Equivalence of Minimal ℓ_0 and ℓ_p Norm Solutions of Linear Equalities, Inequalities and Linear Programs for Sufficiently Small p

G. M. FUNG glenn.fung@siemens.com
R&D Clinical Systems
Siemens Medical Solutions, Inc.
51 Valley Stream Parkway
Malvern, PA 19355

O. L. MANGASARIAN olvi@cs.wisc.edu
Computer Sciences Department
University of Wisconsin
Madison, WI 53706
Department of Mathematics
University of California at San Diego
La Jolla, CA 92093

Abstract. For a bounded system of linear equalities and inequalities we show that the NP-hard ℓ_0 norm minimization problem $\min_x \|x\|_0$ subject to $Ax = a$, $Bx \geq b$ and $\|x\|_\infty \leq 1$, is completely equivalent to the concave minimization $\min_x \|x\|_p$ subject to $Ax = a$, $Bx \geq b$ and $\|x\|_\infty \leq 1$, for a sufficiently small p. A local solution to the latter problem can be easily obtained by solving a provably finite number of linear programs. Computational results frequently leading to a global solution of the ℓ_0 minimization problem and often producing sparser solutions than the corresponding ℓ_1 solution are given. A similar approach applies to finding minimal ℓ_0 solutions of linear programs.

Keywords: ℓ_0 minimization, linear equations, linear inequalities, linear programming

1. Introduction

We consider the solvable system of linear equations in the real vector variable x:

$$Ax = a,\ Bx \geq b, \|x\|_\infty \leq 1, \tag{1}$$

where A is a given real matrix in $R^{m\times n}$, $B \in R^{k\times n}$, $a \in R^m$, $b \in R^k$ and $x \in R^n$. We note immediately that any bounded linear system such as $\tilde{A}y = a$, $\tilde{B}y \geq b$ and $\|y\|_\infty \leq \gamma$ with $\gamma > 0$, can be transformed to the above system (1) by defining $x = y/\gamma$, $A = \gamma\tilde{A}$, $B = \gamma\tilde{B}$ and consequently $\|x\|_\infty \leq 1$. We are here interested in finding a sparsest solution to (1), that is, one with the least number of nonzero components. This is equivalent to the following ℓ_0 norm minimization problem:

$$\min_x \ \|x\|_0 \ \text{ s.t. } \ Ax = a,\ Bx \geq b,\ \|x\|_\infty \leq 1, \tag{2}$$

where $\|x\|_0 = \sum_{i=1}^{n} sign(|x_i|)$. This problem has been studied extensively in the literature especially for linear equalities, for example [4, 6, 3], by relating it to the ℓ_1-minimization problem:

$$\min_{x} \|x\|_1 \text{ s.t. } Ax = a,\ Bx \geq b,\ \|x\|_\infty \leq 1. \tag{3}$$

Independently and predating the above approaches there have have been studies in the machine learning literature that utilized concave minimization for obtaining sparse linear and nonlinear classifiers [1, 9, 11] as well as in the approximation literature [2]. We shall use this concave minimization approach here to establish the fact that the ℓ_0 norm minimization problem is equivalent to the concave ℓ_p norm minimization problem for sufficiently small $p \leq 1$:

$$\min_{x} \|x\|_p \text{ s.t. } Ax = a,\ Bx \geq b,\ \|x\|_\infty \leq 1. \tag{4}$$

The advantage of this approach is that a linear-programming-based successive linearization algorithm (SLA) consisting of minimizing a linearization of a concave function on a polyhedral set is a *finitely* terminating stepless Frank-Wolfe algorithm [7]. In [9] finite termination of the SLA was established for a differentiable concave function, and in [10] for a nondifferentiable concave function using its supergradient.

We briefly describe the contents of our paper. In Section 2 we establish the equivalence between the ℓ_0 norm minimization problem (2) and the concave ℓ_p norm minimization problem (4) for sufficiently small p. We also show how to find minimal ℓ_0 norm solutions of linear programs. In Section 3 we state our SLA algorithm and establish its termination in a finite number of steps at a point satisfying the minimum principle necessary optimality condition. In Section 4 we present our computational results while Section 5 concludes the paper.

