ℓ_1-HOUDINI: A New Homotopy Method for ℓ_1-Minimization

Christoph Brauer and Dirk A. Lorenz
Institute for Analysis and Algebra, TU Braunschweig
Pockelsstraße 14, D-38106 Braunschweig, Germany

Andreas M. Tillmann
Research Group Optimization, TU Darmstadt
Dolivostr. 15, D-64293 Darmstadt, Germany

***Abstract*—We propose a primal-dual homotopy method for ℓ_1-minimization problems with infinity norm constraints, where the natural homotopy parameter is the value of the bound for the constraints. Motivated by primal-dual optimality conditions, each iteration of our method decomposes into two relatively small linear programs. The effectiveness and the competitiveness of our method are demonstrated in numerical experiments.**

I. Problem and Optimality Conditions

We consider the problem

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

with $\boldsymbol{A}\in\mathbb{R}^{m\times n}$, $\boldsymbol{b}\in\mathbb{R}^m$ and $\delta\ge 0$. It is well-known that $\boldsymbol{x}^*$ is an optimal solution of (P$_\delta$) if and only if there exists a $\boldsymbol{y}^*$ such that

$$-\boldsymbol{A}^\top\boldsymbol{y}^*\in\partial\|\boldsymbol{x}^*\|_1 \quad\text{and}\quad \boldsymbol{Ax}^*-\boldsymbol{b}\in\delta\partial\|\boldsymbol{y}^*\|_1 \tag{1}$$

hold. Each such $\boldsymbol{y}^*$ is by construction an optimal solution to the dual problem of (P$_\delta$). Therefore, we sometimes refer to $\boldsymbol{x}^*$ as a *primal solution*, to $\boldsymbol{y}^*$ as a *dual solution* and to $(\boldsymbol{x}^*,\boldsymbol{y}^*)$ as an *optimal pair*. For a thorough understanding of the conditions (1), it is helpful to define the *primal support* $S:=\{j:\boldsymbol{x}_j^*\neq 0\}$, the *primal active set* $W:=\{i:|\boldsymbol{a}_i^\top\boldsymbol{x}^*-\boldsymbol{b}_i|=\delta\}$, the *dual support* $\Omega:=\{i:\boldsymbol{y}_i^*\neq 0\}$ and the *dual active set* $\Sigma:=\{j:|A_j^\top\boldsymbol{y}^*|=1\}$. Since $\partial\|\boldsymbol{x}^*\|_1=\{\boldsymbol{w}\in[-1,1]^n:\boldsymbol{w}_S=\operatorname{sign}(\boldsymbol{x}_S^*)\}$, we can partition (1) as follows:

$$\begin{aligned}
-\boldsymbol{A}_S^\top\boldsymbol{y}^* &= \operatorname{sign}(\boldsymbol{x}_S^*) & \boldsymbol{A}^\Omega\boldsymbol{x}^*-\boldsymbol{b}_\Omega &= \delta\operatorname{sign}(\boldsymbol{y}_\Omega^*)\\
-\mathbb{1}\le -\boldsymbol{A}_{S^c}^\top\boldsymbol{y}^* &\le \mathbb{1} & -\delta\mathbb{1}\le \boldsymbol{A}^{\Omega^c}\boldsymbol{x}^*-\boldsymbol{b}_{\Omega^c} &\le \delta\mathbb{1}\\
0 &= \boldsymbol{x}_{\Sigma^c}^* & 0 &= \boldsymbol{y}_{W^c}^*
\end{aligned} \tag{2}$$

Although (P$_\delta$) is motivated by various applications, such as *sparse dequantization* [1], *sparse linear discriminant analysis* [2] and *sparse precision matrix estimation* [3], our subsequent numerical experiments focus on the *Dantzig selector* problem [4] which has itself numerous applications in statistical estimation. In the following, we describe our method which we call ℓ_1-HOUDINI (ℓ_1-norm HOmotopy UnDer Infinity-Norm constraInts).

