Technische
Universität
Braunschweig

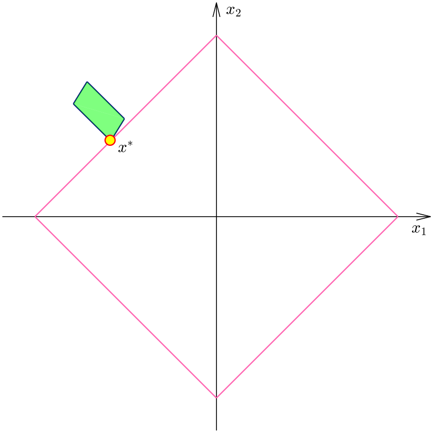
ISMP2018
BORDEAUX
23rd International Symposium on Mathematical Programming | July 1–6

A primal-dual homotopy algorithm for sparse recovery with infinity norm constraints

Christoph Brauer, Dirk Lorenz and Andreas Tillmann

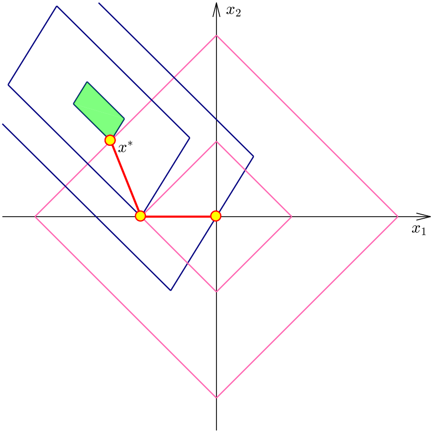
Introduction

$$\min_{x \in \mathbb{R}^n} \quad \|x\|_1$$
$$\text{s.t.} \quad \|Ax - b\|_\infty \leq \delta^\dagger$$



Introduction

$$\min_{x \in \mathbb{R}^n} \quad \|x\|_1$$
$$\text{s.t.} \quad \|Ax - b\|_\infty \leq \delta^\dagger$$



Homotopy path:

$$\mathcal{P} := \{x^*(\delta) \mid \delta \in [\delta_{\min}, \infty)\}$$

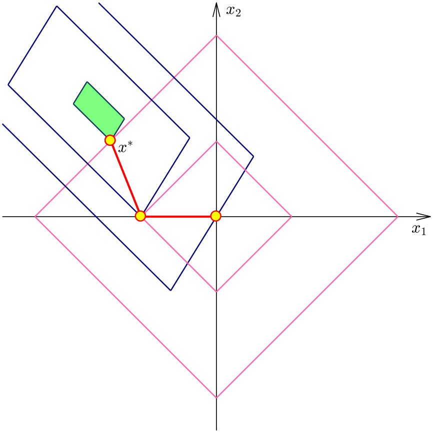
Introduction

$$\min_{x \in \mathbb{R}^n} \quad \|x\|_1 \qquad \text{s.t.} \quad \|Ax - b\|_\infty \le \delta^\dagger$$

x^* is an optimal solution

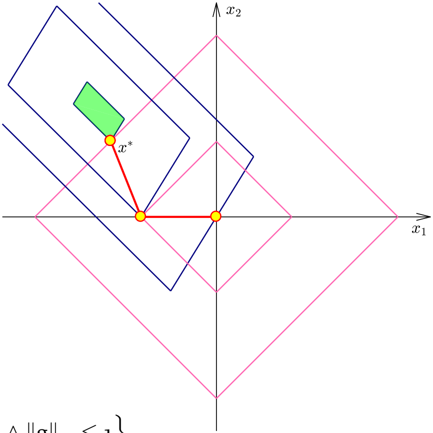
$$\Updownarrow$$

$$\exists y^* \in \mathbb{R}^m : \quad -A^\top y^* \in \operatorname{Sign}(x^*), \quad Ax^* - b \in \delta \operatorname{Sign}(y^*)$$



Introduction

$$\min_{x \in \mathbb{R}^n} \quad \|x\|_1$$
$$\text{s.t.} \quad \|Ax - b\|_\infty \leq \delta^\dagger$$



x^* is an optimal solution

$$\Updownarrow$$

$$\exists y^* \in \mathbb{R}^m : \quad -A^\top y^* \in \operatorname{Sign}(x^*)$$
$$Ax^* - b \in \delta \operatorname{Sign}(y^*)$$

$$\operatorname{Sign}(x) = \left\{ g \in \mathbb{R}^n : g_{\operatorname{supp}(x)} = \operatorname{sign}\left(x_{\operatorname{supp}(x)}\right) \wedge \|g\|_\infty \leq 1 \right\}$$

Basic idea

Solve a sequence of problems with

$$\delta^0 > \delta^1 > \cdots > \delta^{K-1} > \delta^K = \delta$$

and optimal pairs

$$(x^0, y^0),\ (x^1, y^1),\ \ldots,\ (x^{K-1}, y^{K-1}),\ (x^K, y^K) = (x^*, y^*).$$

Motivation:

1. Transitions $(x^k, y^k) \to (x^{k+1}, y^{k+1})$ are easy.
2. $(x^0, y^0) = (0, 0)$ is optimal for $\delta^0 \geq \|b\|_\infty$.