A word about our terminology and notation now. All vectors will be column vectors unless transposed to a row vector by a prime $'$. For a vector $x \in R^n$ the notation x_j will signify the j-th component and $\|x\|_p$ denotes the p-th norm of x for $p \in [0, \infty]$. The scalar (inner) product of two vectors x and y in the n-dimensional real space R^n will be denoted by $x'y$. The notation $A \in R^{m\times n}$ will signify a real $m \times n$ matrix. For such a matrix, A' will denote the transpose of A, A_i will denote the i-th row, $A_{\cdot j}$ the jth column and A_{ij} the ijth element. A vector of ones in a real space of arbitrary dimension will be denoted by e. Thus for $e \in R^n$ and $x \in R^n$ the notation $e'x$ will denote the sum of the components of x. A vector of zeros in a real space of arbitrary dimension will be denoted by 0. The star function $(\cdot)_*$ will signify the the sign function $sign(\cdot)$ which returns 1, 0 or -1, depending on whether the argument is positive, zero or negative. For a nonnegative vector t and a real number p, the expression t^p denotes a vector whose components are the components of t raised to the power p. The abbreviation "s.t." stands for "subject to". For simplicity in presenting our computational results we utilize the notation $\#(x)$ to denote the cardinality of a real vector x, that is the number of nonzero components of x. Thus $\#(x) = \|x\|_0$.

2. Equivalence of the ℓ_0 and the ℓ_p Norm Minimization Problems for Sufficiently Small p

We shall establish in this section the equivalence between the ℓ_0 norm minimization problem (2) and the concave ℓ_p norm minimization problem (4) for a sufficiently small p. Our proof is similar to that of [2, Theorem 2.1] where equivalence between a concave minimization problem and a general step function minimization is established.

We note first that the ℓ_0 norm (2) minimization problem can be stated in the following equivalent form, where the bound constraint $\|x\|_\infty \leq 1$ has been restated as $-e \leq x \leq e$ and the objective function replaced as follows:

$$\min_{(x,t)\in R^{2n}} e't_* \text{ s.t. } Ax = a,\ Bx \geq b,\ -t \leq x \leq t,\ -e \leq x \leq e. \tag{5}$$

Here, as defined in the Introduction, t_* denotes $\text{sign}(t)$. Note that since $t \geq 0$ it follows that each component of t_* is either 1 or 0 depending on whether the component of t is positive or zero. Hence $e't_* = \|t\|_0 \geq \|x\|_0$. Similarly, we can rewrite the concave minimization problem (4) in the following equivalent form:

$$\min_{(x,t)\in R^{2n}} e't^{\frac{1}{q}} \text{ s.t. } Ax = a,\ Bx \geq b,\ -t \leq x \leq t,\ -e \leq x \leq e, \text{ where } q = \frac{1}{p} \geq 1. \tag{6}$$

We define now a subset of the bounded feasible region of the two above problems as follows:

$$T = \{(x,t) \in R^{2n} \mid Ax = a,\ Bx \geq b,\ -t \leq x \leq t,\ -e \leq x \leq e,\ 0 \leq t \leq e\}, \tag{7}$$

where the inequalities $0 \leq t \leq e$ follow from the inequalities $-t \leq x \leq t,\ -e \leq x \leq e$, and the monotonicity of the objective functions of the minimization problems (5) and (6). We are now ready to state our principal result.

Proposition 1 Equivalence of the ℓ_0 and the ℓ_p Norm Minimization Problems (2) & (4) for Sufficiently Small p *The ℓ_0 norm minimization problem (2) is equivalent to the ℓ_p norm minimization problem (4) for some $p_0 \leq 1$. Furthermore, there exists a vertex of T that is an exact solution of the ℓ_0 norm minimization problem (2), equivalently (5),* **and** *is a global solution of the concave ℓ_p norm minimization problem (4), equivalently (6), for some $q_0 \in Q$ where:*

$$Q = \{1, 2, 3, \ldots\}, \tag{8}$$

and $p_0 = \frac{1}{q_0}$.

