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

Stephen Boyd

Prof. S. Boyd, EE364b, Stanford University
Outline

• problems involving cardinality
• the $\ell_1$-norm heuristic
• convex relaxation and convex envelope interpretations
• examples
• recent results
• total variation
• iterated weighted $\ell_1$ heuristic
• matrix rank constraints
\textbf{$\ell_1$-norm heuristics for cardinality problems}

- cardinality problems arise often, but are hard to solve exactly

- a simple heuristic, that relies on $\ell_1$-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 $x \in \mathbb{R}^n$, denoted $\text{card}(x)$, is the number of nonzero components of $x$
- $\text{card}$ is separable; for scalar $x$, $\text{card}(x) = \begin{cases} 0 & x = 0 \\ 1 & x \neq 0 \end{cases}$
- $\text{card}$ is quasiconcave on $\mathbb{R}_+^n$ (but not $\mathbb{R}^n$) since
  \[ \text{card}(x + y) \geq \min\{\text{card}(x), \text{card}(y)\} \]
  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 \text{card} in objective or constraints.

Examples (with $C$, $f$ convex):

- Convex minimum cardinality problem:
  \[
  \begin{align*}
  \text{minimize} & \quad \text{card}(x) \\
  \text{subject to} & \quad x \in C
  \end{align*}
  \]

- Convex problem with cardinality constraint:
  \[
  \begin{align*}
  \text{minimize} & \quad f(x) \\
  \text{subject to} & \quad x \in C, \quad \text{card}(x) \leq k
  \end{align*}
  \]
Solving convex-cardinality problems

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

• 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:
  \[
  \begin{align*}
  \text{minimize} & \quad c^T x \\
  \text{subject to} & \quad Ax \preceq b, \quad x_i \in \{0, 1\}
  \end{align*}
\]

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

- can be expressed as

  \[
  \begin{align*}
  \text{minimize} & \quad c^T x \\
  \text{subject to} & \quad Ax \preceq b, \quad \text{card}(x) + \text{card}(1 - x) \leq n
  \end{align*}
\]

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

- conclusion: general convex-cardinality problem is hard
Sparse design

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

• 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 \( b \in \mathbb{R}^m \) as a linear combination of \( k \) regressors (chosen from \( n \) possible regressors)

\[
\begin{align*}
\text{minimize} & \quad \|Ax - b\|_2 \\
\text{subject to} & \quad \text{card}(x) \leq k
\end{align*}
\]

- 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 \( \text{card}(x) \) subject to \( \|Ax - b\|_2 \leq \epsilon \)
  - minimize \( \|Ax - b\|_2 + \lambda \text{card}(x) \)
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 $\text{card}(x) \leq k$

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

\[
\begin{align*}
\text{minimize} & \quad \|Ax - y\|_2 \\
\text{subject to} & \quad \text{card}(x) \leq k
\end{align*}
\]
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: $\text{card}(w) \leq k$
- $\mathcal{B} = \{ i \mid w_i \neq 0 \}$ is set of bad measurements or outliers
- maximum likelihood estimate of $x$ found by solving
  
  $\begin{align*}
  \text{minimize} & \quad \sum_{i \not\in \mathcal{B}} (y_i - a_i^T x)^2 \\
  \text{subject to} & \quad |\mathcal{B}| \leq k
  \end{align*}$

  with variables $x$ and $\mathcal{B} \subseteq \{1, \ldots, m\}$
- equivalent to
  
  $\begin{align*}
  \text{minimize} & \quad \|y - Ax - w\|_2^2 \\
  \text{subject to} & \quad \text{card}(w) \leq k
  \end{align*}$
Minimum number of violations

• set of convex inequalities

\[ f_1(x) \leq 0, \ldots, f_m(x) \leq 0, \quad x \in C \]

• choose \( x \) to minimize the number of violated inequalities:

\[
\begin{align*}
\text{minimize} & \quad \text{card}(t) \\
\text{subject to} & \quad f_i(x) \leq t_i, \quad i = 1, \ldots, m \\
& \quad x \in C, \quad t \geq 0
\end{align*}
\]

• determining whether zero inequalities can be violated is (easy) convex feasibility problem
Linear classifier with fewest errors

• given data \((x_1, y_1), \ldots, (x_m, y_m) \in \mathbb{R}^n \times \{-1, 1\}\)

• we seek linear (affine) classifier \(y \approx \text{sign}(w^T x + v)\)

• classification error corresponds to \(y_i(w^T x + v) \leq 0\)

• to find \(w, v\) that give fewest classification errors:

\[
\begin{align*}
\text{minimize} & \quad \text{card}(t) \\
\text{subject to} & \quad y_i(w^T x_i + v) + t_i \geq 1, \quad i = 1, \ldots, m
\end{align*}
\]