Transitions

1. Dual update:

$$\begin{aligned} y^{k+1} \in \underset{y \in \mathbb{R}^m}{\arg\min} \quad & \psi^\top y \\ \text{s.t.} \quad & -A^\top y \in \operatorname{Sign}(x^k) \\ & Ax^k - b \in \delta^k \operatorname{Sign}(y) \end{aligned}$$

2. Primal update:

$$\begin{aligned} (x^{k+1}, t^{k+1}) \in \underset{(x,t) \in \mathbb{R}^n \times \mathbb{R}}{\arg\max} \quad & t \\ \text{s.t.} \quad & -A^\top y^{k+1} \in \operatorname{Sign}(x) \\ & Ax - b \in (\delta^k - t) \operatorname{Sign}(y^{k+1}) \end{aligned}$$

3. Parameter update:

$$\delta^{k+1} := \delta^k - t^{k+1}$$

Support and active set

- $S := \{j : x_j \neq 0\}$
  (primal support)
- $W := \{i : |a_i^\top x - b_i| = \delta\}$
  (primal active set)
- $\Sigma := \{j : |A_j^\top y| = 1\}$
  (dual active set)
- $\Omega := \{i : y_i \neq 0\}$
  (dual support)
- $\mathrm{Sign}(x) = \{g \in [-1, 1]^n : g_S = \mathrm{sign}\,(x_S)\}$

Dual update

LP with $|W|$ variables and $2n - |S|$ constraints:

$$
\begin{aligned}
y^{k+1} \in \operatorname*{arg\,min}_{y \in \mathbb{R}^m} \quad & \psi^\top y \\
\text{s.t.} \quad & -A_S^\top y = \operatorname{sign}(x_S^k) \\
& -\mathbb{1} \leq -A_{S^c}^\top y \leq \mathbb{1} \\
& -\operatorname{sign}(A^W x^k - b_W) \odot y_W \leq 0 \\
& y_{W^c} = 0
\end{aligned}
$$

Question: How must we choose ψ?

Not good: Many constraints!

Primal update

LP with $|\Sigma|$ variables and $2m - |\Omega| + 1$ constraints:

$$\begin{aligned}
x^{k+1} \in \underset{(x,t)\in\mathbb{R}^n\times\mathbb{R}}{\arg\max} \quad & t \\
\text{s.t.} \quad & A^{\Omega}x - b_{\Omega} = (\delta^k - t)\operatorname{sign}(y_{\Omega}^{k+1}) \\
& -(\delta^k - t)\mathbb{1} \leq A^{\Omega^c}x - b_{\Omega^c} \leq (\delta^k - t)\mathbb{1} \\
& A_{\Sigma}^{\top}y^{k+1} \odot x_{\Sigma} \leq 0 \\
& x_{\Sigma^c} = 0 \\
& t \leq \delta^k - \delta
\end{aligned}$$

Not good: Many constraints!

Theorem of the alternative

y^k is optimal in the dual update
with x^{k-1} and δ^{k-1} fixed

$\Updownarrow$

there exists no feasible descent
direction w.r.t. ψ at y^k

$(x^{k-1}, 0)$ is not optimal in the primal update
with y^k fixed

$\Updownarrow$

there exists a feasible ascent
direction w.r.t. t at $(x^{k-1}, 0)$

Theorem of the alternative

y^k is optimal in the dual update
with x^{k-1} and δ^{k-1} fixed

$\Updownarrow$

there exists no feasible descent direction w.r.t. ψ at y^k

$\Longleftrightarrow$

there exists a feasible ascent direction w.r.t. t at $(x^{k-1}, 0)$

$\Updownarrow$

$(x^{k-1}, 0)$ is not optimal in the primal update
with y^k fixed

$\uparrow$

Farkas' Lemma
with $\boldsymbol{\psi} = -\mathbf{sign}(\mathbf{A}\mathbf{x}^{k-1} - \mathbf{b})$

We find $x^k \neq x^{k-1}$ and $t^k > 0$ in the primal update.

