ℓ_1-norm Methods for Convex-Cardinality Problems

Part II

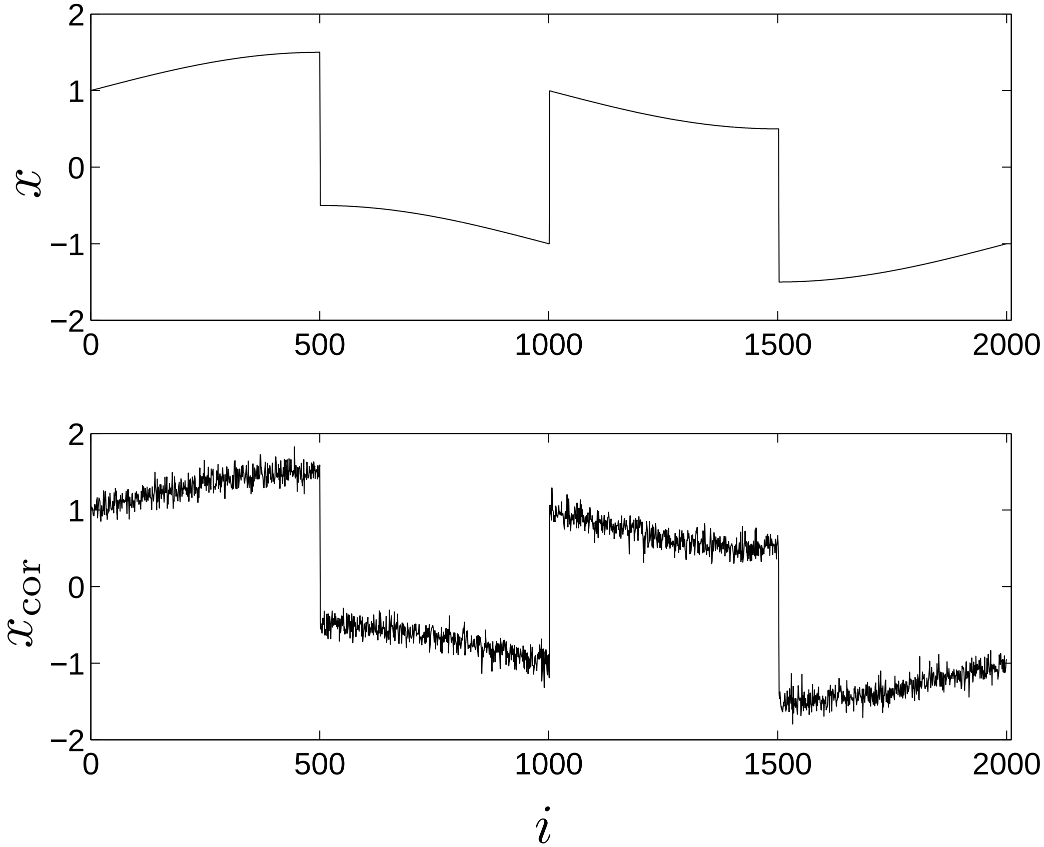
- total variation
- iterated weighted ℓ_1 heuristic
- matrix rank constraints

Total variation reconstruction

- fit x_{cor} with piecewise constant $\hat{x}$, no more than k jumps
- convex-cardinality problem: minimize $\|\hat{x} - x_{\mathrm{cor}}\|_2$ subject to $\mathbf{card}(Dx) \leq k$ (D is first order difference matrix)
- heuristic: minimize $\|\hat{x} - x_{\mathrm{cor}}\|_2 + \gamma \|Dx\|_1$; 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

