Unique solutions for l_1 problems

Lin Zhang, Joe Dellinger & Francis Muir

ABSTRACT

The l_p solution to a set of overdetermined linear equations is unique if $p > 1$, but may not be if $p = 1$. We resolve this ambiguity by defining an l_{1+} solution which is the limit of l_p solutions over p as $p \to 1$ from above. This solution is unique and in the l_1 solution set. We present a conceptual algorithm to find this solution, and illustrate the method with two examples.

INTRODUCTION

Claerbout and Muir (1973) have given numerous examples to show the advantages of the absolute value norm over least squares for solving overdetermined systems of linear equations. Renewed interest in travel-time inversion methods for statics and velocity estimation problems has encouraged the development of new techniques for computing l_p solutions. Scales et al. (1988) have reviewed recent work in this field. They describe a fast and stable iterative l_p algorithm for sparse systems and apply it to seismic travel time tomography using $p = 1$.

l_1 solutions

The l_1 norm is robust in the presence of large data errors, and is thus the norm of choice for certain kinds of problems. But an objective function based on this norm does not necessarily have a unique minimal solution. From a computational point of view this is troublesome. Among all the possible solutions we do not know which we are going to get, and different versions of the same algorithm may produce different solutions. Although all the solutions give the same minimum error measure, it is reasonable to redefine the problem by adding extra conditions and come up with an unambiguous solution.

l_{1^+}: the end of the line

We choose the limit of l_p solutions as p goes to 1 from above. This limiting solution exists, is unique, and lies inside the original l_1 solution set. Dellinger (1984) worked on this notion in the one-dimensional case and used it to compute the unique median of a set of real numbers. He showed that the solution can be obtained first by finding the line segment that covers the set of absolute value norm solutions, and then solving a non-linear equation in that segment. We extend this result to many dimensions. We first find the convex region defining the l_1 solution set, and then solve a set of non-linear equations within that region.

Organization

We organize this paper in the following way. First we review l_p problems. Then we discuss the convexity of the objective functions and give a formal definition of the limiting solution. Then we discuss the uniqueness and existence of the solution, and follow this with a conceptual algorithm. Finally, we illustrate the method with two examples.

REVIEW

In seismic data processing, many problems can be formulated as a set of over-determined linear equations

$$\mathbf{A}\mathbf{x} = \mathbf{d}, \tag{1}$$

where

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ . & . & \dots & . \\ . & . & \dots & . \\ . & . & \dots & . \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix},$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_m \end{bmatrix} \quad \text{and} \quad \mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ . \\ . \\ . \\ d_n \end{bmatrix}.$$

The error vector is

$$\mathbf{e} = \mathbf{d} - \mathbf{A}\mathbf{x}. \tag{2}$$

The solution is the vector $\mathbf{x}$ which minimizes the magnitude of the error vector as measured by some well-defined objective function. The exact concept of what a large error is depends on the particular objective function used. The objective function should have a single global minimum in order for the problem to have a unique solution.

The definitions

Let $\mathbf{v}$ be a vector in an n-dimensional space. For any $p \geq 1$

$$E_p(\mathbf{v}) = \left(\sum_{i=1}^{n} |v_i|^p \right)^{\frac{1}{p}} \tag{3}$$

defines the l_p norm of $\mathbf{v}$, where v_i is the ith component of the vector $\mathbf{v}$.

l_p optimization problems

The objective function is often chosen to be the l_p norm of the error vector. The solution to such an l_p optimization problem is defined as

$$\mathbf{x}_p := \{ \mathbf{x} \mid \min_{\mathbf{x}} E_p(\mathbf{e}) \} \tag{4}$$

for $p \geq 1$.

For different p values, the objective function emphasizes different aspects of the error vector. $p = 2$, least squares, is seemingly ubiquitous in seismic data processing, but the extremal cases, $p = \infty$ and $p = 1$, find use for certain problems. The l_∞ norm measures the maximum error. The l_1 norm is especially good for problems with erratic data because of its robust properties. As previously stated, there is a problem in using the l_1 norm because the objective function does not always define a unique solution.

The non-uniqueness can be removed by adding extra constraints. We choose a solution that is equal to the limit of $\mathbf{x}_p$ as $p \to 1$ from above. Mathematically

$$\mathbf{x}_{1^+} := \lim_{p \to 1^+} \mathbf{x}_p, \tag{5}$$

where $\mathbf{x}_p$ is defined in equation (4). We call this the limiting solution, and will show that this limiting solution is in the l_1 solution set. It should be remembered that the limit must be taken after the minimization.

Convexity of objective functions

Before discussing the existence and uniqueness of the solution, we will state several results on the convexity of objective functions.

