# $\ell_1$-Houdini: A New Homotopy Method for $\ell_1$-Minimization

Christoph Brauer* Dirk Lorenz* Andreas Tillmann†

* TU Braunschweig, Germany
† RWTH Aachen University, Germany

## Problem and Optimality Conditions

- Given $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and $\delta \geq 0$, we consider the problem

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

- It is well-known that $x^\star$ is an optimal solution of $(\mathrm{P}_\delta)$ if and only if there exists a $y^\star$ such that

$$-A^\top y^\star \in \partial\|x^\star\|_1 \quad \text{and} \quad Ax^\star - b \in \delta\partial\|y^\star\|_1. \qquad (1)$$

- Each such $y^\star$ is by construction an optimal solution to the dual problem of $(\mathrm{P}_\delta)$, which is

$$\max_{y\in\mathbb{R}^m} -b^\top y - \delta\|y\|_1 \quad \text{s.t.} \ \|A^\top y\|_\infty \leq 1. \qquad (\mathrm{D}_\delta)$$

## Partitioned Optimality Conditions

- For a thorough understanding of the conditions (1), we define

$$S := \{j : x_j^\star \neq 0\}, \qquad W := \{i : |a_i^\top x^\star - b_i| = \delta\},$$
(*primal support*) (*primal active set*)

$$\Sigma := \{j : |A_j^\top y^\star| = 1\}, \qquad \Omega := \{i : y_i^\star \neq 0\}.$$
(*dual active set*) (*dual support*)

- The optimality conditions (1) are then equivalent to

$$\begin{array}{rlrl} -A_S^\top y^\star &= \operatorname{sign}(x_S^\star) & A^\Omega x^\star - b_\Omega &= \delta\operatorname{sign}(y_\Omega^\star) \\ -\mathbb{1} \leq -A_{S^c}^\top y^\star &\leq \mathbb{1} & -\delta\mathbb{1} \leq A^{\Omega^c}x^\star - b_{\Omega^c} &\leq \delta\mathbb{1} \\ y_{W^c}^\star &= 0 & x_{\Sigma^c}^\star &= 0 \end{array}. \qquad (2)$$

## Basic Idea

- We solve a sequence of problems $(\mathrm{P}_{\delta^k})_{k=0,\dots,K}$ with

$$\|b\|_\infty = \delta_0 > \delta_1 > \cdots > \delta_K = \delta.$$

- The starting point $(x^0, y^0) = (0, 0)$ is an *optimal pair* for $(\mathrm{P}_{\delta^0})$.
- The transition from an optimal pair $(x^k, y^k)$ for $(\mathrm{P}_{\delta^k})$ to an optimal pair $(x^{k+1}, y^{k+1})$ for $(\mathrm{P}_{\delta^{k+1}})$ can be done in two steps:

$\underline{U_D}$ : Fix $x^k$ and $\delta^k$ in (2) and search an appropriate $y^{k+1} \neq y^k$ such that the conditions stay valid at $(x^k, y^{k+1})$ and $\delta^k$.

$\underline{U_P}$ : Fix $y^{k+1}$ in (2) and search $x^{k+1} \neq x^k$ and $\delta^{k+1} < \delta^k$ such that the conditions stay satisfied at $(x^{k+1}, y^{k+1})$ and $\delta^{k+1}$.

## Dual Update $U_D$

- $S$ and $W$ now denote the support and active set of $x^k$.
- We solve the following linear program with $|W|$ bounded variables and $2n - |S|$ constraints to obtain a new dual solution:

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

## Properties

- After $K \leq (3^{m+n} + 1)/2$ consecutive dual and primal updates, the method terminates yielding an optimal pair $(x^K, y^K)$ for $(\mathrm{P}_{\delta^K})$.
- The solution path of $(\mathrm{P}_\delta)$ is continuous piecewise linear. Our method implicitly generates an optimal solution for each problem $(\mathrm{P}_{\hat\delta})$ with $\delta \leq \hat\delta \leq \|b\|_\infty$.
- The linear programs in $U_D$ and $U_P$ can be tackled by an arbitrary LP solver. We propose an active set approach that covers two essential aspects:
  1. The iterates $y^k$ and $x^k$ are feasible starting points for $U_D$ and $U_P$, respectively.
  2. Lagrange multipliers certifying optimality of $y^{k+1}$ in $U_D$ qualify as an initial search direction at $x^k$ in $U_P$, and vice versa.