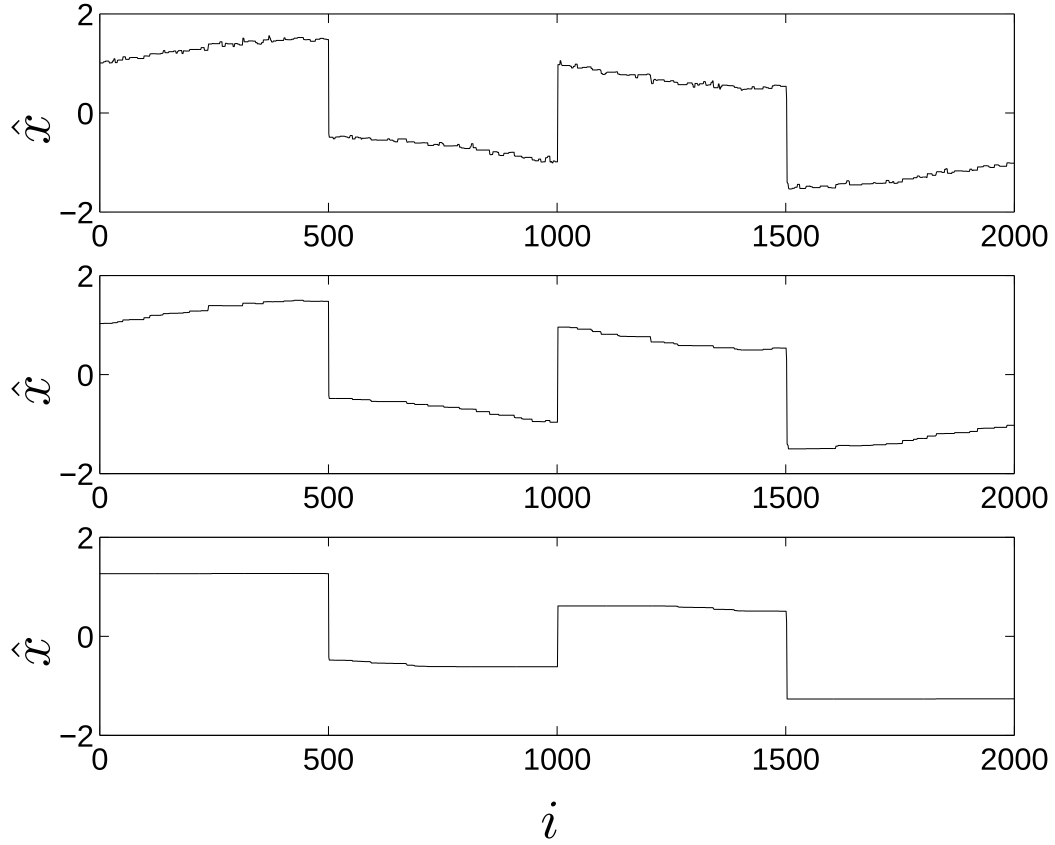
Example (§6.3.3 in BV book)