Theorem of the alternative (2)

y^k is not optimal in the dual update with x^k and δ^k fixed

$\Updownarrow$

there exists a feasible descent direction w.r.t. ψ at y^k

$(x^k, 0)$ is optimal in the primal update with y^k fixed

$\Updownarrow$

there exists no feasible ascent direction w.r.t. t at $(x^k, 0)$

Theorem of the alternative (2)

y^k is <u>not</u> optimal in the dual update
with x^k and δ^k fixed

$\Updownarrow$

there exists a feasible descent
direction w.r.t. ψ at y^k

$\Longleftrightarrow$

$(x^k, 0)$ is optimal in the primal update
with y^k fixed

$\Updownarrow$

there exists <u>no</u> feasible ascent
direction w.r.t. t at $(x^k, 0)$

$\uparrow$

Farkas' Lemma

with $\boldsymbol{\psi} = -\mathbf{sign}(\mathbf{A}\mathbf{x}^k - \mathbf{b})$

We find $y^{k+1} \neq y^k$ in the dual update.

ℓ_1-HOUDINI HOmotopy UnDer Infinity Norm constraInts

Input: $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $0 \leq \delta < \|b\|_\infty$

$\delta^0 \leftarrow \|b\|_\infty$
$x^0 \leftarrow 0$
$S_0 \leftarrow \emptyset$
$W_0 \leftarrow \{i : |b_i| = \delta^0\}$
$k \leftarrow 0$

repeat

$y^{k+1} \leftarrow \mathtt{dual\_lp}(x^k, S_k, W_k)$
$\Omega_{k+1} \leftarrow \{i : y_i^{k+1} \neq 0\}$
$\Sigma_{k+1} \leftarrow \{j : |A_j^\top y^{k+1}| = 1\}$
$[x^{k+1}, t^{k+1}] \leftarrow \mathtt{primal\_lp}(y^{k+1}, \Sigma_{k+1}, \Omega_{k+1})$
$\delta^{k+1} \leftarrow \delta^k - t^{k+1}$
$S_{k+1} \leftarrow \{j : x_j^{k+1} \neq 0\}$
$W_{k+1} \leftarrow \{i : |a_i^\top x^{k+1} - b_i| = \delta^{k+1}\}$
$k \leftarrow k + 1$

until $\delta^k = \delta$ or $t^k = 0$

return $\{x^0, \ldots, x^k\}$ and $\{\delta^0, \ldots, \delta^k\}$

Finite termination

Theorem (B., Lorenz and Tillmann 2018)

ℓ_1-HOUDINI *returns an optimal solution after finitely many iterations.*

Proof idea.

Use the above Theorem of the alternatives to show that each combination of support S, active set W and associated sign patterns $\operatorname{sign}(x_S)$ and $\operatorname{sign}(A^W x - b_W)$ can only occur once among all iterates of ℓ_1-HOUDINI. □

Upper Bound

Theorem (B. 2018)

The number of iterations in ℓ_1-HOUDINI is bounded above by $(3^{m+n}+1)/2$.

Proof idea.

Show that the same combination of support S and active set W cannot occur in combination with opposing sign patterns $\operatorname{sign}(x_S^k) = -\operatorname{sign}(x_S^\ell)$ and $\operatorname{sign}(A^W x^k - b_W) = -\operatorname{sign}(A^W x^\ell - b_W)$. □

Worst case

Theorem (B. 2018)

In the worst case, ℓ_1-HOUDINI has to perform at least $(3^n + 1)/2$ iterations.

Proof idea (similar to Mairal 2012).

For arbitrary $n \in \mathbb{N}$, construct $A^{(n)} \in \mathbb{R}^{n\times n}$ and $b^{(n)} \in \mathbb{R}^n$ recursively:

$$A^{(n)} := \begin{bmatrix} A^{(n-1)} & 2\alpha_n b^{(n-1)} \\ 0 & \alpha_n b_n \end{bmatrix}, \quad b^{(n)} := \begin{pmatrix} b^{(n-1)} \\ b_n \end{pmatrix}, \quad A^{(1)} := \alpha_1 \in \mathbb{R}_+, \quad b^{(1)} := b_1 \in \mathbb{R}_+ .$$

Under appropriate conditions on α_n and b_n, it holds that $K^{(n)} = 3K^{(n-1)} - 1$ for the respective numbers of iterations.

If the statement is true for dimension $n - 1$, then $K^{(n)} = 3 \cdot \frac{3^{n-1}+1}{2} - 1.$ □

Practical aspects

- Linear programs for primal and dual updates can be warm-started with x^k and y^k, and solved efficiently using a dedicated active-set strategy.
- Need $|\mathbb{W}|$ equations in $|S|$ variables for an ascent direction in the primal update, and $|\Sigma|$ equations in $|\Omega|$ variables to compute a descent direction in the dual update.
- Box constraints $\alpha \leq Ax - b \leq \beta$ can be handled as well.
- Modification for problems with arbitrary linear constraints is possible.
- Solution path can be used for the purpose of cross-validation.