Proof: Note first that the objective function of (6) is concave for $t \geq 0$, $q \in Q$, and that:

$$0 \leq e't^{\frac{1}{q}} \leq e't_*, \text{ for } 0 \leq t \leq e. \tag{9}$$

Since $e't^{\frac{1}{q}} \geq 0$ on the bounded polyhedral set T, it follows by [12, Corollaries 32.3.3 and 32.3.4] that (6) has a vertex $(t(q), x(q))$ of T as a solution for each $q \in Q$. Since T has a finite number of vertices, one vertex, say $(\bar{x}, \bar{t})$, will repeatedly solve (6) for some increasing infinite sequence of real numbers in a subset $\bar{Q} = \{q_0, q_1, q_2, \ldots\}$ of Q. Hence for $q_i \in \bar{Q}$:

$$e'\bar{t}^{\frac{1}{q_i}} = e't(q_i)^{\frac{1}{q_i}} = \min_{(x,t)\in T} e't^{\frac{1}{q_i}} \leq \inf_{(x,t)\in T} e't_*, \tag{10}$$

where the last inequality above follows from (9). Letting $i \to \infty$, it follows from (10) that:

$$e'\bar{t}_* = \lim_{i\leftrightarrow\infty} e'\bar{t}^{\frac{1}{q_i}} \leq \inf_{(x,t)\in T} e't_*. \tag{11}$$

Since $(\bar{x},\bar{t}) \in T$, it follows that $(\bar{x},\bar{t})$ solves (5). Furthermore $(\bar{x},\bar{t})$ is a vertex of T. ■

We give now a simple example illustrating the above proposition.

Example 2 Example Illustrating Proposition 1 *Consider the linear system:*

$$\begin{array}{rcl} x_1+4x_2+4x_3+4x_4 &=& 1 \\ 2x_1+2x_2-4x_3+2x_4 &=& 2 \\ 4x_1+6x_2-6x_3+10x_4 &\geq& 4 \end{array} \tag{12}$$

It is easily checked that the minimal ℓ_0 norm solution is $x_1 = 1, x_2 = x_3 = x_4 = 0$, with $\|x\|_0 = 1$, $\|x\|_1 = 1$ and $\|x\|_\infty = 1$. The minimal ℓ_1 norm solution obtained by linear programming is $x_1 = 0, x_2 = 0.5, x_3 = -0.25, x_4 = 0$ with $\|x\|_0 = 2$, $\|x\|_1 = 0.75$ and $\|x\|_\infty = 0.5$. For this problem the minimal $\ell_{1/2}$ norm solution is the same as the minimal ℓ_0 norm solution.

The following remark shows how to obtain minimal ℓ_0 norm solutions of linear programs.

Remark 3 Minimal ℓ_0 Norm Solution of Linear Programs *We note that if we have the linear program*

$$\min_x c'x \text{ s.t. } Ax = a,\ Bx \geq b,\ \|x\|_\infty \leq 1, \tag{13}$$

with a solution, say $\bar{x}$, then finding a minimal ℓ_0 norm solution to this linear program can be stated as:

$$\min_x \|x\|_0 \text{ s.t. } Ax = a,\ Bx \geq b,\ c'x \leq c'\bar{x},\ \|x\|_\infty \leq 1, \tag{14}$$

which can be handled in a manner similar to that of problem (2).

We turn now to our finitely terminating successive linearization algorithm for obtaining a local solution to our concave minimization problem (6).

3. Finitely Terminating Successive Linearization Algorithm (SLA)

Our successive linearization algorithm consists of linearizing the differentiable concave function of the ℓ_p norm minimization problem (6) around a current point (x^i,t^i) and solving the resulting linear program. The algorithm terminates in a finite number of steps at a stationary point as we shall show after adding the constraint $t \geq e\delta$ to the minimization problem (6) for some small $\delta > 0$ to ensure the differentiability of the objective function of (6).