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



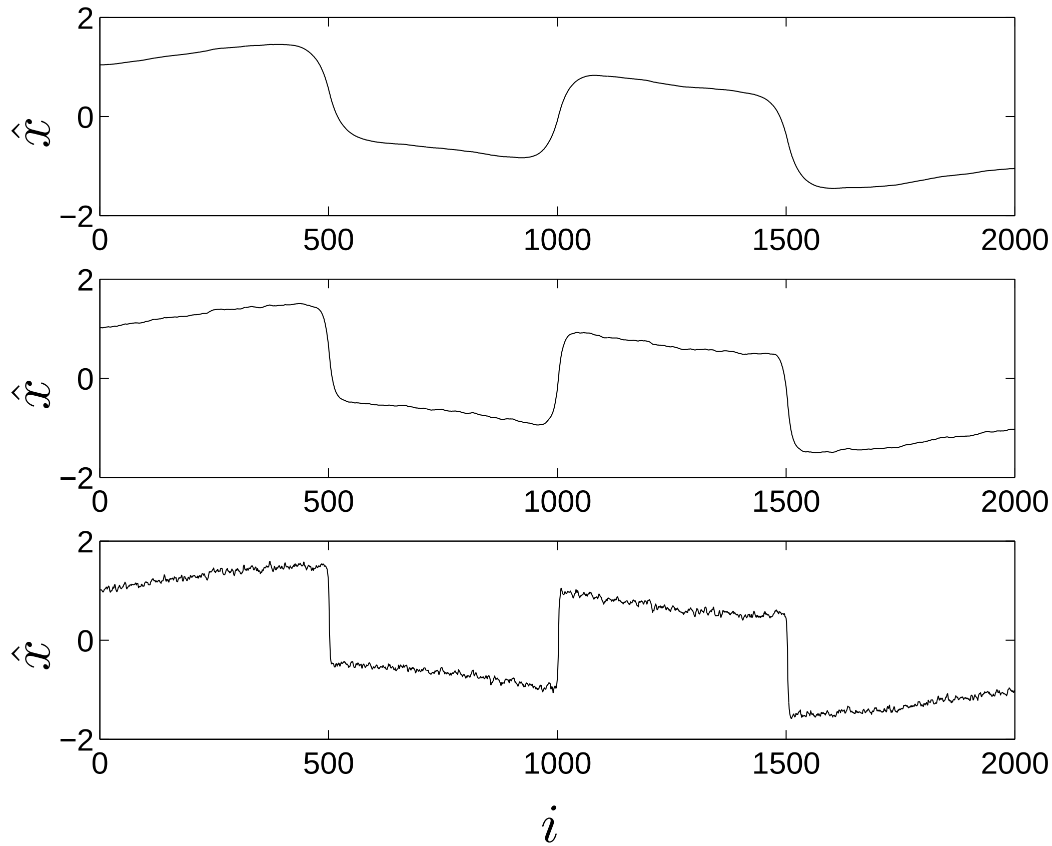
Total variation reconstruction

for three values of γ



ℓ_2 reconstruction

for three values of γ



Example: 2D total variation reconstruction

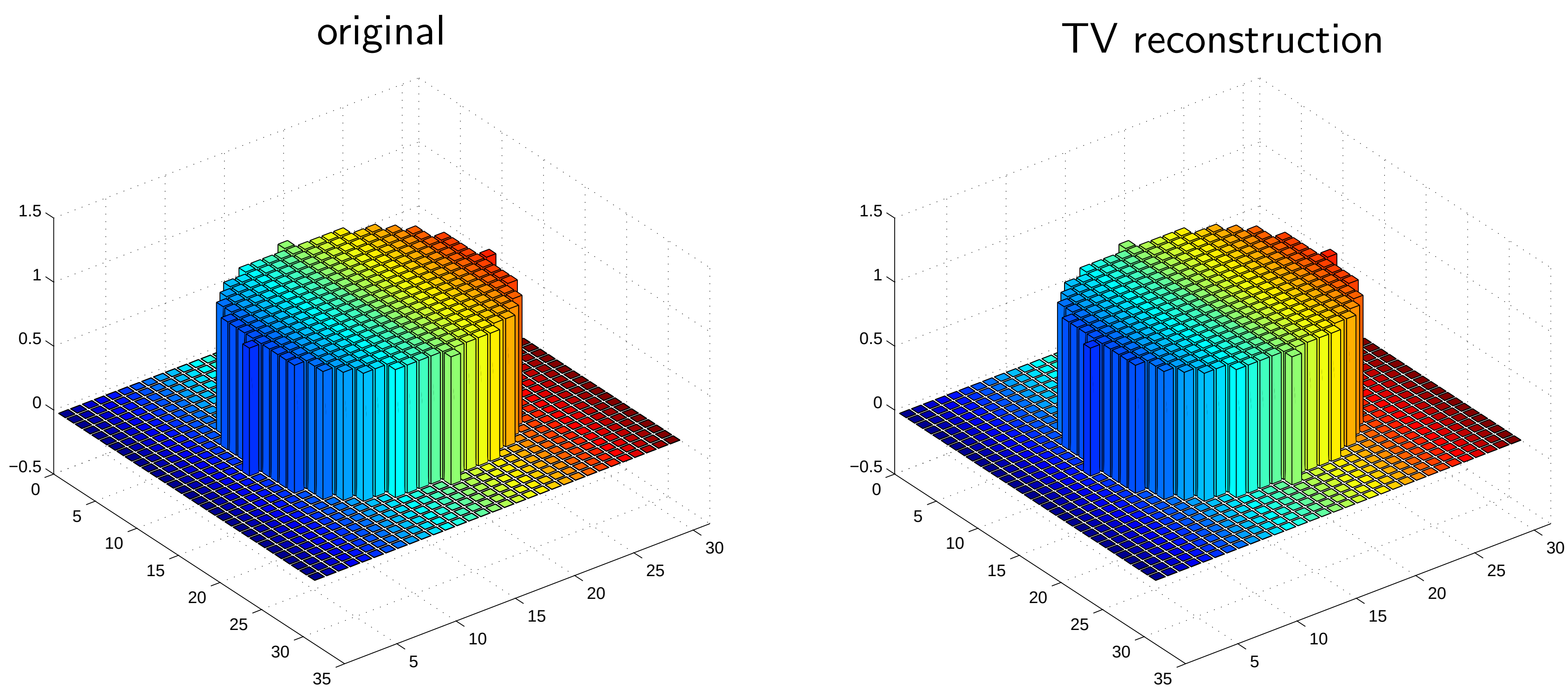
- $x \in \mathbf{R}^n$ are values of pixels on $N \times N$ grid ($N = 31$, so $n = 961$)
- assumption: x has relatively few big changes in value ($i.e.$, boundaries)
- we have $m = 120$ linear measurements, $y = Fx$ ($F_{ij} \sim \mathcal{N}(0, 1)$)
- as convex-cardinality problem:

