Equivalence of Linear Programming and Basis Pursuit

Andreas M. Tillmann[*]

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

In this note, we show that linear programming and the prominent Basis Pursuit problem (i.e., minimizing the ℓ_1-norm of a vector x subject to an underdetermined linear equation system $Ax = b$) are theoretically equivalent, and briefly discuss possible ramifications regarding computational complexity and practical applicability.

1 Introduction

In the context of compressed sensing and sparse signal recovery, ℓ_1-minimization tasks like the popular Basis Pursuit problem

$$\min \|x\|_1 \quad \text{s.t.} \quad Ax = b, \tag{P$_1$}$$

where $A \in \mathbb{R}^{m\times n}$ (with $\operatorname{rank}(A) = m \leq n$) and $b \in \mathbb{R}^m$, have experienced a surge of interest in recent years. It is well-known that (P$_1$) can be recast as a linear program (LP), e.g., as $\min\{\, \mathbb{1}^\top x^+ + \mathbb{1}^\top x^- \;:\; Ax^+ - Ax^- = b;\, x^+, x^- \geq 0 \,\}$. In fact, LP approaches yield some of the fastest and most reliable methods in practice (cf. [1]). In this paper, we show that the converse holds as well: Every (bounded, feasible) LP (with rational data) can be transformed into an equivalent ℓ_1-minimization problem. As our reduction requires only polynomial time and space w.r.t. the input encoding length, this shows polynomial equivalence of the two problem classes.

This work constitutes a revised and highly condensed version of a part of the author's dissertation [2].

1.1 Related Work

An equivalence between linear programs and a (seemingly) different ℓ_1-minimization problem has been noted before: In [3], it was shown that any bounded and feasible LP can be rewritten as the dual of the least absolute value (LAV) regression problem

$$\min \|M\xi - f\|_1 \tag{1}$$

(here, $M \in \mathbb{R}^{p\times n}$ with $p > n$ and $f \in \mathbb{R}^p$). The reduction in [3] is similar in spirit to ours (see Section 2 below); in particular, it involves a certain "Big-M" argument and dualization. However, [3] merely sketches the procedure without giving precise values for several "suitably large" constants needed in the transformation. In our reduction, we will derive explicit values for such constants that retain encoding lengths which depend polynomially on that of the LP.

In fact, (1) is (polynomially) equivalent to a problem of the form (P$_1$). Detailed proofs can be found in [4] or [2, Prop. 3.12], so we only briefly sketch the essentials here: First, note that (P$_1$) can be recast as an unconstrained optimization problem by means of standard variable elimination techniques.

Lemma 1.1 *The Basis Pursuit problem* (P$_1$) *is polynomially equivalent to*

$$\min \|z\|_1 + \|d - Dz\|_1. \tag{P$_1^\dagger$}$$

(Here, if $A \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(A) = m \leq n$ in (P$_1$), $D \in \mathbb{R}^{m\times(n-m)}$, $d \in \mathbb{R}^m$, and minimization is performed over z.) Obviously, (P$_1^\dagger$) can be stated in the form (1); combined with Lemma 1.1, this shows the first direction of the following result; the converse direction can also be derived straightforwardly (see [2, 4]).

Proposition 1.2 *The Basis Pursuit problem* (P$_1$) *is polynomially equivalent to the LAV regression problem* (1).

By these results, a given linear program could be transformed to (P$_1$) by using the reformulation sketched in [3], dualizing, and then applying Proposition 1.2. However, the reduction we present below avoids the intermediate step involving (1) and builds on Lemma 1.1 directly.

---

[*] E-mail tillmann@mathematik.tu-darmstadt.de, phone +49 6151 1670868. During GAMM 2015, the author was with the Institute for Mathematical Optimization at TU Braunschweig, Pockelsstr. 14, D-38106 Braunschweig, Germany (on leave from TU Darmstadt).