II. Homotopy Method

Suppose that $\delta^k>\delta$ and that $(\boldsymbol{x}^k,\boldsymbol{y}^k)$ is an optimal pair for (P$_{\delta^k}$). Hence, the conditions (2) hold for $\boldsymbol{x}^k$, $\boldsymbol{y}^k$ and δ^k. As a first step in each iteration, we construct $\boldsymbol{y}^{k+1}\neq\boldsymbol{y}^k$ such that $(\boldsymbol{x}^k,\boldsymbol{y}^{k+1})$ is still an optimal pair for (P$_{\delta^k}$). To that end, we fix $\boldsymbol{x}^k$ and δ^k in (2) and use the resulting conditions as constraints to reduce the dimension of the search space in the following linear program

$$\begin{aligned}
\min_{\boldsymbol{y}_W\in\mathbb{R}^{|W|}} \quad & -\operatorname{sign}(\boldsymbol{A}^W\boldsymbol{x}^k-\boldsymbol{b}_W)^\top\boldsymbol{y}_W\\
\text{s.t.} \quad & -(\boldsymbol{A}_S^W)^\top\boldsymbol{y}_W=\operatorname{sign}(\boldsymbol{x}_S^k)\\
& -\mathbb{1}\le-(\boldsymbol{A}_{S^c}^W)^\top\boldsymbol{y}_W\le\mathbb{1}\\
& -\operatorname{sign}(\boldsymbol{A}^W\boldsymbol{x}^k-\boldsymbol{b}_W)\odot\boldsymbol{y}_W\le 0.
\end{aligned} \tag{3}$$

We define $\boldsymbol{y}_W^{k+1}$ as a solution of (3) and set $\boldsymbol{y}_{W^c}^{k+1}:=0$. In a second step, we construct $\boldsymbol{x}^{k+1}\neq\boldsymbol{x}^k$ and $t>0$ such that $(\boldsymbol{x}^{k+1},\boldsymbol{y}^{k+1})$ is an optimal pair for (P$_{\delta^k-t}$). This time, we fix $\boldsymbol{y}^{k+1}$ in (2) which gives rise to the linear program

$$\begin{aligned}
\max_{(\boldsymbol{x}_\Sigma,t)\in\mathbb{R}^{|\Sigma|}\times\mathbb{R}} \quad & t\\
\text{s.t.} \quad & \boldsymbol{A}_\Sigma^\Omega\boldsymbol{x}_\Sigma-\boldsymbol{b}_\Omega=(\delta^k-t)\operatorname{sign}(\boldsymbol{y}_\Omega^{k+1})\\
& -(\delta^k-t)\mathbb{1}\le\boldsymbol{A}_\Sigma^{\Omega^c}\boldsymbol{x}_\Sigma-\boldsymbol{b}_{\Omega^c}\le(\delta^k-t)\mathbb{1}\\
& (\boldsymbol{A}_\Sigma^\top\boldsymbol{y}^{k+1})\odot\boldsymbol{x}_\Sigma\le 0\\
& t\le\delta^k-\delta.
\end{aligned} \tag{4}$$

Finally, we define $\boldsymbol{x}_\Sigma^{k+1}$ as a solution of (4), set $\boldsymbol{x}_{\Sigma^c}^{k+1}:=0$ and $\delta^{k+1}:=\delta^k-t$. Note that the sets S, W, Ω and Σ now refer to the respective primal and dual iterates and need to be updated in parallel to those.

After at most $(3^{m+n}+1)/2$ iterations we reach $\delta^{k+1}=\delta$ and $\boldsymbol{x}^{k+1}$ is an optimal solution of (P$_\delta$), while each intermediate iterate $\boldsymbol{x}^k$ is an optimal solution of the related problem (P$_{\delta^k}$). The choice of the objective functions in (3) and (4) is motivated by a theorem of the alternative and plays a key role in view of convergence, see [5] for a thorough convergence analysis and a proof of the following theorem.

Theorem 1: Starting at $\boldsymbol{x}^0:=0$ and $\delta^0:=\|b\|_\infty$, ℓ_1-HOUDINI terminates after a finite number of iterations and returns an optimal solution of (P$_\delta$).