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

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

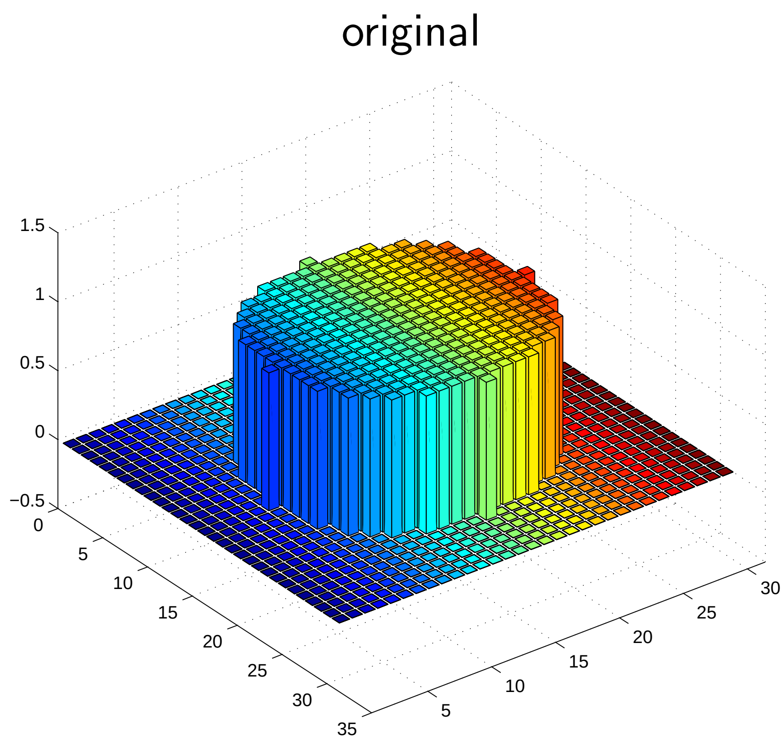
$$\begin{array}{ll} \text{minimize} & \sum |x_{i,j} - x_{i+1,j}| + \sum |x_{i,j} - x_{i,j+1}| \\ \text{subject to} & y = Fx \end{array}$$

