}_{\rm ml}$ is solution of

$$\begin{array}{ll} \text{minimize} & \|Ax - y\|_2 \\ \text{subject to} & \mathbf{card}(x) \leq k \end{array}$$

Estimation with outliers

- we have measurements $y_i = a_i^T x + v_i + w_i$, $i = 1, \ldots, m$
- noises $v_i \sim \mathcal{N}(0, \sigma^2)$ are independent
- only assumption on w is sparsity: $\mathbf{card}(w) \leq k$
- $\mathcal{B} = \{i \mid w_i \neq 0\}$ is set of bad measurements or *outliers*
- maximum likelihood estimate of x found by solving

$$\begin{array}{ll} \text{minimize} & \sum_{i \notin \mathcal{B}} (y_i - a_i^T x)^2 \\ \text{subject to} & |\mathcal{B}| \leq k \end{array}$$

  with variables x and $\mathcal{B} \subseteq \{1, \ldots, m\}$

- equivalent to

$$\begin{array}{ll} \text{minimize} & \|y - Ax - w\|_2^2 \\ \text{subject to} & \mathbf{card}(w) \leq k \end{array}$$

Minimum number of violations

- set of convex inequalities

$$f_1(x) \leq 0, \; \ldots, \; f_m(x) \leq 0, \qquad x \in \mathcal{C}$$

- choose x to minimize the number of violated inequalities:

$$\begin{array}{ll} \text{minimize} & \mathbf{card}(t) \\ \text{subject to} & f_i(x) \leq t_i, \quad i = 1, \ldots, m \\ & x \in \mathcal{C}, \quad t \geq 0 \end{array}$$

- determining whether zero inequalities can be violated is (easy) convex feasibility problem

Linear classifier with fewest errors

- given data $(x_1, y_1), \ldots, (x_m, y_m) \in \mathbf{R}^n \times \{-1, 1\}$
- we seek linear (affine) classifier $y \approx \mathbf{sign}(w^T x + v)$
- classification error corresponds to $y_i(w^T x + v) \leq 0$
- to find w, v that give fewest classification errors:

$$\begin{array}{ll} \text{minimize} & \mathbf{card}(t) \\ \text{subject to} & y_i(w^T x_i + v) + t_i \geq 1, \quad i = 1, \ldots, m \end{array}$$

with variables w, v, t (we use homogeneity in w, v here)

Smallest set of mutually infeasible inequalities

- given a set of mutually infeasible convex inequalities $f_1(x) \leq 0, \ldots, f_m(x) \leq 0$

- find smallest (cardinality) subset of these that is infeasible

- certificate of infeasibility is $g(\lambda) = \inf_x(\sum_{i=1}^m \lambda_i f_i(x)) \geq 1$, $\lambda \succeq 0$

- to find smallest cardinality infeasible subset, we solve

$$\begin{array}{ll} \text{minimize} & \mathbf{card}(\lambda) \\ \text{subject to} & g(\lambda) \geq 1, \quad \lambda \succeq 0 \end{array}$$

(assuming some constraint qualifications)

Portfolio investment with linear and fixed costs

- we use budget B to purchase (dollar) amount $x_i \geq 0$ of stock i
- trading fee is fixed cost plus linear cost: $\beta\,\mathbf{card}(x) + \alpha^T x$
- budget constraint is $\mathbf{1}^T x + \beta\,\mathbf{card}(x) + \alpha^T x \leq B$
- mean return on investment is $\mu^T x$; variance is $x^T \Sigma x$
- minimize investment variance (risk) with mean return $\geq R_{\min}$:

$$
\begin{array}{ll}
\text{minimize} & x^T \Sigma x \\
\text{subject to} & \mu^T x \geq R_{\min}, \quad x \succeq 0 \\
& \mathbf{1}^T x + \beta\,\mathbf{card}(x) + \alpha^T x \leq B
\end{array}
$$

Piecewise constant fitting

- fit corrupted x_{cor} by a piecewise constant signal $\hat{x}$ with k or fewer jumps
- problem is convex once location (indices) of jumps are fixed
- $\hat{x}$ is piecewise constant with $\leq k$ jumps $\Longleftrightarrow \mathbf{card}(D\hat{x}) \leq k$, where

$$D = \begin{bmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \end{bmatrix} \in \mathbf{R}^{(n-1)\times n}$$

- as convex-cardinality problem:

$$\begin{array}{ll} \text{minimize} & \|\hat{x} - x_{\mathrm{cor}}\|_2 \\ \text{subject to} & \mathbf{card}(D\hat{x}) \leq k \end{array}$$

Piecewise linear fitting

- fit $x_{\rm cor}$ by a piecewise linear signal $\hat{x}$ with k or fewer kinks
- as convex-cardinality problem:

$$
\begin{array}{ll}
\mbox{minimize} & \|\hat{x} - x_{\rm cor}\|_2 \\
\mbox{subject to} & \mathbf{card}(\nabla^2 \hat{x}) \leq k
\end{array}
$$

where

$$
\nabla^2 = \left[\begin{array}{ccccccc}
-1 & 2 & -1 & & & & \\
& -1 & 2 & -1 & & & \\
& & \ddots & \ddots & \ddots & & \\
& & & -1 & 2 & -1 &
\end{array}\right]
$$

ℓ_1-norm heuristic

- replace $\mathbf{card}(z)$ with $\gamma\|z\|_1$, or add regularization term $\gamma\|z\|_1$ to objective

- $\gamma > 0$ is parameter used to achieve desired sparsity (when $\mathbf{card}$ appears in constraint, or as term in objective)

- more sophisticated versions use $\sum_i w_i|z_i|$ or $\sum_i w_i(z_i)_+ + \sum_i v_i(z_i)_-$, where w, v are positive weights

Example: Minimum cardinality problem

- start with (hard) minimum cardinality problem

$$\begin{array}{ll} \text{minimize} & \mathbf{card}(x) \\ \text{subject to} & x \in \mathcal{C} \end{array}$$

  ($\mathcal{C}$ convex)

- apply heuristic to get (easy) ℓ_1-norm minimization problem

$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & x \in \mathcal{C} \end{array}$$

Example: Cardinality constrained problem

- start with (hard) cardinality constrained problem (f, $\mathcal{C}$ convex)

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathcal{C}, \quad \mathbf{card}(x) \leq k \end{array}$$

- apply heuristic to get (easy) ℓ_1-constrained problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathcal{C}, \quad \|x\|_1 \leq \beta \end{array}$$

  or ℓ_1-regularized problem

$$\begin{array}{ll} \text{minimize} & f(x) + \gamma\|x\|_1 \\ \text{subject to} & x \in \mathcal{C} \end{array}$$

  β, γ adjusted so that $\mathbf{card}(x) \leq k$

Polishing

- use ℓ_1 heuristic to find $\hat{x}$ with required sparsity
- fix the sparsity pattern of $\hat{x}$
- re-solve the (convex) optimization problem with this sparsity pattern to obtain final (heuristic) solution

Interpretation as convex relaxation

- start with

$$\begin{array}{ll} \text{minimize} & \mathbf{card}(x) \\ \text{subject to} & x \in \mathcal{C}, \quad \|x\|_\infty \leq R \end{array}$$