Algorithm 1 SLA: Successive Linearization Algorithm *Choose a $q \in Q$ and δ sufficiently small, typically $\delta = 1e-6$. Start with an $(x^0,t^0))$ that solves the ℓ_1 norm minimization problem (3). Having (x^i,t^i) determine (x^{i+1},t^{i+1}) by solving the following linear program:*

$$\min_{(x,t)\in T,\, t\geq e\delta} (t^i)'^{(\frac{1}{q}-1)}t. \tag{15}$$

Stop when

$$(t^i)'^{(\frac{1}{q}-1)}(t^i - t^{i+1}) = 0. \tag{16}$$

By [9, Theorem 4.2] we have the following finite termination result for the SLA algorithm.

Proposition 2 SLA Finite termination *The SLA 1 generates a finite sequence $\{(x^i, t^i)\}$ with strictly decreasing objective function values for the ℓ_p norm minimization problem (6) with $p = \frac{1}{q}$, and terminating at an $\bar{i}$ satisfying the following minimum principle necessary optimality condition [8, Theorem 9.3.3]:*

$$(t^{\bar{i}})'^{(\frac{1}{q}-1)}(t - t^{\bar{i}}) \geq 0, \ \forall\ (x,t) \in T \cap \{t \mid t \geq e\delta\}, \tag{17}$$

which states that the linearized objective function of (6) has a global minimum at $(x^{\bar{i}}, t^{\bar{i}})$.

We note that since the SLA terminates at a local solution which is not necessarily a global solution, we introduce the following heuristic into the SLA 1 which attempts to drive down large-valued components of $t^{\bar{i}}$ to δ while ignoring δ-valued components of $t^{\bar{i}}$ as follows.

Algorithm 3 SLA Heuristic *Once Algorithm 1 terminates at a local solution $(x^{\bar{i}}, t^{\bar{i}})$, rerun the linear program (13) while deleting δ-valued components of $(x^{\bar{i}}, t^{\bar{i}})$ from the objective function of (15) as well as the corresponding columns $A_{\cdot j}$ and $B_{\cdot j}$ from A and B in the definition of T (7). Continue with the new point $(x^{\bar{i}+1}, t^{\bar{i}+1})$ only if it has more δ-valued components than $(x^{\bar{i}}, t^{\bar{i}})$, else stop and declare the stationary point $(x^{\bar{i}}, t^{\bar{i}})$ as the final solution.*

We turn now to our computational results which utilize both the SLA Algorithm 1 as well as the SLA Heuristic Algorithm 3.

4. Computational Results

For all the experiments, we generated our problems by starting with random matrices of appropriate dimensions and a random solution of predetermined sparsity whose cardinality is denoted by $\#(L_0)$. For our experiments the sparsity of our initial random solution was set to 5% of the value of n (the number of features). Then we generated the appropriate right hand side for each problem, whether be it a system of linear equalities, linear inequalities or a linear program. One hundred instances were solved for each problem type and size. For all our experiments, the initial point for our SLA algorithm was generated by a linear program resulting from a minimal ℓ_1 norm solution. Since sparsity through an ℓ_1 regularization is commonly used and is considered a state-of-the-art for a wide range of applications, we present comparisons with this approach under the heading of L_1.

We tested our proposed SLA Algorithm 1 by performing experiments on three types of problems as follows.

1. For a system of linear inequalities $Bx \geq b$ we solved:

$$\min_x \ \|x\|_0 \ \text{ s.t. } \ Bx \geq b, \|x\|_\infty \leq 1.$$

by SLA to obtain a solution with a minimal number of nonzero components.

2. For the linear program (13) with a given solution $\bar{x}$ we used SLA to solve:

$$\min_{x} \|x\|_0 \text{ s.t. } Ax = a,\ Bx \geq b,\ c'x \leq c'\bar{x},\ \|x\|_\infty \leq 1,$$

which is the transformation given in (14). This problem is solved by SLA in a manner similar to the problem above.

3. Finally, for a system of linear equalities $Ax = a$ we solved:

$$\min_{x} \|x\|_0 \text{ s.t. } Ax = a, \|x\|_\infty \leq 1.$$

by using SLA to find a solution with a minimal number of nonzero components.