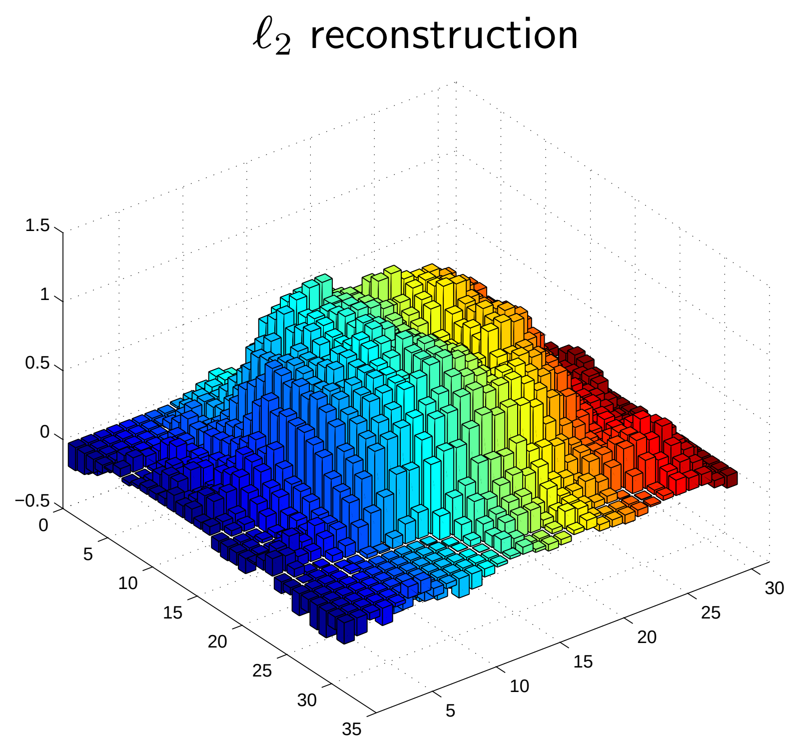
TV reconstruction



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

ℓ_2 reconstruction





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

Iterated weighted ℓ_1 heuristic

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

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

- 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

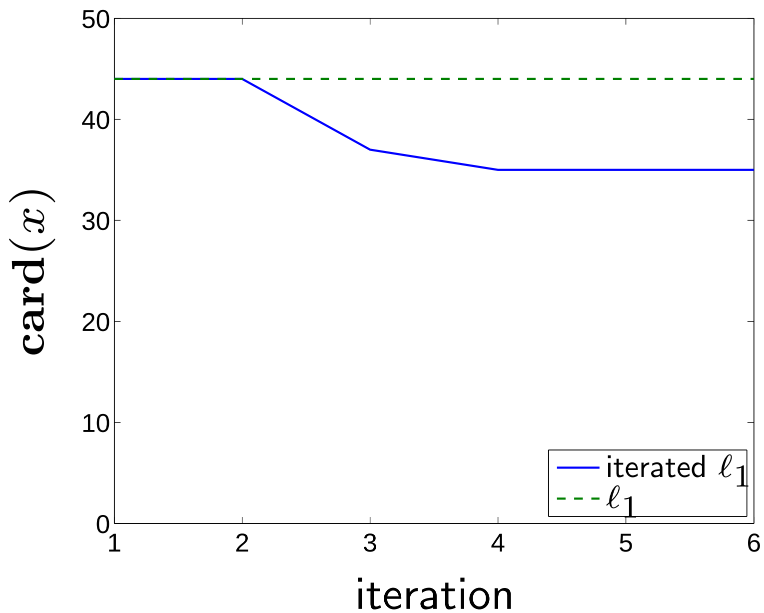
$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{n} w_i x_i \\ \text{subject to} & x \in \mathcal{C}, \quad x \succeq 0 \end{array}$$