- equivalent to mixed Boolean convex problem

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T z \\ \text{subject to} & |x_i| \leq R z_i, \quad i = 1, \ldots, n \\ & x \in \mathcal{C}, \quad z_i \in \{0, 1\}, \quad i = 1, \ldots, n \end{array}$$

with variables x, z

- now relax $z_i \in \{0,1\}$ to $z_i \in [0,1]$ to obtain

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T z \\ \text{subject to} & |x_i| \leq R z_i, \quad i = 1, \ldots, n \\ & x \in \mathcal{C} \\ & 0 \leq z_i \leq 1, \quad i = 1, \ldots, n \end{array}$$

  which is equivalent to

$$\begin{array}{ll} \text{minimize} & (1/R)\|x\|_1 \\ \text{subject to} & x \in \mathcal{C} \end{array}$$

  the ℓ_1 heuristic

- optimal value of this problem is lower bound on original problem

Interpretation via convex envelope

- convex envelope $f^{\rm env}$ of a function f on set $\mathcal{C}$ is the largest convex function that is an underestimator of f on $\mathcal{C}$
- $\mathbf{epi}(f^{\rm env}) = \mathbf{Co}(\mathbf{epi}(f))$
- $f^{\rm env} = (f^*)^*$ (with some technical conditions)
- for x scalar, $|x|$ is the convex envelope of $\mathbf{card}(x)$ on $[-1, 1]$
- for $x \in \mathbf{R}^n$, $(1/R)\|x\|_1$ is convex envelope of $\mathbf{card}(x)$ on $\{z \mid \|z\|_\infty \leq R\}$

Weighted and asymmetric ℓ_1 heuristics

- minimize $\mathbf{card}(x)$ over convex set $\mathcal{C}$
- suppose we know lower and upper bounds on x_i over $\mathcal{C}$

$$x \in \mathcal{C} \implies l_i \leq x_i \leq u_i$$

(best values for these can be found by solving $2n$ convex problems)

- if $u_i < 0$ or $l_i > 0$, then $\mathbf{card}(x_i) = 1$ ($i.e.$, $x_i \neq 0$) for all $x \in \mathcal{C}$
- assuming $l_i < 0$, $u_i > 0$, convex relaxation and convex envelope interpretations suggest using

$$\sum_{i=1}^{n} \left(\frac{(x_i)_+}{u_i} + \frac{(x_i)_-}{-l_i} \right)$$

as surrogate (and also lower bound) for $\mathbf{card}(x)$

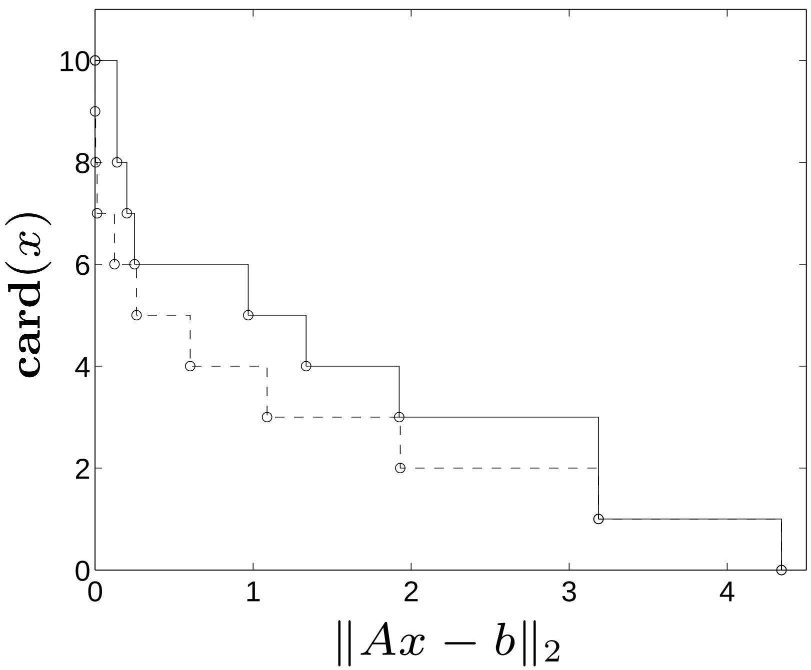
Regressor selection

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|_2 \\ \text{subject to} & \mathbf{card}(x) \leq k \end{array}$$

- heuristic:
  - minimize $\|Ax - b\|_2 + \gamma\|x\|_1$
  - find smallest value of γ that gives $\mathbf{card}(x) \leq k$
  - fix associated sparsity pattern ($i.e.$, subset of selected regressors) and find x that minimizes $\|Ax - b\|_2$

Example (6.4 in BV book)

- $A \in \mathbf{R}^{10\times 20}$, $x \in \mathbf{R}^{20}$, $b \in \mathbf{R}^{10}$
- dashed curve: exact optimal (via enumeration)
- solid curve: ℓ_1 heuristic with polishing



Sparse signal reconstruction

- convex-cardinality problem:

$$\begin{array}{ll} \text{minimize} & \|Ax - y\|_2 \\ \text{subject to} & \mathbf{card}(x) \leq k \end{array}$$

- ℓ_1 heuristic:

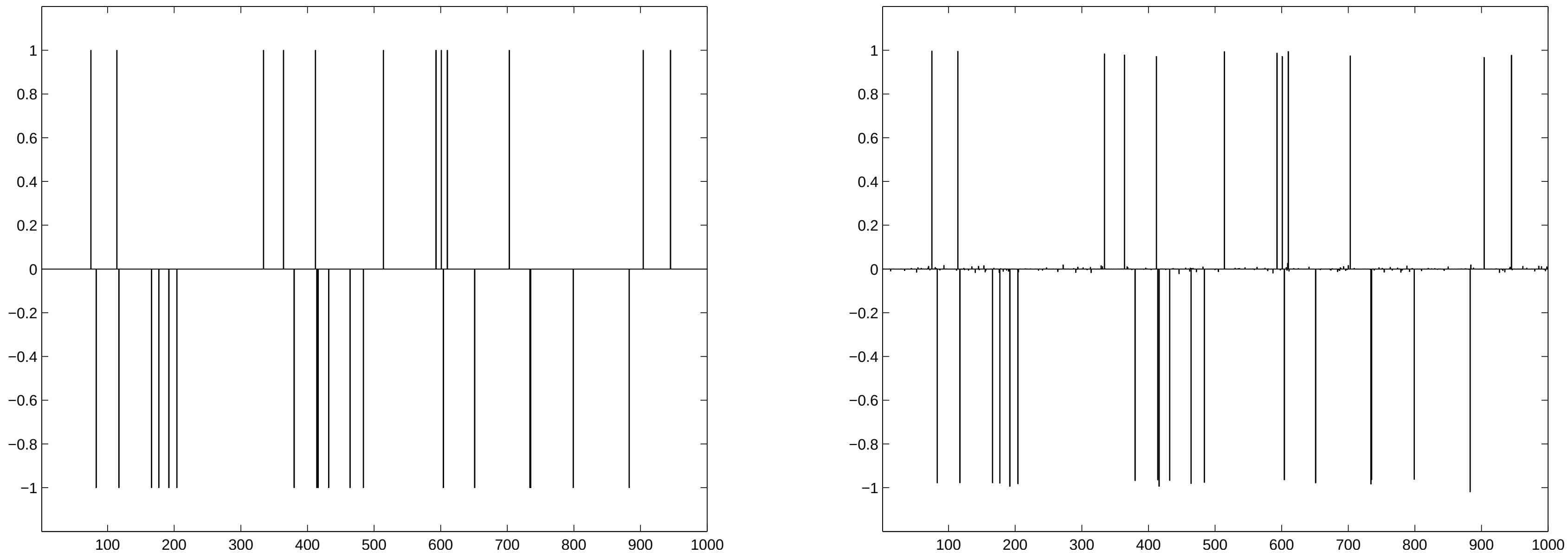
$$\begin{array}{ll} \text{minimize} & \|Ax - y\|_2 \\ \text{subject to} & \|x\|_1 \leq \beta \end{array}$$