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
  \[ g(\lambda) = \inf_{x} \left( \sum_{i=1}^{m} \lambda_i f_i(x) \right) \geq 1, \lambda \succeq 0 \]

- to find smallest cardinality infeasible subset, we solve

\[
\begin{align*}
\text{minimize} & \quad \text{card}(\lambda) \\
\text{subject to} & \quad g(\lambda) \geq 1, \quad \lambda \succeq 0
\end{align*}
\]

(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: $\beta \text{card}(x) + \alpha^T x$
- budget constraint is $1^T x + \beta \text{card}(x) + \alpha^T x \leq B$
- mean return on investment is $\mu^T x$; variance is $x^T \Sigma x$
- minimize investment variance (risk) with mean return $\geq R_{\text{min}}$:

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

$l_1$-norm methods for convex-cardinality problems
Piecewise constant fitting

- fit corrupted $x_{\text{cor}}$ 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 $\iff \text{card}(D\hat{x}) \leq k$, where

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

- as convex-cardinality problem:

$$\begin{align*}
\text{minimize} & \quad \|\hat{x} - x_{\text{cor}}\|_2 \\
\text{subject to} & \quad \text{card}(D\hat{x}) \leq k
\end{align*}$$
Piecewise linear fitting

• fit $x_{cor}$ by a piecewise linear signal $\hat{x}$ with $k$ or fewer kinks

• as convex-cardinality problem:

$$\begin{align*}
\text{minimize} & \quad \|\hat{x} - x_{cor}\|_2 \\
\text{subject to} & \quad \text{card}(\nabla^2 \hat{x}) \leq k
\end{align*}$$

where

$$\nabla^2 = \begin{bmatrix}
-1 & 2 & -1 \\
-1 & 2 & -1 \\
\vdots & \vdots & \vdots \\
-1 & 2 & -1
\end{bmatrix}$$
**$\ell_1$-norm heuristic**

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

- $\gamma > 0$ is parameter used to achieve desired sparsity
  (when $\text{card}$ 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

\[
\begin{align*}
\text{minimize} & \quad \text{card}(x) \\
\text{subject to} & \quad x \in C \\
\end{align*}
\]

(C convex)

• apply heuristic to get (easy) \( \ell_1 \)-norm minimization problem

\[
\begin{align*}
\text{minimize} & \quad \|x\|_1 \\
\text{subject to} & \quad x \in C \\
\end{align*}
\]
**Example: Cardinality constrained problem**

• start with (hard) cardinality constrained problem \((f, C \text{ convex})\)
  \[
  \begin{align*}
  &\text{minimize} & f(x) \\
  &\text{subject to} & x \in C, \quad \text{card}(x) \leq k
  \end{align*}
  \]

• apply heuristic to get (easy) \(\ell_1\)-constrained problem
  \[
  \begin{align*}
  &\text{minimize} & f(x) \\
  &\text{subject to} & x \in C, \quad \|x\|_1 \leq \beta
  \end{align*}
  \]
  or \(\ell_1\)-regularized problem
  \[
  \begin{align*}
  &\text{minimize} & f(x) + \gamma\|x\|_1 \\
  &\text{subject to} & x \in C
  \end{align*}
  \]
  \(\beta, \gamma\) adjusted so that \(\text{card}(x) \leq k\)
Polishing

- use $\ell_1$ 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

\[
\begin{align*}
\text{minimize} & \quad \text{card}(x) \\
\text{subject to} & \quad x \in C, \quad \|x\|_\infty \leq R
\end{align*}
\]

• equivalent to mixed Boolean convex problem

\[
\begin{align*}
\text{minimize} & \quad 1^T z \\
\text{subject to} & \quad |x_i| \leq Rz_i, \quad i = 1, \ldots, n \\
& \quad x \in C, \quad z_i \in \{0, 1\}, \quad i = 1, \ldots, n
\end{align*}
\]

with variables \(x, z\)
• now relax $z_i \in \{0, 1\}$ to $z_i \in [0, 1]$ to obtain

\[
\begin{align*}
\text{minimize} & \quad 1^T z \\
\text{subject to} & \quad |x_i| \leq R z_i, \quad i = 1, \ldots, n \\
& \quad x \in C \\
& \quad 0 \leq z_i \leq 1, \quad i = 1, \ldots, n
\end{align*}
\]

which is equivalent to

\[
\begin{align*}
\text{minimize} & \quad (1/R) \|x\|_1 \\
\text{subject to} & \quad x \in C
\end{align*}
\]

the $\ell_1$ heuristic

• optimal value of this problem is lower bound on original problem
**Interpretation via convex envelope**

- Convex envelope $f_{\text{env}}$ of a function $f$ on set $C$ is the largest convex function that is an underestimator of $f$ on $C$.