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

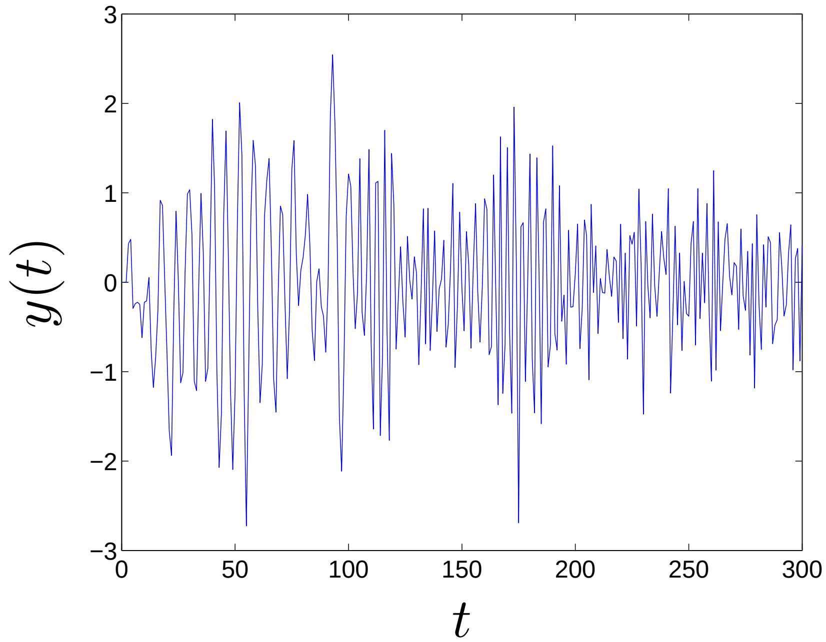
$$y(t+2) = a(t)y(t+1) + b(t)y(t) + v(t), \quad v(t) \text{ IID } \mathcal{N}(0, 0.5^2)$$

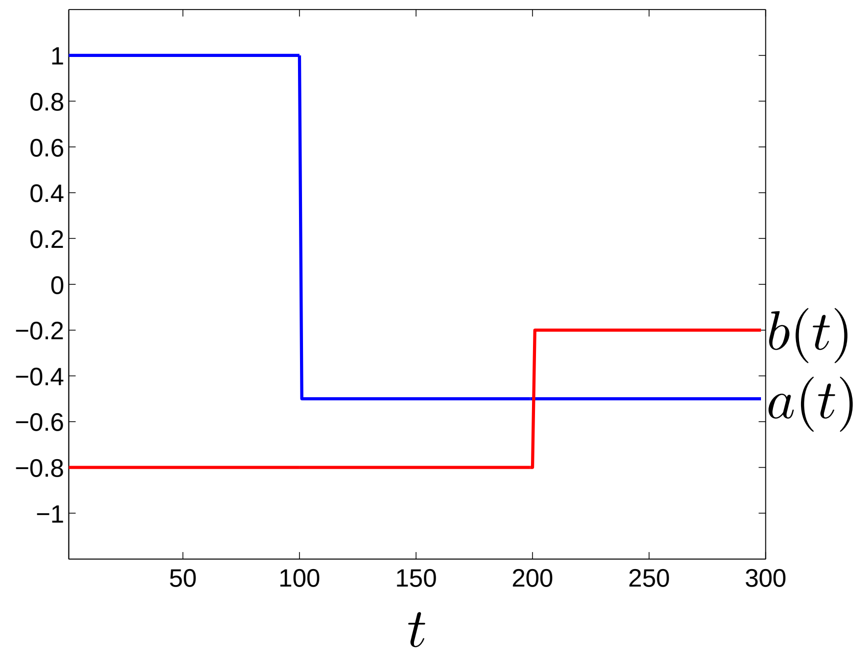
- assumption: $a(t)$ and $b(t)$ are piecewise constant, change infrequently
- given $y(t)$, $t = 1, \ldots, T$, estimate $a(t)$, $b(t)$, $t = 1, \ldots, T-2$
- heuristic: minimize over variables $a(t)$, $b(t)$, $t = 1, \ldots, T-1$