The fact that an arbitrary LP solver can be used to tackle (3) and (4) can be considered an advantage of our method as it makes it adaptable and easy to implement. Nevertheless, one can exploit the structure of (3) and (4) in order to design efficient methods. In [5], we propose an active set method that covers two essential aspects. First, the previous iterates $\boldsymbol{y}^k$ and $\boldsymbol{x}^k$ are feasible starting points in the subproblems to determine $\boldsymbol{y}^{k+1}$ and $\boldsymbol{x}^{k+1}$, respectively. Second, Lagrange multipliers certifying optimality of $\boldsymbol{y}^{k+1}$ in (3) qualify as an initial search direction at $\boldsymbol{x}^k$ in (4), and vice versa.

III. Numerical Experiments

The *Dantzig selector* problem is a special case of (P$_\delta$), where the constraint is replaced by $\|\boldsymbol{A}^\top(\boldsymbol{Ax}-\boldsymbol{b})\|_\infty\le\delta$. A specific homotopy scheme for this problem, called *Primal Dual pursuit* (PDP), was proposed in [6]. We compare our method to PDP and to the commercial LP solver GUROBI, where we apply the latter to the LP reformulation of (P$_\delta$). Our test set includes random instances according to [4] and instances from [7]. Table II provides an overview.

The first part of the comparison in Table I shows that the runtimes of ℓ_1-HOUDINI and PDP often lie in the same magnitude while the respective runtimes of GUROBI are significantly larger. We can further observe that ℓ_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 [7], GUROBI is the fastest algorithm in the majority of cases, while PDP fails to find an optimal solution in three out of seven cases. Table I finally shows that ℓ_1-HOUDINI is the only algorithm that works with high accuracy on the whole test set.

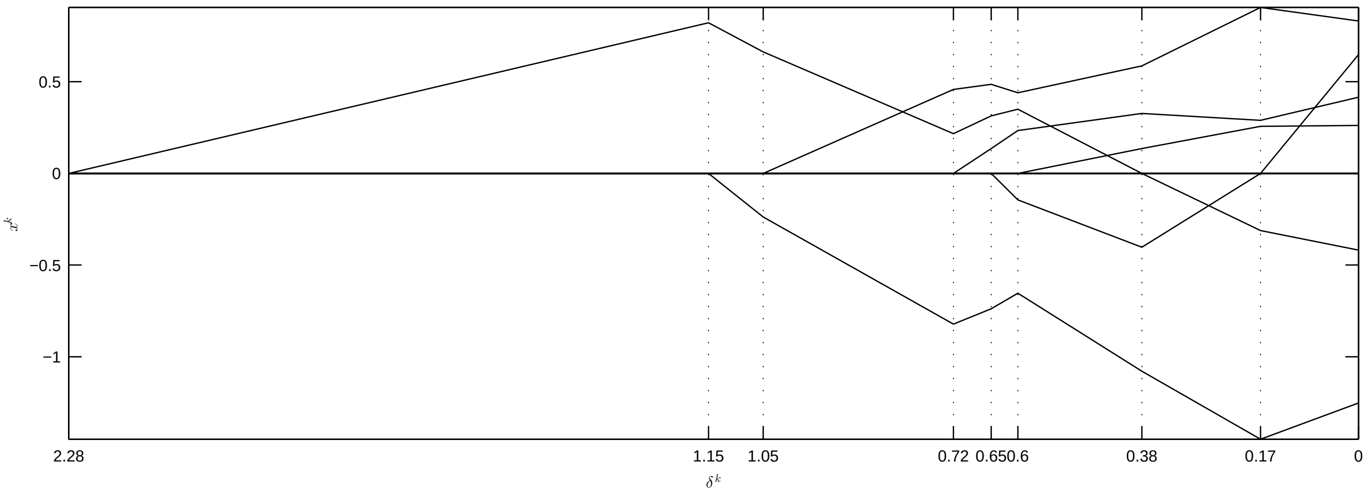

Fig. 1. Examplary run of ℓ_1-HOUDINI (using active set) with $A \in \mathbb{R}^{6\times 12}$ and $b \in \mathbb{R}^6$ randomly generated and $\delta = 0$. The algorithm needed 9 iterations to solve the problem. Horizontal labels display the value of the homotopy parameter δ^k after each iteration. The plots represent the solution paths of x_j^k for $j = 1, \dots, 12$. The optimal solution has 6 nonzero entries.