  (called LASSO)

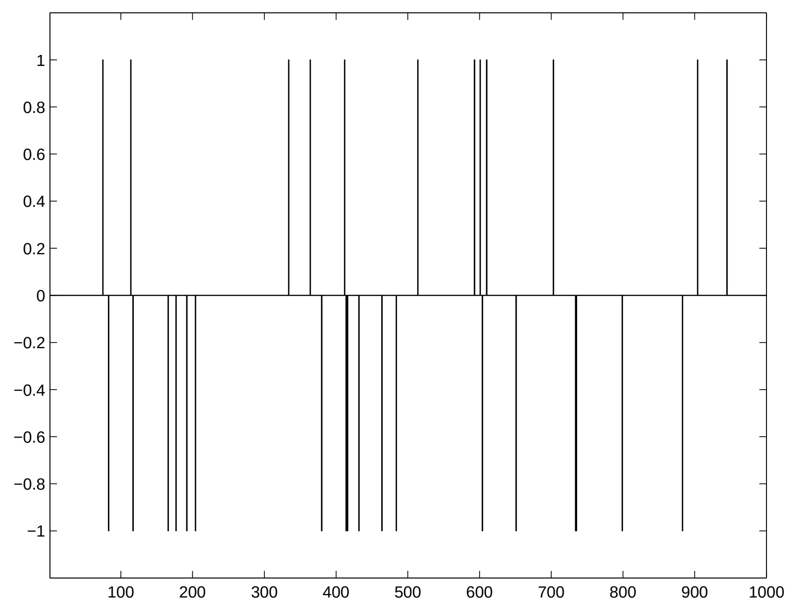
- another form: minimize $\|Ax - y\|_2 + \gamma\|x\|_1$
  (called basis pursuit denoising)

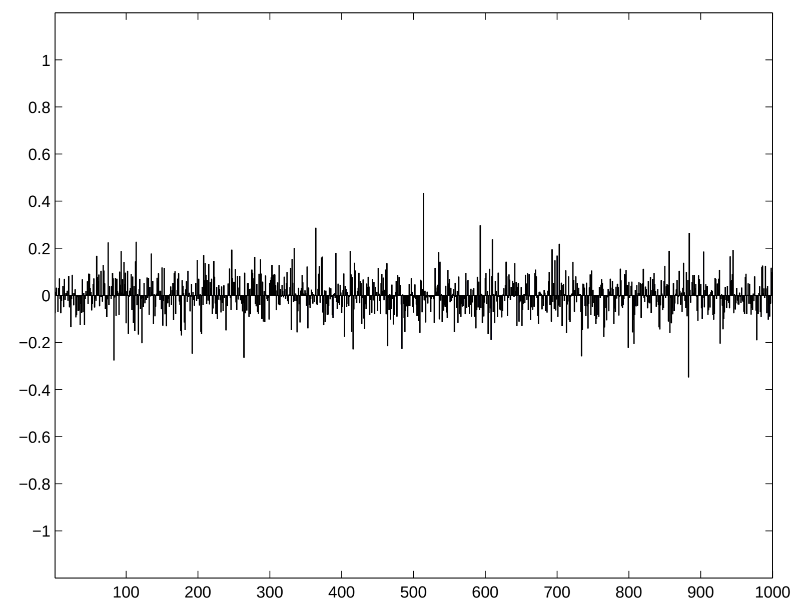
Example

- signal $x \in \mathbf{R}^n$ with $n = 1000$, $\mathbf{card}(x) = 30$
- $m = 200$ (random) noisy measurements: $y = Ax + v$, $v \sim \mathcal{N}(0, \sigma^2 I)$, $A_{ij} \sim \mathcal{N}(0, 1)$
- *left*: original; *right*: ℓ_1 reconstruction with $\gamma = 10^{-3}$



- ℓ_2 reconstruction; minimizes $\|Ax - y\|_2 + \gamma\|x\|_2$, where $\gamma = 10^{-3}$
- *left*: original; *right*: ℓ_2 reconstruction





Some recent theoretical results

- suppose $y = Ax$, $A \in \mathbf{R}^{m \times n}$, $\mathbf{card}(x) \leq k$
- to reconstruct x, clearly need $m \geq k$
- if $m \geq n$ and A is full rank, we can reconstruct x without cardinality assumption
- when does the ℓ_1 heuristic (minimizing $\|x\|_1$ subject to $Ax = y$) reconstruct x (exactly)?

recent results by Candès, Donoho, Romberg, Tao, . . .

- (for some choices of A) if $m \geq (C \log n)k$, ℓ_1 heuristic reconstructs x exactly, with overwhelming probability
- C is absolute constant; valid A's include
  - $A_{ij} \sim \mathcal{N}(0, \sigma^2)$
  - Ax gives Fourier transform of x at m frequencies, chosen from uniform distribution

Total variation reconstruction

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

- convex-cardinality problem: minimize $\|\hat{x} - x_{\text{cor}}\|_2$ subject to $\mathbf{card}(Dx) \leq k$ (D is first order difference matrix)

- heuristic: minimize $\|\hat{x} - x_{\text{cor}}\|_2 + \gamma\|Dx\|_1$; vary γ to adjust number of jumps