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

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

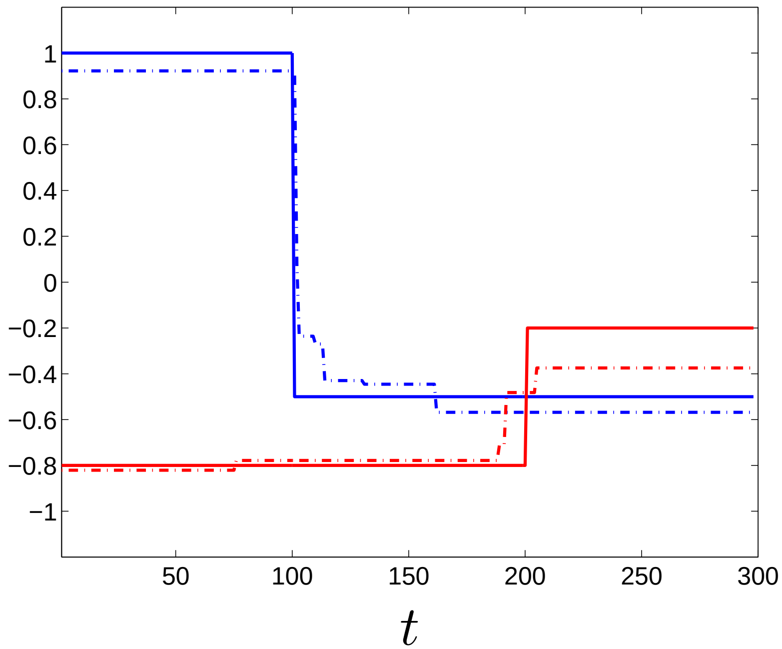
Time series and true coefficients

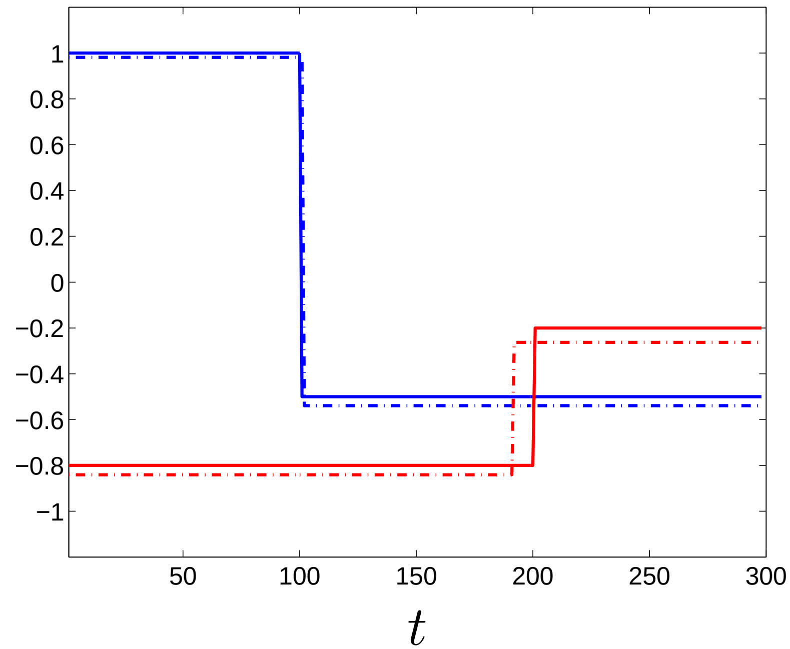




TV heuristic and iterated TV heuristic

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

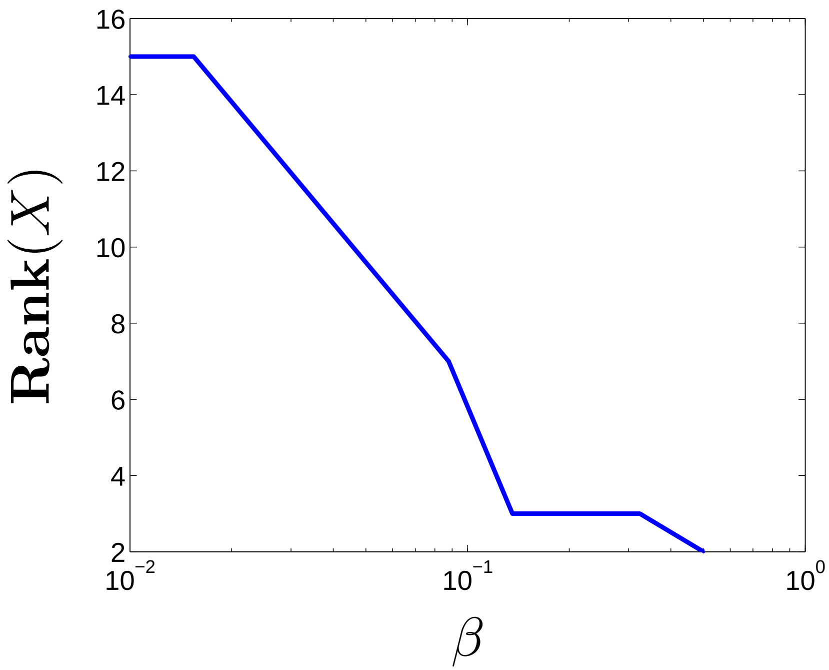