inst.	runtime in seconds			$\|\boldsymbol{x}^*\|_1$			constraint violation		
	ℓ_1-HOU.	PDP	GUR.	ℓ_1-HOU.	PDP	GUR.	ℓ_1-HOU.	PDP	GUR.
1	0.19	0.14	2.22	97.09	97.09	97.09	$3\cdot10^{-15}$	$4\cdot10^{-15}$	$3\cdot10^{-15}$
2	1.02	0.64	2.36	154.93	154.93	154.93	$3\cdot10^{-15}$	$7\cdot10^{-15}$	$4\cdot10^{-15}$
3	0.34	0.27	8.93	96.41	96.41	96.41	$3\cdot10^{-15}$	$3\cdot10^{-15}$	$4\cdot10^{-15}$
4	2.74	1.48	9.19	188.03	188.03	188.03	$4\cdot10^{-15}$	$1\cdot10^{-14}$	$6\cdot10^{-15}$
5	0.21	0.26	2.26	98.68	98.68	98.68	$3\cdot10^{-15}$	$5\cdot10^{-15}$	$2\cdot10^{-15}$
6	0.47	0.52	2.35	152.03	152.03	152.03	$5\cdot10^{-15}$	$1\cdot10^{-14}$	$5\cdot10^{-15}$
7	0.44	0.41	9.11	95.73	95.73	95.73	$5\cdot10^{-15}$	$6\cdot10^{-15}$	$5\cdot10^{-15}$
8	0.84	0.86	9.22	186.19	186.19	186.19	$5\cdot10^{-15}$	$1\cdot10^{-14}$	$5\cdot10^{-15}$
9	0.03	0.02	< 0.01	44.64	44.64	9.36	$3\cdot10^{-10}$	$3\cdot10^{-4}$	$2\cdot10^{-2}$
10	0.03	0.02	< 0.01	304.27	304.27	6.03	$1\cdot10^{-8}$	$4\cdot10^{-3}$	$2\cdot10^{-1}$
11	0.02	0.01	< 0.01	316.35	316.35	316.35	$7\cdot10^{-8}$	$1\cdot10^{-4}$	$1\cdot10^{-7}$
12	0.04	0.02	< 0.01	64.18	64.18	64.18	$3\cdot10^{-9}$	$6\cdot10^{-7}$	$7\cdot10^{-10}$
13	0.02	-	0.03	0.79	-	$2\cdot10^{5}$	$7\cdot10^{-7}$	-	$4\cdot10^{-9}$
14	0.21	3.47	0.52	0.67	1.88	634.89	$7\cdot10^{-7}$	$1\cdot10^{-7}$	$1\cdot10^{-11}$
15	176.76	5.52	1.11	998.72	157.41	998.72	$8\cdot10^{-7}$	$4\cdot10^{4}$	$4\cdot10^{-7}$

TABLE I
RUNTIME AND ACCURACY COMPARISON.

inst.	description	m	n	δ	$\|S\|$
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) [7]	1599	12	0.00	12
10	Wine (white) [7]	4898	12	0.00	12
11	Airfoil Self-Noise [7]	1503	6	0.00	6
12	Housing [7]	506	14	0.00	14
13	Online News Popularity [7]	39644	59	0.00	6
14	Blog Feedback [7]	52396	280	0.00	11
15	Relative location of CT sclices on axial axis [7]	53500	385	0.00	385

TABLE II
TEST INSTANCES.

REFERENCES

[1] 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.

[2] T. Cai and W. Liu, "A Direct Estimation Approach to Sparse Linear Discriminant Analysis", *Journal of the American Statistical Association*, vol. 106, no. 496, pp. 1566–1577, 2011.

[3] T. Cai, W. Liu and X. Luo, "A Constrained ℓ_1 Minimization Approach to Sparse Precision Matrix Estimation", *Journal of the American Statistical Association*, vol. 106, no. 494, pp. 594–607, 2011.

[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] C. Brauer, D. A. Lorenz and A. M. Tillmann, "A Primal-Dual Homotopy Algorithm for ℓ_1-Minimization with ℓ_∞-Constraints", submitted, Oct. 2016. arXiv:1610.10022 [math.OC].

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

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