$\ell_1$-norm Methods for Convex-Cardinality Problems

Part II

- total variation
- iterated weighted $\ell_1$ heuristic
- matrix rank constraints

Prof. S. Boyd, EE364b, Stanford University
Total variation reconstruction

• fit \( x_{\text{cor}} \) with piecewise constant \( \hat{x} \), no more than \( k \) jumps

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

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

• \( \| Dx \|_1 \) is total variation of signal \( \hat{x} \)

• method is called total variation reconstruction

• unlike \( \ell_2 \) based reconstruction, TVR filters high frequency noise out while preserving sharp jumps
Example ($\S$6.3.3 in BV book)

signal $x \in \mathbb{R}^{2000}$ and corrupted signal $x_{\text{cor}} \in \mathbb{R}^{2000}$
Total variation reconstruction

for three values of $\gamma$

Prof. S. Boyd, EE364b, Stanford University
$\ell_2$ reconstruction

for three values of $\gamma$

Prof. S. Boyd, EE364b, Stanford University
Example: 2D total variation reconstruction

- $x \in \mathbb{R}^n$ 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:

\[
\begin{align*}
\text{minimize} & \quad \text{card}(x_{i,j} - x_{i+1,j}) + \text{card}(x_{i,j} - x_{i,j+1}) \\
\text{subject to} & \quad y = Fx
\end{align*}
\]

- $\ell_1$ heuristic (objective is a 2D version of total variation)

\[
\begin{align*}
\text{minimize} & \quad \sum |x_{i,j} - x_{i+1,j}| + \sum |x_{i,j} - x_{i,j+1}| \\
\text{subject to} & \quad y = Fx
\end{align*}
\]
TV reconstruction

original

TV reconstruction

... not bad for $8 \times$ more variables than measurements!
\( \ell_2 \) reconstruction

... this is what you’d expect with \( 8 \times \) more variables than measurements
Iterated weighted $\ell_1$ heuristic

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

  \[
  w := 1
  \]

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

- first iteration is basic $\ell_1$ heuristic
- increases relative weight on small $x_i$
- typically converges in 5 or fewer steps
- often gives a modest improvement (i.e., reduction in $\text{card}(x)$) over basic $\ell_1$ heuristic
Interpretation

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

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

- using this approximation, we get (nonconvex) problem

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

- 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

\[
\begin{align*}
\text{minimize} & \quad \sum_{i=1}^{n} w_i x_i \\
\text{subject to} & \quad x \in \mathcal{C}, \quad x \succeq 0
\end{align*}
\]

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 $\text{card}(x)$ over polyhedron $\{ x \mid Ax \preceq b \}$, $A \in \mathbb{R}^{100 \times 50}$
- $\ell_1$ heuristic finds $x \in \mathbb{R}^{50}$ with $\text{card}(x) = 44$
- iterated weighted $\ell_1$ heuristic finds $x$ with $\text{card}(x) = 36$
  (global solution, via branch & bound, is $\text{card}(x) = 32$)
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 \( a(t), b(t), t = 1, \ldots, T - 1 \)

\[
\sum_{t=1}^{T-2} (y(t + 2) - a(t)y(t + 1) - b(t)y(t))^2 \\
+ \gamma \sum_{t=1}^{T-2} (|a(t + 1) - a(t)| + |b(t + 1) - b(t)|)
\]

• vary \( \gamma \) to trade off fit versus number of changes in \( a, b \)
Time series and true coefficients

\[ y(t) \]

\[ a(t) \quad b(t) \]
TV heuristic and iterated TV heuristic

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

- **Rank** is natural analog of **card** for matrices

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

- rank problem reduces to card problem when matrices are diagonal: 
  \[ \text{Rank}(\text{diag}(x)) = \text{card}(x) \]

- analog of \( \ell_1 \) heuristic: use **nuclear norm**, 
  \[ \|X\|_* = \sum_i \sigma_i(X) \]  
  (sum of singular values; dual of spectral norm)

- for \( X \succeq 0 \), reduces to \( \text{Tr } X \)  
  (for \( x \succeq 0 \), \( \|x\|_1 \) reduces to \( 1^T x \))
Factor modeling

- given matrix $\Sigma \in \mathbb{S}^n_+$, find approximation of form $\hat{\Sigma} = FF^T + D$, where $F \in \mathbb{R}^{n \times r}$, $D$ 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:

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

with variables $D$, $X \in \mathbb{S}^n$

$\mathcal{C}$ is convex set of acceptable approximations to $\Sigma$
• e.g., via KL divergence

\[ C = \left\{ \hat{\Sigma} \mid - \log \det(\Sigma^{-1/2}\hat{\Sigma}\Sigma^{-1/2}) + \text{Tr}(\Sigma^{-1/2}\hat{\Sigma}\Sigma^{-1/2}) - n \leq \epsilon \right\} \]

• trace heuristic:

\[
\begin{align*}
\text{minimize} & \quad \text{Tr } X \\
\text{subject to} & \quad X \succeq 0, \quad D \succeq 0 \text{ diagonal} \\
& \quad X + D \in C
\end{align*}
\]

with variables \( d \in \mathbb{R}^n, \; X \in \mathbb{S}^n \)
Example

• \( x = Fz + v, \ z \sim \mathcal{N}(0, I), \ v \sim \mathcal{N}(0, D), \ D \) diagonal; \( F \in \mathbb{R}^{20 \times 3} \)

• \( \Sigma \) is empirical covariance matrix from \( N = 3000 \) samples

• set of acceptable approximations

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

• trace heuristic

\[
\begin{align*}
\text{minimize} & \quad \text{Tr } X \\
\text{subject to} & \quad X \succeq 0, \quad d \succeq 0 \\
& \quad \| \Sigma^{-1/2}(X + \text{diag}(d) - \Sigma)\Sigma^{-1/2} \| \leq \beta 
\end{align*}
\]
Trace approximation results

![Graphs showing rank and eigenvalues vs beta]
• for $\beta = 0.1357$ (knee of the tradeoff curve) we find

- $\angle (\text{range}(X), \text{range}(F F^T)) = 6.8^\circ$
- $\|d - \text{diag}(D)\| / \|\text{diag}(D)\| = 0.07$

• i.e., we have recovered the factor model from the empirical covariance