- $x = Fz + v$, $z \sim \mathcal{N}(0, I)$, $v \sim \mathcal{N}(0, D)$, D diagonal; $F \in \mathbf{R}^{20\times 3}$
- Σ is empirical covariance matrix from $N = 3000$ samples
- set of acceptable approximations

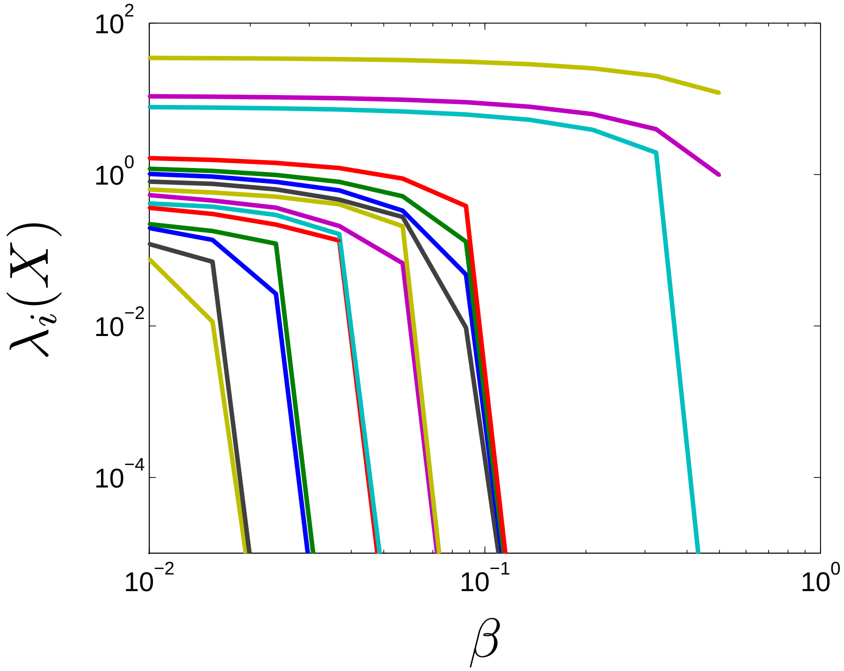
$$\mathcal{C} = \{\hat{\Sigma} \mid \|\Sigma^{-1/2}(\hat{\Sigma} - \Sigma)\Sigma^{-1/2}\| \leq \beta\}$$

- 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