Experimental results for the three type of problems listed above are summarized in Tables 1, 2 and 3 where $\#(SLA))$ stands for the cardinality, and hence ℓ_0 norm, of the SLA solution. Similarly for the other # functions appearing in these three tables. For all problems considered, three values were used for the number of variables n (500,750,1000). For each value of n we considered three values for the number of rows m of equalities/inequalities which consisted of approximately 60%, 80% and 100% of the value of the corresponding n. This resulted in 9 different values for the pair (m,n).

The two first columns of each table show the values for m and n respectively while following columns show the number of times (out of a 100) that the cardinality (#) of the solution obtained by the SLA algorithm or the L_1 formulation, is equal or smaller than the original sparse L_0 solution used to generate the random problems.

Table (1) shows results for the linear inequality problem: $\min_x \|x\|_0$ s.t. $Bx \geq b, \|x\|_\infty \leq 1$. These results clearly show that our SLA approach outperforms the L_1 approach by producing sparser solutions with cardinality closer to the L_0 solution.

Similarly to the results obtained for the linear inequality case, Table (2) shows results for the sparse linear programming formulation problem (13). Again, the results clearly show that our SLA approach frequently outperforms the L_1 approach by producing sparser solutions with cardinality closer to that of the L_0 solution. We also note cases where the cardinality of the solution generated by the SLA approach was at most equal to the cardinality of the ℓ_1 solutions.

Table (3), shows interesting results for the problem $\min_x \|x\|_0$ s.t. $Ax = a, \|x\|_\infty \leq 1$, for a system of linear equalities. In this case, we can see that the cardinality of all the solutions (for all the (m,n) cases) were the same for the SLA algorithm and for the L_1 formulation for our randomly generated sparse ℓ_0 solution. This somewhat puzzling result is essentially justified by the following statement from [5, Corollary 1.5] and the way we generated our test problems:

"Let $y = Ax_0$, where x_0 contains nonzeros at k sites (fewer than $.49n$) selected uniformly at random, with signs chosen uniformly at random (amplitudes can have any distribution), and where A is a uniform random orthoprojector from R^n to R^m. With overwhelming probability for large n, the minimum 1-norm solution to $y = Ax$ is also the sparsest solution, and is precisely x_0."

Table 1. Results for the problem: $\min_x \|x\|_0$ s.t. $Bx \geq b, \|x\|_\infty \leq 1$. Columns show the number of times (out of a 100) that the cardinality of the solution obtained by the SLA algorithm or the L_1 formulation, is equal or smaller than the original sparse solution used to generate the random problems (L_0).

No. of Rows	No. of Variables	$\#(SLA) = \#(L_0)$	$\#(L_1) = \#(L_0)$	$\#(SLA) < \#(L_1)$	$\#(SLA) = \#(L_1)$
300	500	11	0	100	0
400	500	88	0	100	0
500	500	94	0	100	0
450	750	13	0	100	0
600	750	89	0	100	0
750	750	92	0	100	0
600	1,000	9	0	100	0
800	1,000	79	0	100	0
1,000	1,000	87	0	100	0

This statement is related to what is generally referred to as the ℓ_1/ℓ_0 equivalence, where under certain conditions the cardinality of the ℓ_1 and the ℓ_0 minimization problems are the same. Hence, these results confirm that both our SLA algorithm and the L_1 formulation coincide with an optimal global solution to the L_0 formulation.

5. Conclusion and Outlook

We have presented a new result which shows that the NP-hard problem of minimizing the ℓ_0 norm solution for linear equations, inequalities and linear programs, is equivalent to minimizing the ℓ_p norm solution for the same problems for sufficiently small p. Although the latter concave minimization problem is still NP-hard, a successive linearization algorithm applied to it terminates in a finite number of steps at a local solution which is often a global solution as indicated by the computational results presented. As such, the proposed equivalence has practical significance for the problem types presented, as well as for problems in other fields requiring sparsity such as machine learning and data mining. Hopefully these problems will be addressed in future work.

Acknowledgments

Research described here is available as Data Mining Institute Report 11-02, April 2011: ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/11-02.pdf.