1.2 Notation & Preliminaries

Before turning to the announced reduction, we provide further auxiliary results and establish some notation.

Lemma 1.3 *The dual problem of* ($P_1^\dagger$) *reads*

$$\max\ -d^\top q \quad \text{s.t.} \quad \|q\|_\infty \le 1,\ \|D^\top q\|_\infty \le 1. \tag{$D_1^\dagger$}$$

P r o o f. This well-known fact can be obtained easily, e.g., by standard LP duality arguments. □

Remark 1.4 Since strong duality holds for ($P_1^\dagger$) and ($D_1^\dagger$), the optimum for one problem can be obtained by solving the other and the two problems are in fact polynomially equivalent. By Lemma 1.1, the same holds for (P_1) and ($D_1^\dagger$).

Recall the (binary) encoding length of an integer z, $\langle z\rangle := 1 + \lceil \log_2(|z|) + 1\rceil$; for a rational number r, $\langle r\rangle := \langle s\rangle + \langle t\rangle$ where s, t are mutually prime integers such that $r = s/t$. For matrices and vectors, the respective encoding length is the sum of those of their entries. Some estimates involving encoding lengths are gathered in the following lemma; for a proof, we refer to [2, Lem. 3.15] and the references therein (in particular, [5]).

Lemma 1.5 *Let $r, s \in \mathbb{Q}$, $A \in \mathbb{Q}^{m\times n}$, $B \in \mathbb{Q}^{n\times p}$, $C \in \mathbb{Q}^{n\times n}$ nonsingular, $b \in \mathbb{Q}^m$, and $z \in \mathbb{Z}$. It holds that*

$$\langle C^{-1}\rangle \le 4n^2 \langle C\rangle, \tag{2}$$

$$|r| \le 2^{\langle r\rangle - 1} - 1. \tag{3}$$

Moreover, for every vertex $v = (v_1, \dots, v_n)^\top$ of a polyhedron in one of the forms $\{\, x : Ax \le b \,\}$, $\{\, x : Ax = b \,\}$ or $\{\, x : Ax \le b,\ x \ge 0 \,\}$, it holds that

$$|v_j| \le 2^{2\langle A\rangle + \langle b\rangle - 2n^2} =: K \qquad \textit{for all } j \in [n] := \{1, 2, \dots, n\}. \tag{4}$$

2 The Reduction

We consider a linear program with rational data ($A \in \mathbb{Q}^{m\times n}$, $b \in \mathbb{Q}^m$ and $c \in \mathbb{Q}^n$) in standard inequality form:

$$\max\ c^\top x \quad \text{s.t.} \quad Ax \le b,\ x \ge 0, \tag{LP}$$

Throughout, we assume w.l.o.g. that A does not contain any all-zero rows or columns. To simplify the main proof (i.e., that of Theorem 2.2), we additionally make the following assumptions; our result can easily be adapted to accommodate the excluded cases as well, see Proposition 2.4 later.

Assumption 2.1 *The inequality system of* (LP) *describes a* nonempty *polyhedron, and the objective function $c^\top x$ remains* bounded *over this polyhedron.*

Theorem 2.2 *Under Assumption 2.1, linear programming is polynomially equivalent to the Basis Pursuit problem* (P_1).

P r o o f. We already know that (P_1) can be reformulated as an LP; the same clearly holds for ($D_1^\dagger$). Moreover, it is well-known that every linear program can be stated in the above standard inequality form (LP). Thus, because (P_1) and ($D_1^\dagger$) are polynomially equivalent by Remark 1.4, it suffices to show that (LP) can be transformed (using polynomial time and space only) into an instance of ($D_1^\dagger$) (under Assumption 2.1).