## Primal Update $U_P$

- In the following, $\Omega$ and $\Sigma$ denote the support and active set of $y^{k+1}$.
- For the primal update, we solve the following linear program with $|\Sigma| + 1$ bounded variables an $2m - |\Omega|$ constraints:

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

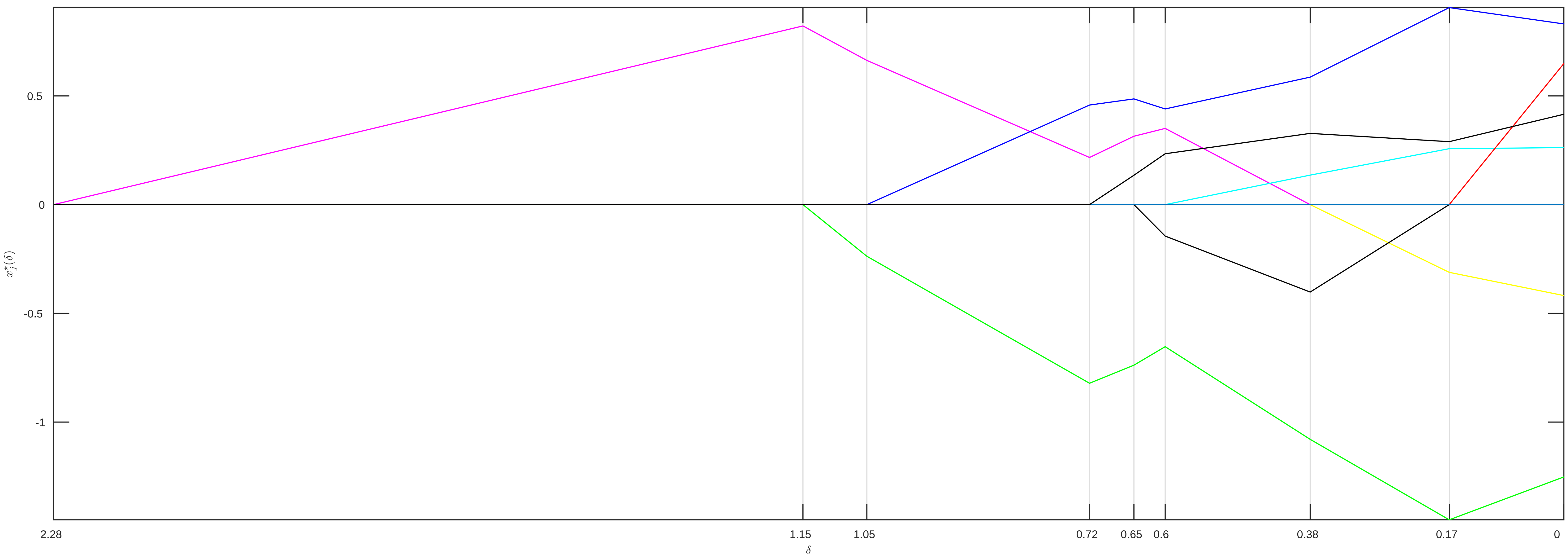
- The choice of the objective functions in $U_D$ and $U_P$ is motivated by a theorem of the alternative and plays a key role in view of finite termination.

**SPARS 2017** Lisbon, Portugal June 5-8, 2017

RWTH AACHEN UNIVERSITY

CAROLO-WILHELMINA BRAUNSCHWEIG Technische Universität Braunschweig

## Exemplary Solution Path




Exemplary run of $\ell_1$-Houdini (using active set) with $A \in \mathbb{R}^{6\times 12}$ and $b \in \mathbb{R}^6$ randomly generated and $\delta = 0$. The algorithm needed 9 iterations to solve the problem. Horizontal labels display the value of the homotopy parameter $\delta^k$ after each iteration. The plots represent the solution paths of $x_j^k$ for $j = 1, \ldots, 12$. The optimal solution has 6 nonzero entries.