Chebyshew estimation

- $S = \{(b_i, a_i)\}_{i=1}^{m} \subseteq \mathbb{R} \times \mathbb{R}^n$ samples
- $b = Ax + \eta \in \mathbb{R}^m$ linear model
- $\eta_i \sim \mathcal{U}([-\delta^\dagger, \delta^\dagger])$ i.i.d. noise
- $\delta^\dagger > 0$ unknown

Goal: Find a sparse linear predictor $\hat{x}$.

If $\delta^\dagger$ was known a priori,

$$\hat{x}^\dagger \in \operatorname*{arg\,min}_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{s.t. } \|Ax - b\|_\infty \leq \delta^\dagger$$

would be a standard approach. We need to do something else!

Cross-validation

- $S = S_1 \cup \cdots \cup S_K$
- $I_k = \{i \mid (b_i, a_i) \in S_k\}$
- $x_k(\delta)$ homotopy path for

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{s.t. } \|A^{I_k^c} x - b_{I_k^c}\|_\infty \leq 0$$

- cross-validation error

$$\varepsilon(\delta) := \frac{1}{K} \sum_{k=1}^{K} \|A^{I_k} x_k(\delta) - b_{I_k}\|_\infty$$

- cross-validation parameter

$$\delta_{cv} := \operatorname*{arg\,min}_{\delta \in [\delta_{min}, \infty)} \varepsilon(\delta)$$

Cross-validation vs. true parameter

Final predictor:

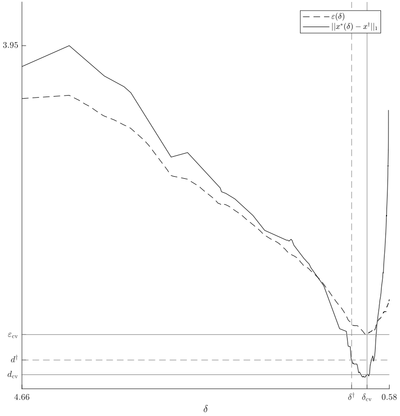
$$\hat{x}_{cv} \in \arg\min_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{s.t.} \quad \|Ax - b\|_\infty \le \delta_{cv}$$

Distances to ground truth:

$$\|\hat{x}_{cv} - x^\dagger\|_1 \approx 0.5983$$
$$\|\hat{x}^\dagger - x^\dagger\|_1 \approx 0.7480$$

Smallest distance to ground truth:

$$\|\hat{x}^* - x^\dagger\|_1 \approx 0.5728$$



More applications

$$\min_{x \in \mathbb{R}^n} \quad \|x\|_1$$
$$\text{s.t.} \quad \|Ax - b\|_\infty \leq \delta$$

- Sparse dequantization
- Sparse linear discriminant analysis
- Sparse precision matrix estimation
- Dantzig selector

More applications

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{s.t.} \quad \|Ax - b\|_\infty \leq \delta$$

$$\min_{a \in \mathbb{R}^n} \|a\|_1 \quad \text{s.t.} \quad \|\Psi a - q\|_\infty \leq \frac{\Delta}{2}$$

- Sparse dequantization
- Sparse linear discriminant analysis
- Sparse precision matrix estimation
- Dantzig selector

More applications

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{s.t.} \quad \|Ax - b\|_\infty \leq \delta$$

$$\min_{\beta \in \mathbb{R}^p} \|\beta\|_1 \quad \text{s.t.} \quad \|\hat{\Sigma}\beta - (\bar{X} - \bar{Y})\|_\infty \leq \lambda$$

- Sparse dequantization
- Sparse linear discriminant analysis
- Sparse precision matrix estimation
- Dantzig selector

More applications

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{s.t.} \quad \|Ax - b\|_\infty \leq \delta$$

$$\min_{\beta \in \mathbb{R}^p} \|\beta\|_1 \quad \text{s.t.} \quad \|\hat{\Sigma}\beta - e_i\|_\infty \leq \lambda$$

- Sparse dequantization
- Sparse linear discriminant analysis
- Sparse precision matrix estimation
- Dantzig selector

More applications

$$\min_{x\in\mathbb{R}^n} \|x\|_1 \quad \text{s.t.} \quad \|Ax-b\|_\infty \le \delta$$

$$\min_{\beta\in\mathbb{R}^p} \|\beta\|_1 \quad \text{s.t.} \quad \|X^\top(X\beta - Y)\|_\infty \le \lambda$$

- Sparse dequantization
- Sparse linear discriminant analysis
- Sparse precision matrix estimation
- Dantzig selector

Thank you!

Christoph Brauer, Dirk Lorenz and Andreas Tillmann. **A Primal-Dual Homotopy Algorithm for ℓ_1-Minimization with ℓ_∞-Constraints.** *Computational Optimization and Applications, February 2018.*

Christoph Brauer. **Homotopy Methods for Linear Optimization Problems with Sparsity Penalty and Applications.** *PhD thesis, TU Braunschweig, March 2018.*

`https://github.com/chrbraue/l1Houdini`