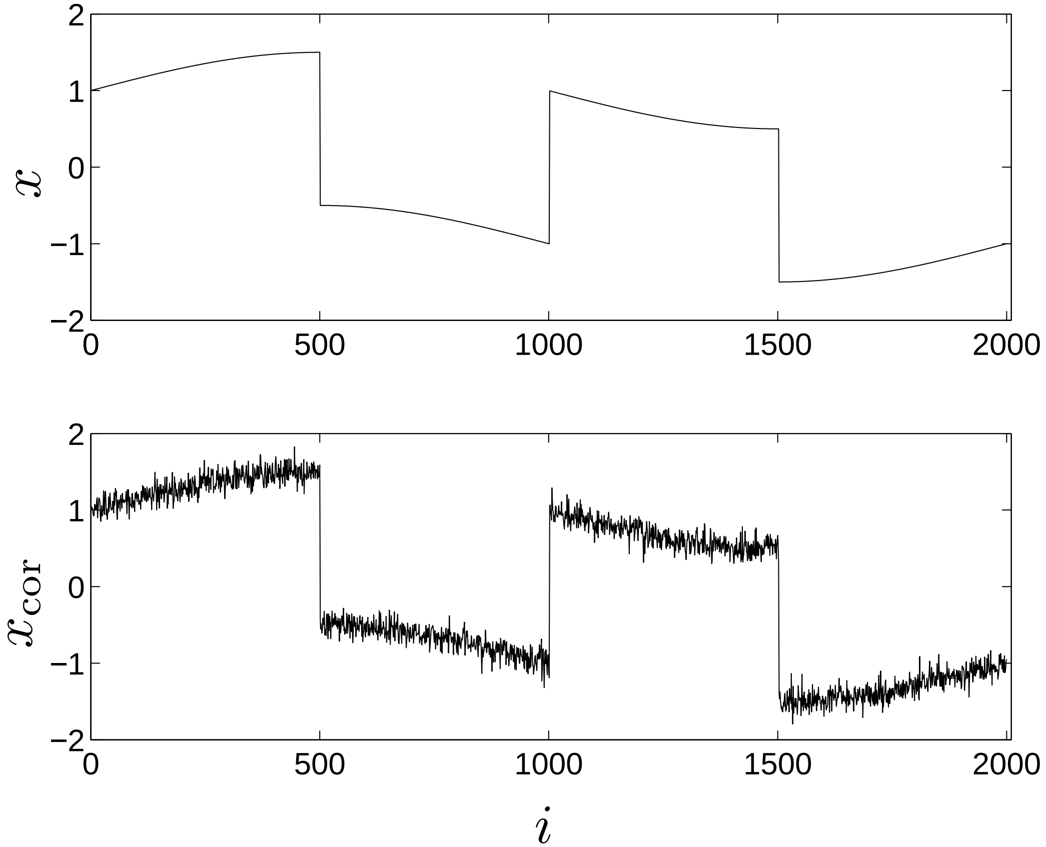
- $\|Dx\|_1$ is *total variation* of signal $\hat{x}$

- method is called *total variation reconstruction*

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

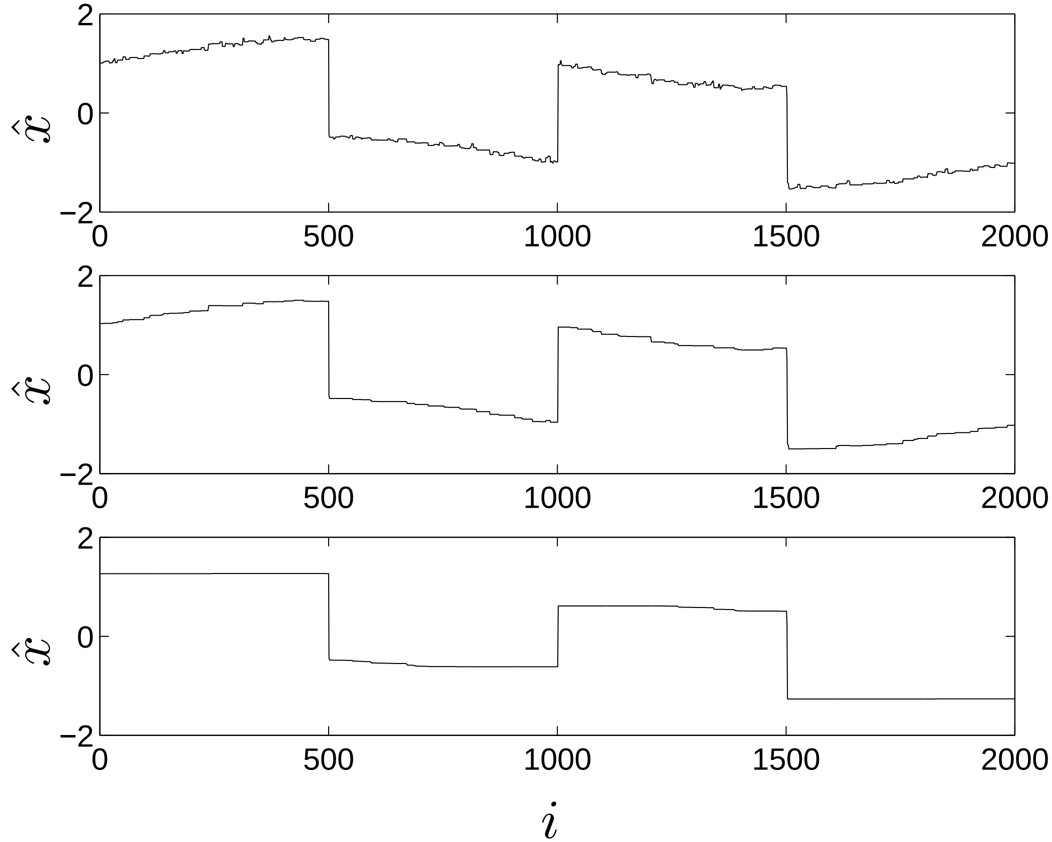
Example (§6.3.3 in BV book)

signal $x \in \mathbf{R}^{2000}$ and corrupted signal $x_{\mathrm{cor}} \in \mathbf{R}^{2000}$