## Runtime and Accuracy Comparison for the Dantzig Selector [4]

| inst. | runtime in seconds | | | $\|x^\star\|_1$ | | | constraint violation | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\ell_1$-Hou. | PDP | Gur. | $\ell_1$-Hou. | PDP | Gur. | $\ell_1$-Hou. | PDP | Gur. |
| 1 | 0.19 | **0.14** | 2.22 | 97.09 | 97.09 | 97.09 | $3 \cdot 10^{-15}$ | $4 \cdot 10^{-15}$ | $3 \cdot 10^{-15}$ |
| 2 | 1.02 | **0.64** | 2.36 | 154.93 | 154.93 | 154.93 | $3 \cdot 10^{-15}$ | $7 \cdot 10^{-15}$ | $4 \cdot 10^{-15}$ |
| 3 | 0.34 | **0.27** | 8.93 | 96.41 | 96.41 | 96.41 | $3 \cdot 10^{-15}$ | $3 \cdot 10^{-15}$ | $4 \cdot 10^{-15}$ |
| 4 | 2.74 | **1.48** | 9.19 | 188.03 | 188.03 | 188.03 | $4 \cdot 10^{-15}$ | $1 \cdot 10^{-14}$ | $6 \cdot 10^{-15}$ |
| 5 | **0.21** | 0.26 | 2.26 | 98.68 | 98.68 | 98.68 | $3 \cdot 10^{-15}$ | $5 \cdot 10^{-15}$ | $2 \cdot 10^{-15}$ |
| 6 | **0.47** | 0.52 | 2.35 | 152.03 | 152.03 | 152.03 | $5 \cdot 10^{-15}$ | $1 \cdot 10^{-14}$ | $5 \cdot 10^{-15}$ |
| 7 | 0.44 | **0.41** | 9.11 | 95.73 | 95.73 | 95.73 | $5 \cdot 10^{-15}$ | $6 \cdot 10^{-15}$ | $5 \cdot 10^{-15}$ |
| 8 | **0.84** | 0.86 | 9.22 | 186.19 | 186.19 | 186.19 | $5 \cdot 10^{-15}$ | $1 \cdot 10^{-14}$ | $5 \cdot 10^{-15}$ |
| 9 | 0.03 | 0.02 | $<$ **0.01** | 44.64 | 44.64 | **9.36** | $3 \cdot 10^{-10}$ | $3 \cdot 10^{-4}$ | $\mathbf{2 \cdot 10^{-2}}$ |
| 10 | 0.03 | 0.02 | $<$ **0.01** | 304.27 | 304.27 | **6.03** | $1 \cdot 10^{-8}$ | $4 \cdot 10^{-3}$ | $\mathbf{2 \cdot 10^{-1}}$ |
| 11 | 0.02 | 0.01 | $<$ **0.01** | 316.35 | 316.35 | 316.35 | $7 \cdot 10^{-8}$ | $1 \cdot 10^{-4}$ | $1 \cdot 10^{-7}$ |
| 12 | 0.04 | 0.02 | $<$ **0.01** | 64.18 | 64.18 | 64.18 | $3 \cdot 10^{-9}$ | $6 \cdot 10^{-7}$ | $7 \cdot 10^{-10}$ |
| 13 | **0.02** | — | 0.03 | 0.79 | — | $\mathbf{2 \cdot 10^5}$ | $7 \cdot 10^{-7}$ | — | $4 \cdot 10^{-9}$ |
| 14 | **0.21** | 3.47 | 0.52 | 0.67 | 1.88 | **634.89** | $7 \cdot 10^{-7}$ | $1 \cdot 10^{-7}$ | $1 \cdot 10^{-11}$ |
| 15 | 176.76 | 5.52 | **1.11** | 998.72 | **157.41** | 998.72 | $8 \cdot 10^{-7}$ | $\mathbf{4 \cdot 10^4}$ | $4 \cdot 10^{-7}$ |

The first part of the comparison shows that the runtimes of $\ell_1$-Houdini [3] and PDP [1] often lie in the same magnitude while the respective runtimes of Gurobi are significantly larger. We can further observe that $\ell_1$-Houdini is fastest in case $m > n$ which is of interest in many *machine learning* applications, where the number of training examples is much larger than the number of features. Applied to the empirical data from [5], Gurobi is the fastest algorithm in the majority of cases, while PDP fails to find an optimal solution in three out of seven cases. The table finally shows that $\ell_1$-Houdini is the only algorithm that works with high accuracy on the whole test set.

| inst. | description | $m$ | $n$ | $\delta$ | $\|S\|$ |
|---|---|---|---|---|---|
| 1 | random [4] | 1024 | 1024 | 0.39 | 66 |
| 2 | random [4] | 1024 | 1024 | 0.51 | 152 |
| 3 | random [4] | 1024 | 2048 | 0.27 | 69 |
| 4 | random [4] | 1024 | 2048 | 0.39 | 166 |
| 5 | random [4] | 2048 | 1024 | 0.35 | 65 |
| 6 | random [4] | 2048 | 1024 | 0.55 | 128 |
| 7 | random [4] | 2048 | 2048 | 0.29 | 64 |
| 8 | random [4] | 2048 | 2048 | 0.39 | 130 |
| 9 | Wine (red) [5] | 1599 | 12 | 0.00 | 12 |
| 10 | Wine (white) [5] | 4898 | 12 | 0.00 | 12 |
| 11 | Airfoil Self-Noise [5] | 1503 | 6 | 0.00 | 6 |
| 12 | Housing [5] | 506 | 14 | 0.00 | 14 |
| 13 | Online News Popularity [5] | 39644 | 59 | 0.00 | 6 |
| 14 | Blog Feedback [5] | 52396 | 280 | 0.00 | 11 |
| 15 | Relative location of CT sclices on axial axis [5] | 53500 | 385 | 0.00 | 385 |

## References


[1] M. S. Asif and J. Romberg, "Dantzig selector homotopy with dynamic measurements", *Proc. SPIE*, vol. 7246, pp. 72460E–72460E–11, Feb. 2009.

[2] C. Brauer, T. Gerkmann and D. A. Lorenz, "Sparse Reconstruction of Quantized Speech Signals", *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5940–5944, 2016.

[3] C. Brauer, D. A. Lorenz and A. M. Tillmann, "A Primal-Dual Homotopy Algorithm for $\ell_1$-Minimization with $\ell_\infty$-Constraints", submitted, Oct. 2016. arXiv:1610.10022 [math.OC].

[4] E. J. Candès and T. Tao, "The Dantzig Selector: Statistical Estimation When $p$ Is Much Larger than $n$", *The Annals of Statistics*, vol. 35, no. 6, pp. 2313–2351, Dec. 2007.

[5] M. Lichman, *UCI Machine Learning Repository*, `http://archive.ics.uci.edu/ml`, Irvine, CA: University of California, School of Information and Computer Science, 2013.




RWTH AACHEN UNIVERSITY

Technische Universität Braunschweig