Technische Universität Braunschweig

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

Christoph Brauer, Dirk Lorenz and Andreas Tillmann

Introduction

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

- Sparse dequantization [B., Gerkmann, and Lorenz, 2016]
- Sparse linear discriminant analysis [Cai and Liu, 2011]
- Sparse precision matrix estimation [Cai, Liu, and Luo, 2011]
- Chebyshev estimation [Stiefel, 1959, Appa and Smith, 1973]
- Dantzig selector [Candès and Tao, 2007]

Introduction

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

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

- Sparse dequantization [B., Gerkmann, and Lorenz, 2016]
- Sparse linear discriminant analysis [Cai and Liu, 2011]
- Sparse precision matrix estimation [Cai, Liu, and Luo, 2011]
- Chebyshev estimation [Stiefel, 1959, Appa and Smith, 1973]
- Dantzig selector [Candès and Tao, 2007]

Introduction

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

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

- Sparse dequantization [B., Gerkmann, and Lorenz, 2016]
- Sparse linear discriminant analysis [Cai and Liu, 2011]
- Sparse precision matrix estimation [Cai, Liu, and Luo, 2011]
- Chebyshev estimation [Stiefel, 1959, Appa and Smith, 1973]
- Dantzig selector [Candès and Tao, 2007]

Introduction

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

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

- Sparse dequantization [B., Gerkmann, and Lorenz, 2016]
- Sparse linear discriminant analysis [Cai and Liu, 2011]
- Sparse precision matrix estimation [Cai, Liu, and Luo, 2011]
- Chebyshev estimation [Stiefel, 1959, Appa and Smith, 1973]
- Dantzig selector [Candès and Tao, 2007]

Introduction

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

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

- Sparse dequantization [B., Gerkmann, and Lorenz, 2016]
- Sparse linear discriminant analysis [Cai and Liu, 2011]
- Sparse precision matrix estimation [Cai, Liu, and Luo, 2011]
- Chebyshev estimation [Stiefel, 1959, Appa and Smith, 1973]
- Dantzig selector [Candès and Tao, 2007]

Introduction

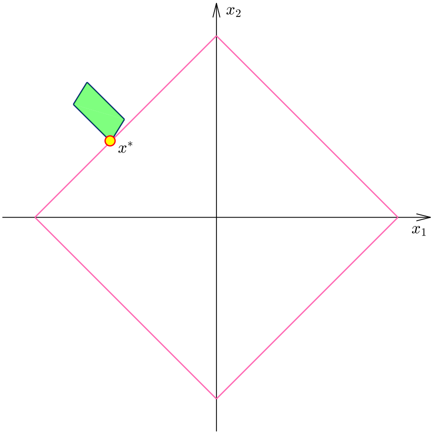
$$\min_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{s.t.} \quad \|Ax - b\|_\infty \leq \delta$$

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

- Sparse dequantization [B., Gerkmann, and Lorenz, 2016]
- Sparse linear discriminant analysis [Cai and Liu, 2011]
- Sparse precision matrix estimation [Cai, Liu, and Luo, 2011]
- Chebyshev estimation [Stiefel, 1959, Appa and Smith, 1973]
- Dantzig selector [Candès and Tao, 2007]

Introduction

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

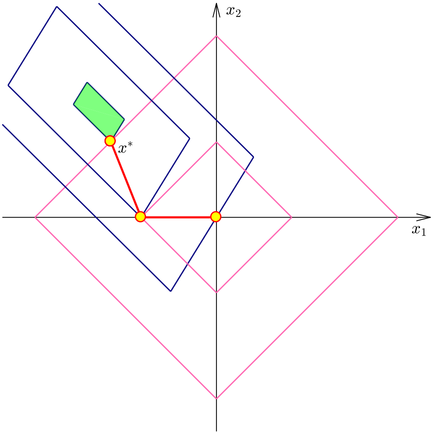


Introduction

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

$x^\star : [\delta_{\min}, \infty) \to \mathbb{R}^n$
(solution mapping)

$x^\star\left([\delta_{\min}, \infty)\right)$
(solution path)



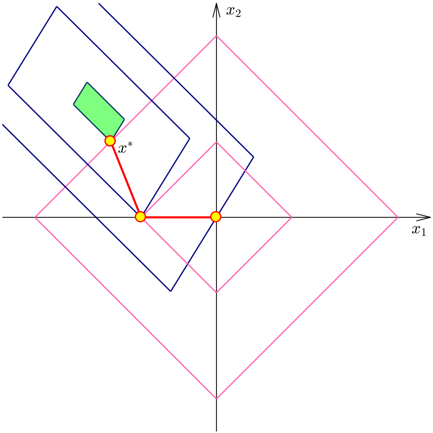
Introduction

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

$(x^\star, y^\star)$ *optimal pair*

$$\Updownarrow$$

$$-A^\top y^\star \in \partial\|x^\star\|_1$$
$$Ax^\star - b \in \delta\, \partial\|y^\star\|_1$$



Outline

- **Introduction**
- **Homotopy method**
- **Theoretical results**
- **Practical aspects**

Basic idea

Solve a sequence of problems with

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

$$\begin{aligned} -A^\top y^\star &\in \partial\|x^\star\|_1 \\ Ax^\star - b &\in \delta\,\partial\|y^\star\|_1 \end{aligned}$$

and optimal pairs

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

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. Fix x^k and δ^k and search y^{k+1} such that:

$$\begin{aligned} -A^\top y^{k+1} &\in \partial\|x^k\|_1 \\ Ax^k - b &\in \delta^k\, \partial\|y^{k+1}\|_1 \end{aligned}$$