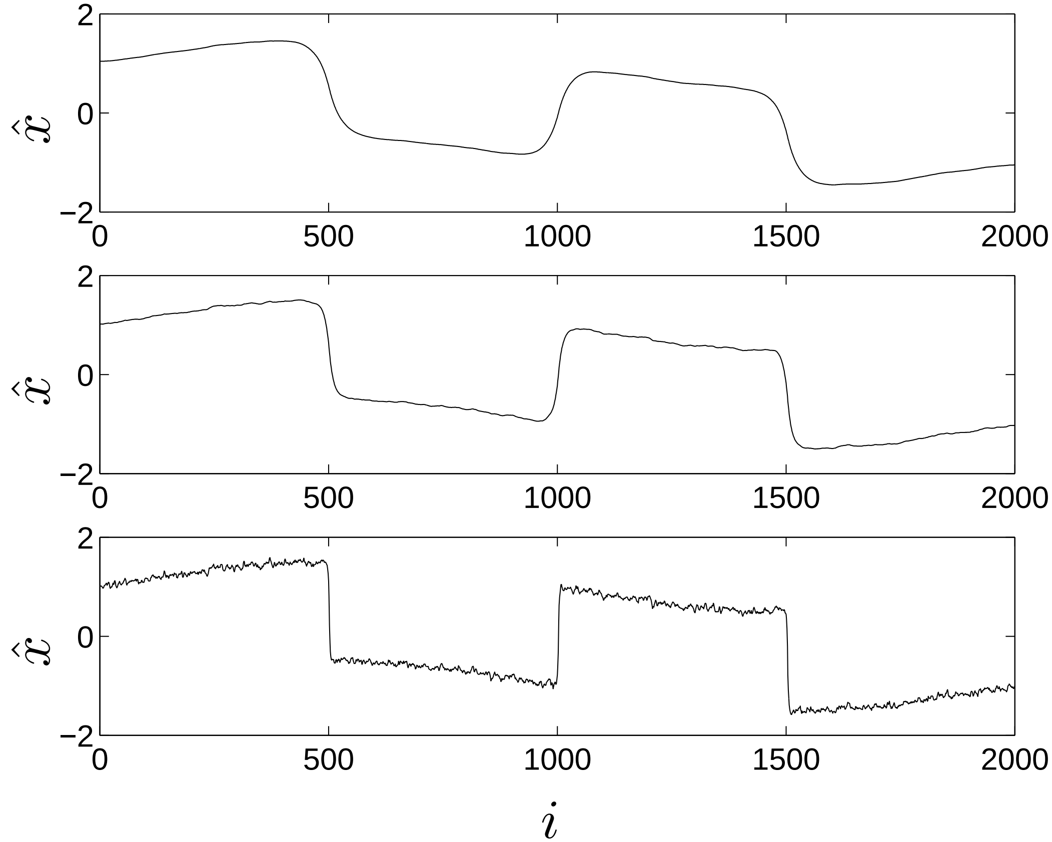
Total variation reconstruction

for three values of γ



ℓ_2 reconstruction

for three values of γ



Example: 2D total variation reconstruction

- $x \in \mathbf{R}^n$ are values of pixels on $N \times N$ grid ($N = 31$, so $n = 961$)

- assumption: x has relatively few big changes in value (*i.e.*, boundaries)

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

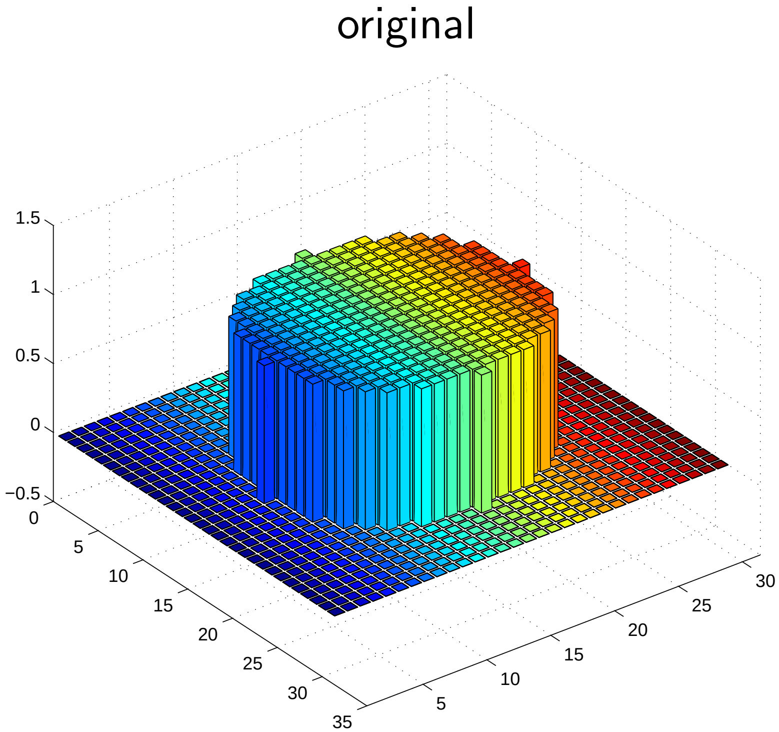
- as convex-cardinality problem:

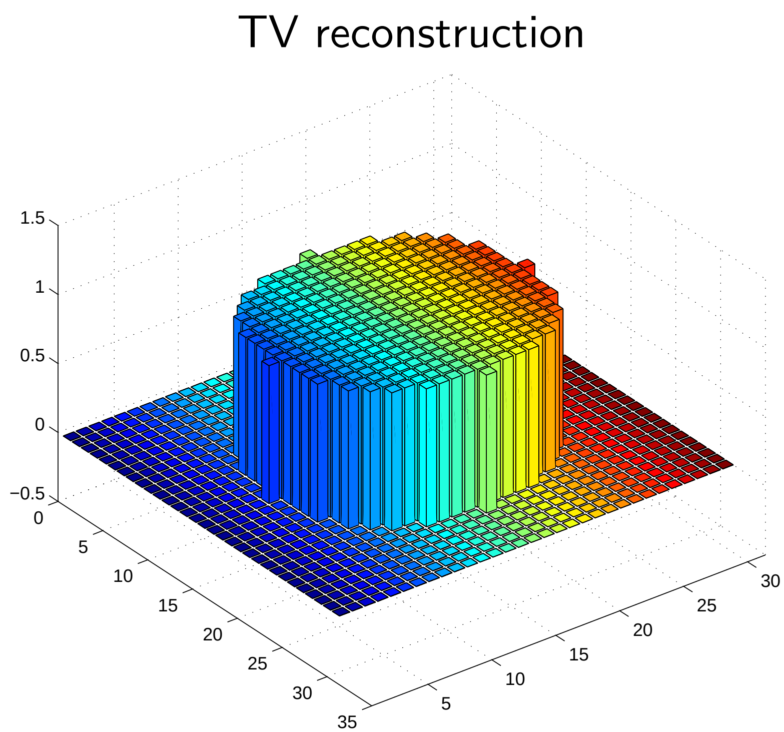
$$\begin{array}{ll} \text{minimize} & \mathbf{card}(x_{i,j} - x_{i+1,j}) + \mathbf{card}(x_{i,j} - x_{i,j+1}) \\ \text{subject to} & y = Fx \end{array}$$

- ℓ_1 heuristic (objective is a 2D version of total variation)

$$\begin{array}{ll} \text{minimize} & \sum |x_{i,j} - x_{i+1,j}| + \sum |x_{i,j} - x_{i,j+1}| \\ \text{subject to} & y = Fx \end{array}$$