Table 2. Results for the LP problem (13): $\min_x c'x$ s.t. $Ax = a,\ Bx \geq b,\ \|x\|_\infty \leq 1$. Columns show the number of times (out of a 100) that the cardinality of the solution obtained by the SLA algorithm or the L_1 formulation, is equal or smaller than the original sparse solution used to generate the random problems (L_0).

No. of Rows	No. of Variables	$\#(SLA) = \#(L_0)$	$\#(L_1) = \#(L_0)$	$\#(SLA) < \#(L_1)$	$\#(SLA) = \#(L_1)$
300	500	79	21	79	21
400	500	100	100	0	100
500	500	100	100	0	100
450	750	78	13	87	13
600	750	100	100	0	100
750	750	100	100	0	100
600	1,000	73	19	81	19
800	1,000	99	100	0	100
1,000	1,000	99	100	0	100

References

1. P. S. Bradley and O. L. Mangasarian. Feature selection via concave minimization and support vector machines. In J. Shavlik, editor, *Proceedings 15th International Conference on Machine Learning*, pages 82–90, San Francisco, California, 1998. Morgan Kaufmann. ftp://ftp.cs.wisc.edu/math-prog/tech-reports/98-03.ps.
2. P. S. Bradley, O. L. Mangasarian, and J. B. Rosen. Parsimonious least norm approximation. *Computational Optimization and Applications*, 11(1):5–21, October 1998. ftp://ftp.cs.wisc.edu/math-prog/tech-reports/97-03.ps.
3. E. Candès, M. B. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier Analysis and Applications*, 14:877–905, 2008.
4. E. J. Candés and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51:4203–4215, 2005.
5. D. L. Donoho. Neighborly polytopes and sparse solutions of underdetermined linear equations. Technical report, Department of Statistics, Stanford University, 2004. http://www-stat.stanford.edu/ donoho/Reports/2005/NPaSSULE-01-28-05.pdf.
6. D. L. Donoho. For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, 59:797–829, 2006. http://www-stat.stanford.edu/ donoho/Reports/2004/l1l0EquivCorrected.pdf.
7. M. Frank and P. Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3:95–110, 1956.
8. O. L. Mangasarian. *Nonlinear Programming.* SIAM, Philadelphia, PA, 1994.
9. O. L. Mangasarian. Machine learning via polyhedral concave minimization. In H. Fischer, B. Riedmueller, and S. Schaeffler, editors, *Applied Mathematics and Parallel Computing - Festschrift for Klaus Ritter*, pages 175–188. Physica-Verlag A Springer-Verlag Company, Heidelberg, 1996. ftp://ftp.cs.wisc.edu/math-prog/tech-reports/95-20.ps.

Table 3. Results for the problem: $\min_x \|x\|_0$ s.t. $Ax = a, \|x\|_\infty \leq 1$. Columns show the number of times (out of a 100) that the cardinality of the solution obtained by the SLA algorithm or the L_1 formulation, is equal or smaller than the original sparse solution used to generate the random problems (L_0).

No. of Rows	No. of Variables	$\#(SLA) = \#(L_0)$	$\#(L_1) = \#(L_0)$	$\#(SLA) < \#(L_1)$	$\#(SLA) = \#(L_1)$
300	500	100	100	0	100
400	500	100	100	0	100
500	500	100	100	0	100
450	750	100	100	0	100
600	750	100	100	0	100
750	750	100	100	0	100
600	1,000	100	100	0	100
800	1,000	100	100	0	100
1,000	1,000	100	100	0	100

10. O. L. Mangasarian. Solution of general linear complementarity problems via nondifferentiable concave minimization. *Acta Mathematica Vietnamica*, 22(1):199–205, 1997. ftp://ftp.cs.wisc.edu/math-prog/tech-reports/96-10.ps.
11. O. L. Mangasarian. Minimum-support solutions of polyhedral concave programs. *Optimization*, 45:149–162, 1999. ftp://ftp.cs.wisc.edu/math-prog/tech-reports/97-05.ps.
12. R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, New Jersey, 1970.