By Assumption 2.1, the optimum of (LP) is finite and therefore attained at a vertex of $\{\, x : Ax \le b,\ x \ge 0 \,\} \ne \emptyset$. Thus, we can restrict our attention to a bounded polyhedron (i.e., a polytope) by explicitly including a constraint of the sort (4) for each variable; see also [6, Theorem 2.2]. Note that this does not cut off any vertices, since these constraints are *implied* by the data. Since in (LP), $x \ge 0$ and hence $|x_j| = x_j$ for all $j \in [n]$, it follows from Lemma 1.5 that (LP) is equivalent to

$$\max\ c^\top x \quad \text{s.t.} \quad Ax \le b,\ 0 \le x \le K\mathbb{1}, \tag{5}$$

with K as in (4) and $\mathbb{1} := (1, \dots, 1)^\top$. Substituting x by $\frac{K}{2}(\mathbb{1} + y)$, i.e., with $y := \frac{2}{K}x - \mathbb{1}$, we see that (5) is equivalent to

$$\max\ c^\top \left(\tfrac{K}{2}\mathbb{1} + \tfrac{K}{2}y\right) \quad \text{s.t.} \quad A\left(\tfrac{K}{2}\mathbb{1} + \tfrac{K}{2}y\right) \le b,\ 0 \le \tfrac{K}{2}\mathbb{1} + \tfrac{K}{2}y \le K\mathbb{1}.$$

Omitting the constant terms and factors in the objective, and denoting $r := \frac{2}{K}b - A\mathbb{1}$, this can be rewritten as

$$\max\{\, c^\top y : Ay \le r,\ -\mathbb{1} \le y \le \mathbb{1} \,\} = \max\{\, c^\top y : Ay \le r,\ \|y\|_\infty \le 1 \,\}. \tag{6}$$

Now, observe that $-\mathbb{1} \le y \le \mathbb{1}$ implies $-\|a_i^\top\|_1 \ \le\ a_i^\top y \ \le\ \|a_i^\top\|_1$ for all $i \in [m]$. Combined with $Ay \le r$, we obtain

$$-\|a_i^\top\|_1 \le a_i^\top y \ \le\ \min\{r_i,\ \|a_i^\top\|_1\} \qquad \text{for all } i \in [m]. \tag{7}$$

In fact, we may assume w.l.o.g. that $\min\{r_i,\, \|a_i^\top\|_1\} = r_i$ holds for all i (otherwise, if $r_i > \|a_i^\top\|_1$ for some $i \in [m]$, the constraint $a_i^\top y \leq r_i$ is redundant and can be omitted) and that $r_i \neq -\|a_i^\top\|_1$ for all i (which can be assured a priori by appropriate scaling, cf. [2, Prop. 3.19]). Moreover, Assumption 2.1 excludes the case $r_i < -\|a_i^\top\|_1$ (as it would imply inconsistency of the constraint system), and because A contains no zero rows, we have $\|a_i^\top\|_1 \neq 0^\top$ for all $i \in [m]$. Thus,

$$(7) \quad \Leftrightarrow \quad -\|a_i^\top\|_1 \leq a_i^\top y \leq r_i \qquad \text{for all } i \in [m] \quad \Leftrightarrow \quad -1 \leq \tilde{a}_i^\top y + \delta_i \leq 1 \qquad \text{for all } i \in [m], \tag{8}$$

where $\tilde{a}_i^\top := \big(2/(\|a_i^\top\|_1 + r_i)\big)a_i^\top$ and $\delta_i := (\|a_i^\top\|_1 - r_i)/(\|a_i^\top\|_1 + r_i)$. Hence, the constraint system $Ay \leq r$ can be replaced by the m box constraints of the form (8). Denoting by $\tilde{A}$ the matrix consisting of rows $\tilde{a}_i^\top$ and by δ the vector containing the numbers δ_i, this shows that (6) is equivalent to

$$\max\{\, c^\top y \,:\, -\mathbb{1} \leq \tilde{A}y + \delta \leq \mathbb{1},\ \|y\|_\infty \leq 1 \,\} = \max\{\, c^\top y \,:\, \|\tilde{A}y + \delta\|_\infty \leq 1,\ \|y\|_\infty \leq 1 \,\}. \tag{9}$$

Note that the only difference between (9) and the desired form ($D_1^\dagger$) is the presence of the constant "shift" δ in the constraint system. We can eliminate this shift by a *lifting* procedure, which we describe in the following.