- $\text{epi}(f_{\text{env}}) = \text{Co}(\text{epi}(f))$.

- $f_{\text{env}} = (f^*)^*$ (with some technical conditions).

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

- For $x \in \mathbb{R}^n$, $(1/R)\|x\|_1$ is convex envelope of $\text{card}(x)$ on $\{z \mid \|z\|_\infty \leq R\}$. 

---

$\ell_1$-norm methods for convex-cardinality problems 23
Weighted and asymmetric $\ell_1$ heuristics

- minimize $\text{card}(x)$ over convex set $C$
- suppose we know lower and upper bounds on $x_i$ over $C$
  \[ x \in C \implies l_i \leq x_i \leq u_i \]

(best values for these can be found by solving $2n$ convex problems)
- if $u_i < 0$ or $l_i > 0$, then $\text{card}(x_i) = 1$ (i.e., $x_i \neq 0$) for all $x \in C$
- 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 $\text{card}(x)$
Regressor selection

minimize $\|Ax - b\|_2$
subject to $\text{card}(x) \leq k$

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

$\ell_1$-norm methods for convex-cardinality problems
Example (6.4 in BV book)

- $A \in \mathbb{R}^{10 \times 20}$, $x \in \mathbb{R}^{20}$, $b \in \mathbb{R}^{10}$
- dashed curve: exact optimal (via enumeration)
- solid curve: $\ell_1$ heuristic with polishing

$\ell_1$-norm methods for convex-cardinality problems
Sparse signal reconstruction

- convex-cardinality problem:

  \[
  \text{minimize} \quad \|Ax - y\|_2 \\
  \text{subject to} \quad \text{card}(x) \leq k
  \]

- \(\ell_1\) heuristic:

  \[
  \text{minimize} \quad \|Ax - y\|_2 \\
  \text{subject to} \quad \|x\|_1 \leq \beta
  \]

  (called LASSO)

- another form: minimize \(\|Ax - y\|_2 + \gamma\|x\|_1\)

  (called basis pursuit denoising)
Example

• signal $x \in \mathbb{R}^n$ with $n = 1000$, 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: $\ell_1$ reconstruction with $\gamma = 10^{-3}$
- $\ell_2$ reconstruction; minimizes $\|Ax - y\|_2 + \gamma\|x\|_2$, where $\gamma = 10^{-3}$
- *left*: original; *right*: $\ell_2$ reconstruction
Some recent theoretical results

• suppose $y = Ax$, $A \in \mathbb{R}^{m \times n}$, $\text{card}(x) \leq k$

• 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 $\ell_1$ heuristic (minimizing $\|x\|_1$ subject to $Ax = y$) reconstruct $x$ (exactly)?
recent results by Candès, Donoho, Romberg, Tao, . . .

- (for some choices of $A$) if $m \geq (C \log n)k$, $\ell_1$ 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

$\ell_1$-norm methods for convex-cardinality problems
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 (§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 $\gamma$

$\ell_1$-norm methods for convex-cardinality problems
$\ell_2$ reconstruction

for three values of $\gamma$

\[ \hat{x} \]

$\ell_1$-norm methods for convex-cardinality problems
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

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

$\ell_1$-norm methods for convex-cardinality problems
... this is what you’d expect with $8 \times$ more variables than measurements
Iterated weighted $\ell_1$ heuristic

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

\[ w := 1 \]

repeat

\[ \text{minimize} \| \text{diag}(w)x \|_1 \text{ over } x \in C \]

\[ w_i := 1/(\epsilon + |x_i|) \]

- 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

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

• we’ll find a local solution by linearizing objective at current point,

$$\begin{aligned}
\sum_{i=1}^{n} \log(1 + x_i/\epsilon) & \approx \sum_{i=1}^{n} \log(1 + x_i^{(k)}/\epsilon) + \sum_{i=1}^{n} \frac{x_i - x_i^{(k)}}{\epsilon + x_i^{(k)}}
\end{aligned}$$
and solving resulting convex problem

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

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 $\text{card}(x) = 32$)

$\ell_1$-norm methods for convex-cardinality problems
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

\[ y(t) \]

\[ a(t) \quad b(t) \]

$\ell_1$-norm methods for convex-cardinality problems
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 $\Sigma \in S^n_+$, 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 $D, X \in S^n$

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

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

• 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

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

- \( \Sigma \) 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

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

\[ \text{Rank}(X) \] vs. \[ \beta \]

\[ \lambda_i(X) \] vs. \[ \beta \]

\[ \ell_1 \]-norm methods for convex-cardinality problems
• for $\beta = 0.1357$ (knee of the tradeoff curve) we find

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

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

• convex-cardinality (and rank) problems arise in many applications

• these problems are hard (to solve exactly, in general)

• heuristics based on $\ell_1$ 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, ... )