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

- problems involving cardinality
- the $\ell_1$-norm heuristic
- convex relaxation and convex envelope interpretations
- examples
- recent results
$\ell_1$-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

- new 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 \( x, y \geq 0 \)

• 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:
  \[
  \begin{align*}
  &\text{minimize} & c^T x \\
  &\text{subject to} & Ax \preceq b, & x_i \in \{0, 1\}
  \end{align*}
\]
includes many famous (hard) problems, e.g., 3-SAT, traveling salesman

- can be expressed as

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

  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

\[
\text{minimize } \quad \text{card}(x) \\
\text{subject to } \quad x \in C
\]

• 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 $\hat{x}_{\text{ml}}$ is solution of

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

\[
\begin{align*}
\text{minimize} \quad & \sum_{i \notin \mathcal{B}} (y_i - a_i^T x)^2 \\
\text{subject to} \quad & |\mathcal{B}| \leq k
\end{align*}
\]

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 \( g(\lambda) = \inf_x (\sum_{i=1}^{m} \lambda_i f_i(x)) \geq 1, \lambda \succeq 0 \)
- 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 $\geq R_{\min}$:

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

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

• as convex-cardinality problem:

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

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

- as convex-cardinality problem:

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

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

  \( (C \text{ convex}) \)

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

  minimize \( \|x\|_1 \)
  subject to \( x \in C \)
Example: Cardinality constrained problem

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

- apply heuristic to get (easy) \(\ell_1\)-constrained problem
  
  \[
  \begin{align*}
  \text{minimize} & \quad f(x) \\
  \text{subject to} & \quad x \in C, \quad \|x\|_1 \leq \beta
  \end{align*}
  \]

  or \(\ell_1\)-regularized problem
  
  \[
  \begin{align*}
  \text{minimize} & \quad f(x) + \gamma\|x\|_1 \\
  \text{subject to} & \quad x \in C
  \end{align*}
  \]

  \(\beta, \gamma\) 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

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

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

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

• 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 $\|Ax - b\|_2$
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
Sparse signal reconstruction

• convex-cardinality problem:

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

• \(\ell_1\) heuristic:

\[
\text{minimize } \|Ax - y\|_2 \\
\text{subject to } \|x\|_1 \leq \beta
\]
(called LASSO)

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

• signal \( x \in \mathbb{R}^n \) with \( n = 1000 \), \( \text{card}(x) = 30 \)
• \( m = 200 \) (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