Our goal is to transform the shift δ into the coefficient vector of a new variable $z \in [-1, 1]$, thereby obtaining the desired unit box constraint form. To that end, suppose we assign an objective function coefficient $\mathcal{M}$ to z and replace δ by δz in (9). If $\mathcal{M}$ is sufficiently large (hence the name "Big-M" for this kind of argument), setting $z = 1$ is superior to any other choice and implicitly enforces the original constraints (8) on y, i.e., the original bounds are restored, and an optimal solution of (9) is indeed in one-to-one correspondence with the y-part of an optimal solution for

$$\max\ c^\top y + \mathcal{M}z \quad \text{s.t.} \quad \left\|\Big(\tilde{A},\, \delta\Big)\begin{pmatrix} y \\ z \end{pmatrix}\right\|_\infty \leq 1,\ \left\|\begin{pmatrix} y \\ z \end{pmatrix}\right\|_\infty \leq 1. \tag{10}$$

Note that with respect to any *fixed* value of z, replacing the constant shift δ by δz can be regarded as changing the box constraint bounds on $\tilde{A}y$. Let $\hat{A}$ and $\hat{b}$ denote the matrix and vector such that $\hat{A}y \leq \hat{b}$ is equivalent to the constraints in (9). Replacing δ by δz (for some fixed z) changes only the $2m$ entries in $\hat{b}$ that correspond to $-\mathbb{1} - \delta \leq \tilde{A}y \leq \mathbb{1} - \delta$; namely, from $\mathbb{1} \pm \delta$ to $\mathbb{1} \pm \delta z$. Resulting variations in the LP objective due to changing $\hat{b}$ to some $\hat{b}'$ can be bounded as follows (cf. [7]):

$$\big|\ \max\{\, c^\top y \,:\, \hat{A}y \leq \hat{b} \,\} - \max\{\, c^\top y \,:\, \hat{A}y \leq \hat{b}' \,\}\ \big| \ \leq\ n\,\beta\,\|c\|_1\,\|\hat{b} - \hat{b}'\|_\infty,$$

where β is an upper bound on the absolute values of all entries in the inverses of all regular submatrices B of $\hat{A}$. In the present case, since each $\delta_i \geq 0$, we have for any $z \in [-1, 1]$ that

$$\|\hat{b} - \hat{b}'\|_\infty \ =\ \max_{i\in[m]}\big\{\, |1 - \delta_i - (1 - \delta_i z)|,\ |1 + \delta_i - (1 + \delta_i z)| \,\big\} \ =\ \max_{i\in[m]}\{\, \delta_i \,\}(1 - z)$$

and obtain the following crude upper bound on $|(B^{-1})_{ij}|$ for any such B (non-singular submatrix of $\hat{A}$):

$$\beta \ :=\ 2^{4n^2\langle\hat{A}\rangle} \ \geq\ 2^{4n^2\langle B\rangle} \ \overset{(2)}{\geq}\ 2^{\langle B^{-1}\rangle} \ \geq\ 2^{\langle|(B^{-1})_{ij}|\rangle - 1} - 1 \ \overset{(3)}{\geq}\ |(B^{-1})_{ij}|.$$

To summarize, due to implicit bound changes for $\tilde{A}y$ caused by introducing $z \in [-1, 1]$, the objective value contribution of y can *increase* (with respect to the optimal value of (9)) by at most

$$n\,\beta\,\|c\|_1 \max_{i\in[m]}\{\, \delta_i \,\}(1 - z) \ =\ 2^{4n^2\langle\hat{A}\rangle}\, n\,\|c\|_1 \max_{i\in[m]}\{\, \delta_i \,\}(1 - z) \ =:\ M \cdot (1 - z).$$