2. Fix y^{k+1} and search x^{k+1} and t^{k+1} such that:

$$\begin{aligned} -A^\top y^{k+1} &\in \partial\|x^{k+1}\|_1 \\ Ax^{k+1} - b &\in (\delta^k - t^{k+1})\, \partial\|y^{k+1}\|_1 \end{aligned}$$

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

Subdifferential

$S := \{j : x_j \neq 0\}$
(*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*)

$$\begin{aligned} -A^\top y^\star &\in \partial\|x^\star\|_1 \\ Ax^\star - b &\in \delta\,\partial\|y^\star\|_1 \end{aligned}$$

$$\boxed{\partial\|x\|_1 = \{g \in [-1,1]^n : g_S = \mathrm{sign}(x_S)\}}$$

Dual update

LP with $|\mathbb{W}|$ variables and $2n - |S|$ constraints:

$$
\begin{aligned}
-A^\top y^{k+1} &\in \partial \|x^k\|_1 \\
Ax^k - b &\in \delta^k \, \partial \|y^{k+1}\|_1
\end{aligned}
$$

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

Question: How must we choose ψ?

Primal update

$$-A^\top y^{k+1} \in \partial \|x^{k+1}\|_1$$
$$Ax^{k+1} - b \in (\delta^k - t^{k+1})\, \partial \|y^{k+1}\|_1$$

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

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

Theorem of the alternative

y^{k+1} is optimal in the dual update

$\Updownarrow$

there exists no feasible descent direction w.r.t. ψ at y^{k+1}

$(x^k, 0)$ is not optimal in the primal update

$\Updownarrow$

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

Theorem of the alternative

y^{k+1} is optimal in the dual update

$\Updownarrow$

there exists no feasible descent direction w.r.t. ψ at y^{k+1}

$\Longleftrightarrow$

$(x^k, 0)$ is not optimal in the primal update

$\Updownarrow$

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

$\uparrow$

Farkas' Lemma

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

Theorem of the alternative (2)

y^{k+1} is not optimal in the dual update

$\Updownarrow$

there exists a feasible descent direction w.r.t. ψ at y^{k+1}

$\Longleftrightarrow$

$(x^{k+1}, 0)$ is optimal in the primal update

$\Updownarrow$

there exists no feasible ascent direction w.r.t. t at $(x^{k+1}, 0)$

$\uparrow$

Farkas' Lemma

mit $\psi = -\mathbf{sign}(\mathbf{Ax}^{k+1} - \mathbf{b})$

ℓ_1-Houdini HOmotopy UnDer Infinity Norm constraInts

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

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

repeat

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

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

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

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 $\mathrm{sign}(x_S)$ and $\mathrm{sign}(A^W x - b_W)$ 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 (Mairal and Yu, 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.

$m \times n$	δ	$\|S\|$	$\|W\|$	ℓ_1-Houdini act.-set	ℓ_1-Houdini Gurobi	Gurobi standal.
512 × 1024	4.09	34	512	0.98	2.58	0.46
			72	0.53	2.64	0.46
512 × 1024	4.54	51	512	1.80	103.35	1.26
			96	1.22	–	1.09
512 × 1536	0.72	14	512	0.24	3.60	0.83
			31	0.24	3.77	0.81
512 × 1536	4.58	22	512	0.42	15.92	1.63
			43	0.29	10.40	1.53
512 × 2048	3.20	51	512	5.86	–	1.13
			141	3.79	–	0.98
512 × 2048	0.58	20	512	0.79	–	1.93
			45	0.44	16.13	1.42
512 × 4096	1.47	10	512	0.20	18.36	1.26
			38	0.16	1.06	1.22
512 × 2048	2.78	10	512	0.14	8.32	1.25
			32	0.09	0.86	1.21
1024 × 2048	4.79	84	1024	0.77	2.13	0.07
			148	0.82	2.02	0.07
1024 × 2048	4.83	27	1024	1.91	–	3.51
			55	0.80	38.83	2.77
1024 × 3072	0.87	18	1024	0.91	19.65	3.36
			47	0.76	17.49	3.42
1024 × 3072	4.86	99	1024	26.57	–	1.79
			234	16.46	–	1.59
1024 × 4096	4.79	97	1024	36.53	–	2.99
			245	27.36	437.62	2.69
1024 × 4096	2.48	26	1024	2.47	–	6.85
			60	1.33	50.41	3.99
1024 × 8192	4.02	20	1024	1.41	23.21	5.34
			64	1.34	20.67	5.29
1024 × 8192	0.80	9	1024	0.82	–	5.12
			43	0.42	–	5.27

References

Gautam Appa and Cyril Smith. On L_1 and Chebyshev estimation. *Mathematical Programming*, 5(1):73–87, 1973.

C. B. The Homotopy Path of ℓ_∞-Constrained ℓ_1-Minimization: Complexity and Application to Model Selection. Under review, February 2018.

C. B., Timo Gerkmann, and Dirk Lorenz. Sparse Reconstruction of Quantized Speech Signals. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5940–5944. IEEE, 2016.

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

Tony Cai and Weidong Liu. A Direct Estimation Approach to Sparse Linear Discriminant Analysis. *Journal of the American Statistical Association*, 106(496):1566–1577, December 2011.

Tony Cai, Weidong Liu, and Xi Luo. A Constrained ℓ_1 Minimization Approach to Sparse Precision Matrix Estimation. *Journal of the American Statistical Association*, 106(494):594–607, June 2011.

Emmanuel Candès and Terence Tao. The Dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics*, 35(6):2313–2351, 2007.

Julien Mairal and Bin Yu. Complexity Analysis of the Lasso Regularization Path. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pages 353–360, July 2012.

Eduard Stiefel. Über diskrete und lineare Tschebyscheff-Approximationen. *Numerische Mathematik*, 1(1):1–28, 1959.

https://github.com/chrbraue/l1Houdini