TV reconstruction

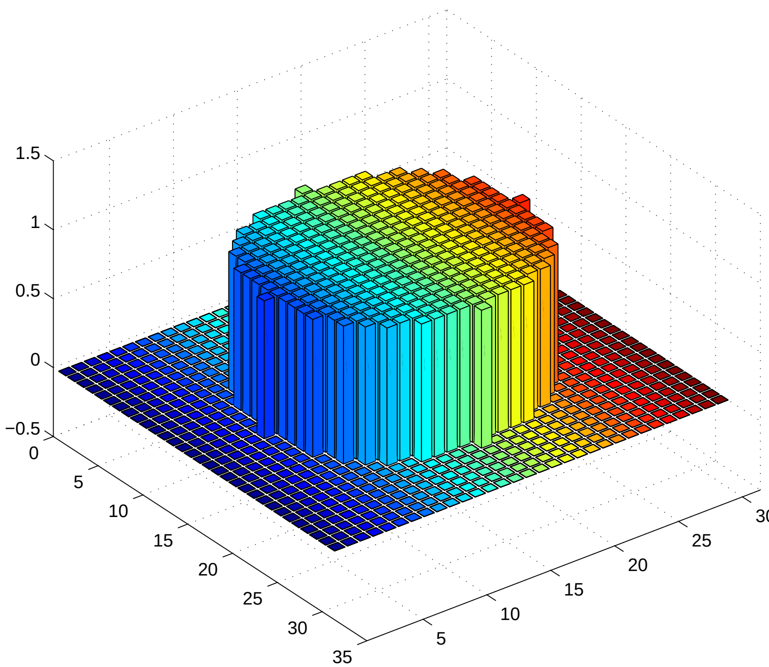




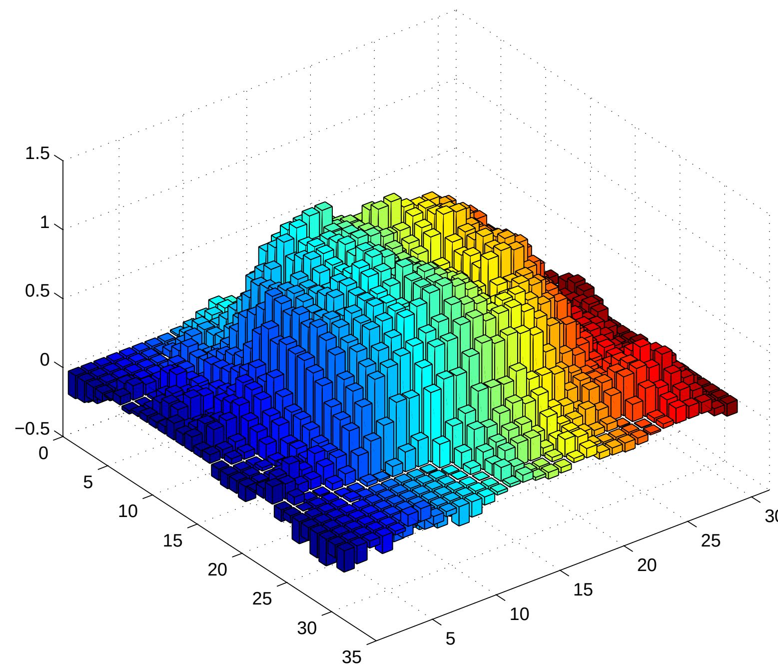
. . . not bad for $8\times$ more variables than measurements!

ℓ_2 reconstruction

original



ℓ_2 reconstruction



. . . this is what you'd expect with $8\times$ more variables than measurements

Iterated weighted ℓ_1 heuristic

- to minimize $\mathbf{card}(x)$ over $x \in \mathcal{C}$

$$
\begin{aligned}
&w := \mathbf{1} \\
&\text{repeat} \\
&\quad \text{minimize } \|\mathbf{diag}(w)x\|_1 \text{ over } x \in \mathcal{C} \\
&\quad w_i := 1/(\epsilon + |x_i|)
\end{aligned}
$$

- first iteration is basic ℓ_1 heuristic
- increases relative weight on small x_i
- typically converges in 5 or fewer steps
- often gives a modest improvement (*i.e.*, reduction in $\mathbf{card}(x)$) over basic ℓ_1 heuristic

Interpretation

- wlog we can take $x \succeq 0$ (by writing $x = x_+ - x_-$, $x_+, x_- \succeq 0$, and replacing $\mathbf{card}(x)$ with $\mathbf{card}(x_+) + \mathbf{card}(x_-)$)

- we'll use approximation $\mathbf{card}(z) \approx \log(1 + z/\epsilon)$, where $\epsilon > 0$, $z \in \mathbf{R}_+$

- using this approximation, we get (nonconvex) problem

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n \log(1 + x_i/\epsilon) \\ \text{subject to} & x \in \mathcal{C}, \quad x \succeq 0 \end{array}$$

- we'll find a local solution by linearizing objective at current point,

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

and solving resulting convex problem

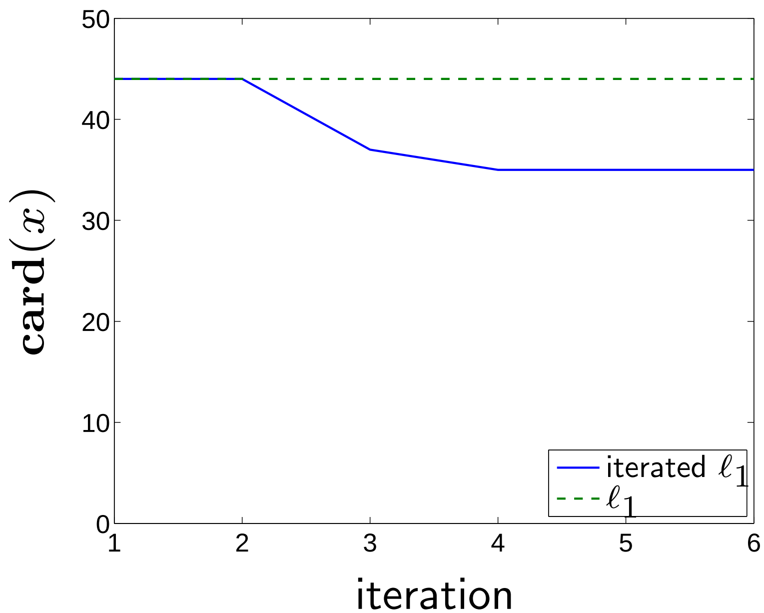
$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n w_i x_i \\ \text{subject to} & x \in \mathcal{C}, \quad x \succeq 0 \end{array}$$

with $w_i = 1/(\epsilon + x_i)$, to get next iterate

- repeat until convergence to get a local solution

Sparse solution of linear inequalities

- minimize $\mathbf{card}(x)$ over polyhedron $\{x \mid Ax \preceq b\}$, $A \in \mathbf{R}^{100\times 50}$
- ℓ_1 heuristic finds $x \in \mathbf{R}^{50}$ with $\mathbf{card}(x) = 44$
- iterated weighted ℓ_1 heuristic finds x with $\mathbf{card}(x) = 36$ (global solution, via branch & bound, is $\mathbf{card}(x) = 32$)



Detecting changes in time series model

- AR(2) scalar time-series model

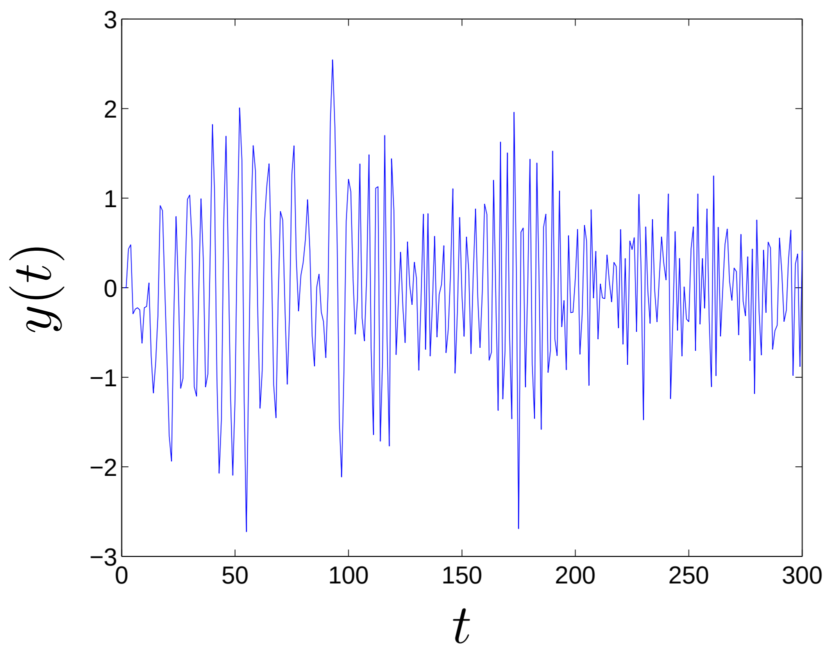
$$y(t+2) = a(t)y(t+1) + b(t)y(t) + v(t), \quad v(t) \text{ IID } \mathcal{N}(0, 0.5^2)$$