Hence, setting the objective function coefficient for z to any value $\mathcal{M} > M$ guarantees that for the problem (10), any optimal point $((y^*)^\top, z^*)^\top$ has $z^* = 1$ and y^* is an optimal solution of (9). Indeed, consider $\mathcal{M} = \mathcal{M}_\epsilon = M + \epsilon$ for some $\epsilon > 0$. Since $-1 \leq z^* \leq 1$ (and hence, $M(1 - z^*) \geq 0$), it then holds for any feasible point $(y^\top, z)^\top$ of (10) that

$$c^\top y + \mathcal{M}_\epsilon z \ \leq\ c^\top y^* + M(1 - z^*) + (M + \epsilon)z^* \ =\ c^\top y^* + M + \epsilon\, z^* \ \leq\ c^\top y^* + \mathcal{M}_\epsilon.$$

It is easily seen that the second inequality holds with equality if and only if $z^* = 1$, in which case the upper bound becomes sharp precisely for $y = y^*$ (i.e., an optimal solution of (9)).

Thus, (9) can be restated *equivalently* as (10), which is clearly an instance of ($D_1^\dagger$). It remains to note that with (say) $\mathcal{M} := M + 1$, all transformations described in the proof can clearly be performed in polynomial time, and the encoding length of (10) is polynomially bounded by that of (LP). Indeed, one can show that the Basis Pursuit instance

$$\min\ \|x\|_1 \quad \text{s.t.} \quad \left(I, \begin{pmatrix} \tilde{A}^\top \\ \delta^\top \end{pmatrix}\right)x = \begin{pmatrix} c \\ \mathcal{M} \end{pmatrix} \tag{11}$$

that is equivalent to (10) (via Lemma 1.1 and Remark 1.4) has encoding length $\mathcal{O}\big(nm(\langle A\rangle + \langle b\rangle) + \langle c\rangle + n^3 m\big) \subseteq \mathcal{O}\big(nm\langle\,(\text{LP})\,\rangle + n^3 m\big)$. The technical and quite lengthy proof of this estimate involves a series of estimations, or bounds, on encoding lengths, see [2] for the details. □

Remark 2.3 Clearly, if the LP contains variable bounds $\ell_j \leq x_j \leq u_j$, these can be used directly instead of $0 \leq x_j \leq K$. As mentioned earlier, the feasibility and boundedness assumptions regarding (LP) are not essential, by the next result:

Proposition 2.4 *Let some $\tilde{K} \geq K$ replace K in the construction of the instance* (10). *Then the following holds true:*

1. (LP) *is infeasible if and only if $b_i < \tilde{K}(a_i^\top \mathbb{1} - \|a_i^\top\|_1)/2$ for some $i \in [m]$ or $z^* < 1$ holds for every optimal solution $((y^*)^\top, z^*)^\top$ of* (10).
2. (LP) *is unbounded if and only if* (10) *constructed with $\tilde{K} > K$ has an optimal solution $((y^*)^\top, 1)^\top$ with $y_i^* = 1$ for some $i \in [m]$.*

P r o o f. Due to space limitations, we omit the proof and refer to [2, Prop. 3.19]. □

Remark 2.5 Due to the strong duality between (P_1) and ($D_1^\dagger$) (cf. Remark 1.4 and Lemma 1.1), the optimal objective function values opt_{BP} and opt_{LP} of (11) and (LP), respectively, are related as

$$c^\top x^* \;=\; \text{opt}_{LP} \;=\; \tfrac{K}{2}(\text{opt}_{BP} - \mathcal{M} + c^\top \mathbb{1}).$$

In contrast, an optimal *solution* x^* for (LP) itself can, in general, not be obtained directly (in closed form) from an optimal solution η^* for (11). Indeed, we first need to construct from η^* an optimal *dual* solution $((y^*)^\top, 1)^\top$ (for (10)) to get $x^* = \frac{K}{2}(y^* + \mathbb{1})$ which optimally solves (LP), see the beginning of the proof of Theorem 2.2. Nevertheless, a closed-form expression for y^* is in fact available in some situations (e.g., if the Basis Pursuit solution is sufficiently sparse); also, many ℓ_1-solvers are of a primal-dual type, i.e., compute a dual optimal solution as well. Moreover, one could employ an LP cross-over method relying only on an (approximate) primal-optimal solution, see, e.g., [8].