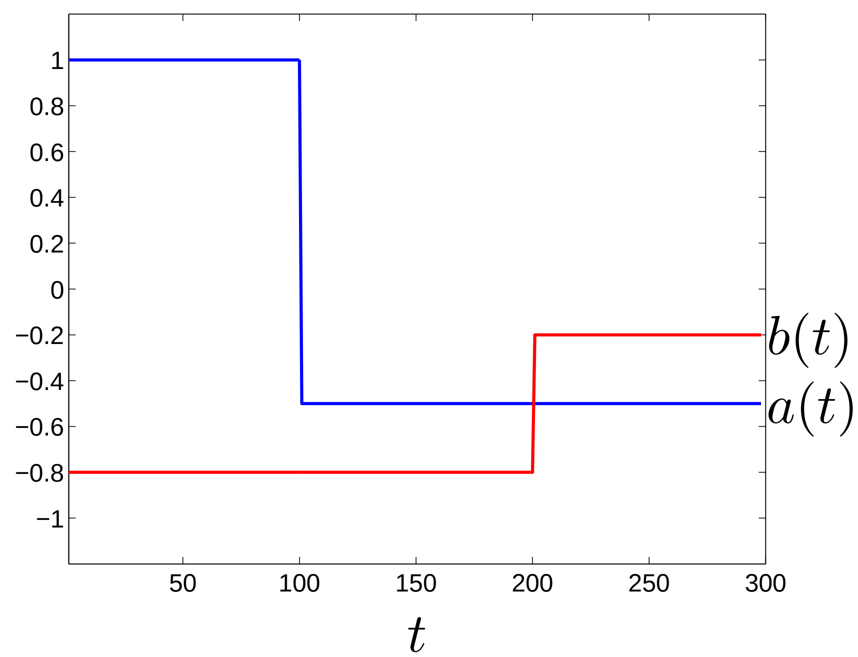
- assumption: $a(t)$ and $b(t)$ are piecewise constant, change infrequently
- given $y(t)$, $t = 1, \ldots, T$, estimate $a(t)$, $b(t)$, $t = 1, \ldots, T-2$
- heuristic: minimize over variables $a(t)$, $b(t)$, $t = 1, \ldots, T-1$

$$\begin{array}{l} \sum_{t=1}^{T-2}(y(t+2) - a(t)y(t+1) - b(t)y(t))^2 \\ \quad +\gamma \sum_{t=1}^{T-2}(|a(t+1) - a(t)| + |b(t+1) - b(t)|) \end{array}$$

- vary γ to trade off fit versus number of changes in a, b

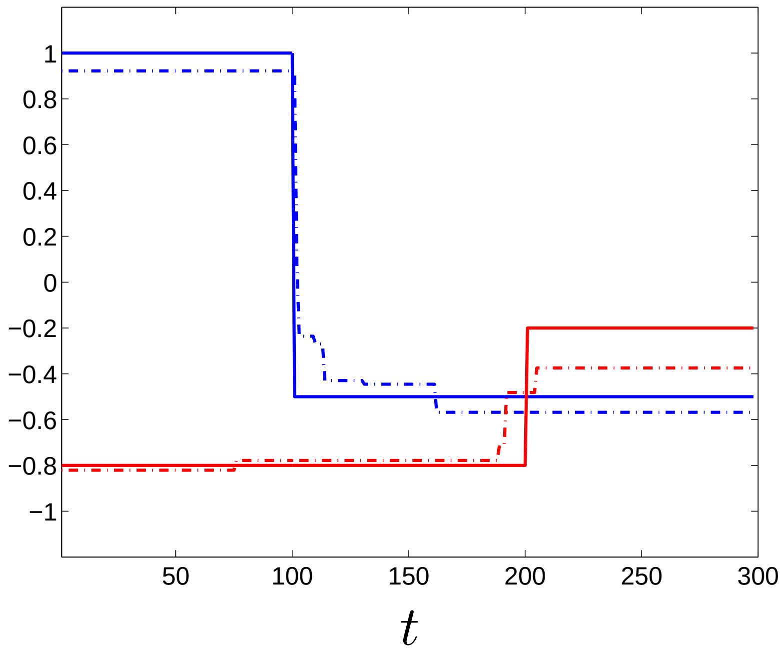
Time series and true coefficients

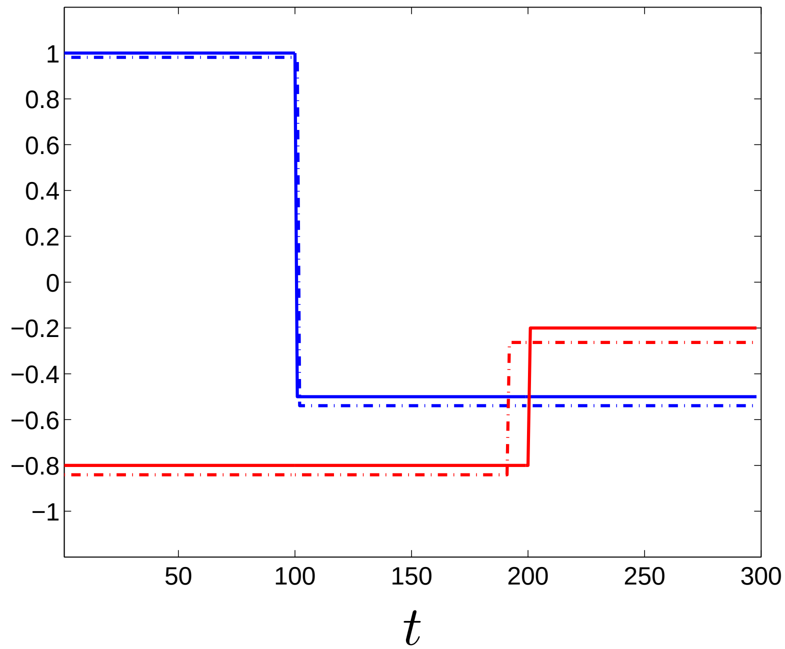




TV heuristic and iterated TV heuristic

left: TV with $\gamma = 10$;   *right:* iterated TV, 5 iterations, $\epsilon = 0.005$