3 Concluding Remarks

Theorem 2.2 opens a possible way to obtain new asymptotical running time bounds for solving LPs, by combining the reduction with theoretical complexities of specialized Basis Pursuit solution methods. However, it seems very unlikely that this approach could improve the current best general bound $\mathcal{O}(n^3\langle\,(\text{LP})\rangle\,)$ for linear programming[1]: Roughly, an ℓ_1-solver would be required to perform fewer than $\mathcal{O}((n^2/m)\langle\,(11)\,\rangle)$ arithmetic operations; none of the known Basis Pursuit solvers are guaranteed to adhere to this bound in general (even if they converge finitely), and such an algorithm may very well not exist. Similarly, the Homotopy method sometimes needs only k iterations if the (unique) optimal Basis Pursuit solution has k nonzero entries and k is sufficiently small, so that if this was the case for the instance (11), then (LP) could in fact be solved in about $\mathcal{O}(nmk)$ time, cf. [9]. Unfortunately, this property can hardly be guaranteed a priori, since neither does one usually have prior knowledge of the solution sparsity for BP problems, nor is it clear what properties of the (LP) solution or instance a certain sparsity of solutions of (11) possibly translates into.

Generally, the reduction from (LP) to (11) will be numerically unstable in practice due to the large number M, for which no better a priori estimates may be known (as variable bounds $\ell \leq x \leq u$ are often available, the choice of K may be considered less problematic, cf. Remark 2.3). It is subject to future research to investigate whether the approach can be used for certain LP classes with particularly "nice" data such as, e.g., those arising from LP relaxation of combinatorial binary integer programs.

References

[1] D. A. Lorenz, M. E. Pfetsch, and A. M. Tillmann, ACM Transactions on Mathematical Software **41**(2), Art. No. 8 (2015).
[2] A. M. Tillmann, Computational Aspects of Compressed Sensing, Doctoral dissertation, TU Darmstadt, Germany, 2013.
[3] R. H. Bartels, A. R. Conn, and J. W. Sinclair, SIAM Journal of Numerical Analysis **15**(2), 224–241 (1978).
[4] E. J. Candès and T. Tao, IEEE Transactions on Information Theory **51**(12), 4203–4215 (2005).
[5] M. Grötschel, L. Lovász, and A. Schrijver, Geometric Algorithms and Combinatorial Optimization, 2nd edition, Algorithms and Combinatorics, Vol. 2 (Springer, Heidelberg, Germany, 1993).
[6] C. H. Papadimitriou and K. Steiglitz, Combinatorial Optimization: Algorithms and Complexity (Prentice-Hall, Upper Saddle River, NJ, USA, 1982).
[7] A. Schrijver, Theory of Linear and Integer Programming, Wiley-Interscience Series in Discrete Mathematics and Optimization (John Wiley & Sons, New York, NY, USA, 1986).
[8] R. E. Bixby and M. J. Saltzman, Operations Research Letters **15**(4), 169–178 (1994).
[9] D. L. Donoho and Y. Tsaig, IEEE Transactions on Information Theory **54**(11), 4789–4812 (2008).

[1] This bound assumes $m \leq n$ and integer LP data, and involves several technical subtleties (such as using $\mathcal{O}(\langle\,(\text{LP})\,\rangle)$-bit precision arithmetic to approximate a solution, then cross-over to an exact basic solution); for a more formal discussion, see [2] and the interior-point/LP references therein.