Extension to matrices

- $\mathbf{Rank}$ is natural analog of $\mathbf{card}$ for matrices
- convex-rank problem: convex, except for $\mathbf{Rank}$ in objective or constraints
- rank problem reduces to card problem when matrices are diagonal: $\mathbf{Rank}(\mathbf{diag}(x)) = \mathbf{card}(x)$
- analog of ℓ_1 heuristic: use *nuclear norm*, $\|X\|_* = \sum_i \sigma_i(X)$ (sum of singular values; dual of spectral norm)
- for $X \succeq 0$, reduces to $\mathbf{Tr}\, X$ (for $x \succeq 0$, $\|x\|_1$ reduces to $\mathbf{1}^T x$)

Factor modeling

- given matrix $\Sigma \in \mathbf{S}_+^n$, find approximation of form $\hat{\Sigma} = FF^T + D$, where $F \in \mathbf{R}^{n \times r}$, D is diagonal nonnegative
- gives underlying factor model (with r factors)

$$x = Fz + v, \quad v \sim \mathcal{N}(0, D), \quad z \sim \mathcal{N}(0, I)$$

- model with fewest factors:

$$\begin{array}{ll} \text{minimize} & \mathbf{Rank}\, X \\ \text{subject to} & X \succeq 0, \quad D \succeq 0 \text{ diagonal} \\ & X + D \in \mathcal{C} \end{array}$$

with variables $D, \ X \in \mathbf{S}^n$
$\mathcal{C}$ is convex set of acceptable approximations to Σ

- $e.g.$, via KL divergence

$$\mathcal{C} = \{\hat{\Sigma} \mid -\log\det(\Sigma^{-1/2}\hat{\Sigma}\Sigma^{-1/2}) + \mathbf{Tr}(\Sigma^{-1/2}\hat{\Sigma}\Sigma^{-1/2}) - n \leq \epsilon\}$$

- trace heuristic:

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}\, X \\ \text{subject to} & X \succeq 0, \quad D \succeq 0 \text{ diagonal} \\ & X + D \in \mathcal{C} \end{array}$$

with variables $d \in \mathbf{R}^n$, $X \in \mathbf{S}^n$

Example

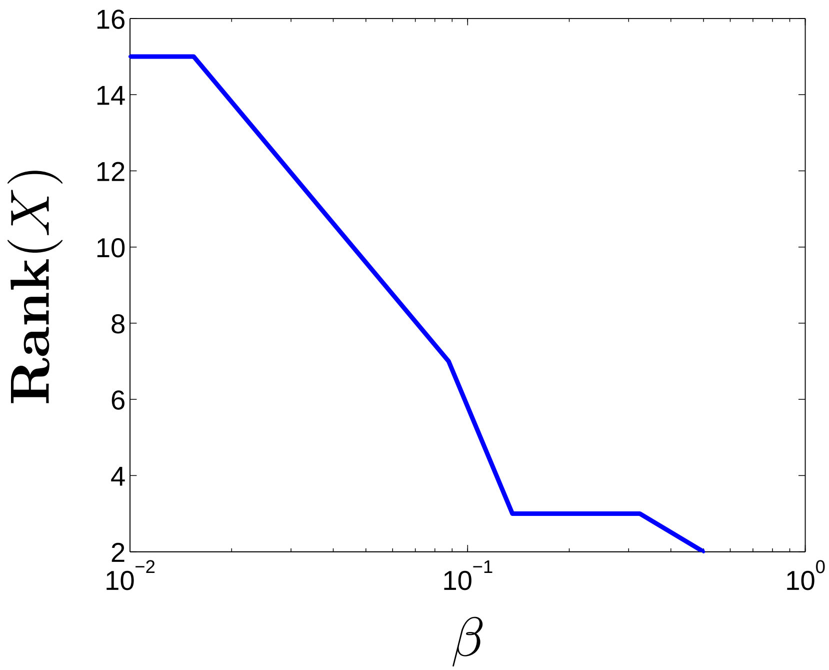
- $x = Fz + v$, $z \sim \mathcal{N}(0, I)$, $v \sim \mathcal{N}(0, D)$, D diagonal; $F \in \mathbf{R}^{20 \times 3}$
- Σ is empirical covariance matrix from $N = 3000$ samples
- set of acceptable approximations

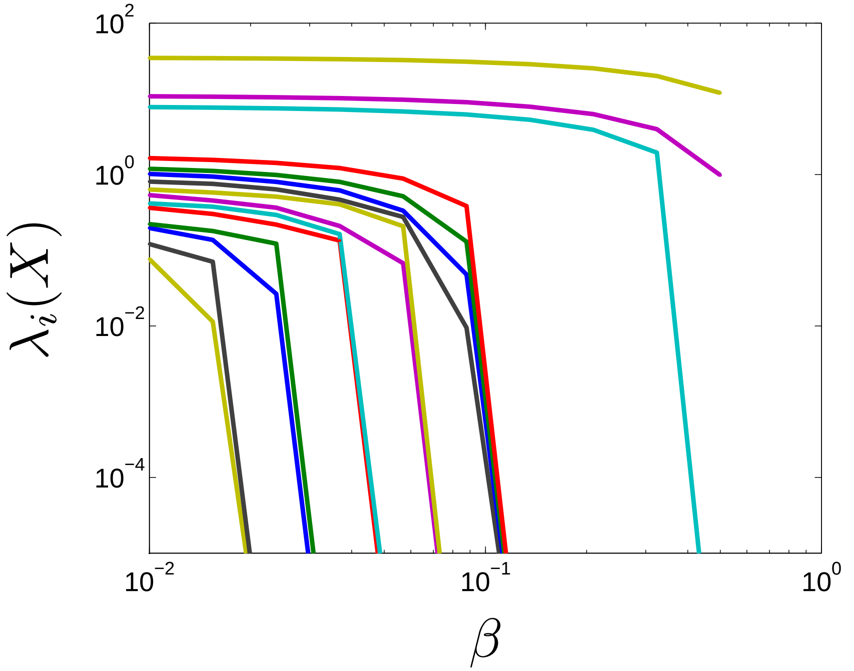
$$\mathcal{C} = \{\hat{\Sigma} \mid \|\Sigma^{-1/2}(\hat{\Sigma} - \Sigma)\Sigma^{-1/2}\| \leq \beta\}$$

- trace heuristic

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}\, X \\ \text{subject to} & X \succeq 0, \quad d \succeq 0 \\ & \|\Sigma^{-1/2}(X + \mathbf{diag}(d) - \Sigma)\Sigma^{-1/2}\| \leq \beta \end{array}$$

Trace approximation results





- for $\beta = 0.1357$ (knee of the tradeoff curve) we find
  - $\angle\left(\mathrm{range}(X), \mathrm{range}(FF^T)\right) = 6.8^\circ$
  - $\|d - \mathbf{diag}(D)\| / \|\mathbf{diag}(D)\| = 0.07$
- *i.e.*, we have recovered the factor model from the empirical covariance

Summary and conclusions

- convex-cardinality (and rank) problems arise in many applications
- these problems are hard (to solve exactly, in general)
- heuristics based on ℓ_1 norm (or nuclear norm for rank)
  - are convex, hence solvable
  - give very good results in practice
- is basis of many well known methods
  (lasso, SVM, compressed sensing, TV denoising, . . .)