For $p = 1$, $E_1(\mathbf{e})$ is a convex upwards function of $\mathbf{x}$. Solutions always exist but they may not be unique. All the solutions, however, lie in one convex region. This means that any linear combination of solutions is also a solution.

If the system is over-determined, i.e. the column rank of matrix $\mathbf{A}$ is full and there is no precise solution because the data are inconsistent, then the objective function $E_p(\mathbf{e})$ for any $p > 1$ is always a strictly convex upwards function. We will always have a unique solution for l_p problems for every $p > 1$. To show the existence and uniqueness of the limiting solution, the remaining question is whether the limit in equation (5) exists.

LIMITING SOLUTION

For any strictly convex upwards function, the necessary and sufficient condition for a point to be the global minimum is that all its partial derivatives are continuous and zero at this point. Clearly, according to the conventional definition, the partial derivatives of our objective function for $p = 1$ do not always exist. We need to have a special definition of derivative so that we can always compute the derivatives of our objective function for any $p \geq 1$.

The directional derivative

The *directional derivative* of $f(\mathbf{x})$ at $\mathbf{x}$ in the direction $\mathbf{u}, where |\mathbf{u}| = 1$, is defined as

$$f_{\mathbf{u}}(\mathbf{x}) = \lim_{h \to 0^+} \frac{f(\mathbf{x} + h\mathbf{u}) - f(\mathbf{x})}{h}. \tag{6}$$

Now we state the theorem, which is easy to prove.

Necessary and sufficient conditions

Let $f(\mathbf{x})$ be a strictly convex upwards and continuous function in the domain $\mathcal{R}^n$. Then $\mathbf{x}$ is a global minimum if and only if the directional derivatives of $f(\mathbf{x})$ at $\mathbf{x}_0$ in all directions are ≥ 0.

Existence of the limit

We rewrite definition (3) into the form of vector components:

$$E_p(\mathbf{e}) = \left(\sum_{i=1}^{n} \left| d_i - \sum_{j=1}^{m} a_{ij} x_j \right|^p \right)^{\frac{1}{p}}. \tag{7}$$

The directional derivatives of $E_p(\mathbf{e})$ exist for any $p \geq 1$ and are

$$[E_p(\mathbf{e})]_{\mathbf{u}} = \left[\sum_{i=1}^{n} \left| d_i - \sum_{j=1}^{m} a_{ij} x_j \right|^p \right]^{\frac{1}{p}-1} \sum_{i=1}^{n} \left(\sum_{j=1}^{m} -a_{ij} u_j \right)$$
$$\lim_{h \to 0^+} \left| d_i - \sum_{j=1}^{m} a_{ij} (x_j + h u_j) \right|^{p-1} \operatorname{sgn} \left(d_i - \sum_{j=1}^{m} a_{ij} (x_j + h u_j) \right). \tag{8}$$

The theorem states $\mathbf{x}_p$ is a minimum of $E_p(\mathbf{e})$ if and only if, for any unit vector $\mathbf{u}$,

$$\sum_{i=1}^{n}\left(\sum_{j=1}^{m}-a_{ij}u_j\right)\lim_{h\to 0^+}\operatorname{sgn}\left(d_i-\sum_{j=1}^{m}a_{ij}\left(x_{jp}+hu_j\right)\right)\geq 0 \qquad (9.a)$$

for $p=1$, and

$$\sum_{i\in R}\left(\sum_{j=1}^{m}-a_{ij}u_j\right)\left|d_i-\sum_{j=1}^{m}a_{ij}x_{jp}\right|^{p-1}\operatorname{sgn}\left(d_i-\sum_{j=1}^{m}a_{ij}x_{jp}\right)\geq 0 \qquad (9.b)$$

for $p>1$, where

$$R=\left\{\, i \;\middle|\; d_i\neq\sum_{j=1}^{m}a_{ij}x_{jp}\,\right\}.$$

From the convexity of $E_p(\mathbf{e})$ we know that for every $p>1$ there always exists a unique $\mathbf{x}_p$ such that the above inequality holds. We show that $\mathbf{x}_p$ is a continuous function of p for $p>1$ by the following argument:

Suppose δ is a small real number such that $p+\delta>1$. Let $\mathbf{x}_{p+\delta}$ be the solution using the norm $l_{p+\delta}$. Then

$$\sum_{i\in R}\left(\sum_{j=1}^{m}-a_{ij}u_j\right)\left|d_i-\sum_{j=1}^{m}a_{ij}x_{jp+\delta}\right|^{p+\delta-1}\operatorname{sgn}\left(d_i-\sum_{j=1}^{m}a_{ij}x_{jp+\delta}\right)\geq 0. \qquad (10)$$

Now let $\delta\to 0^+$. The limit of the left hand side should also be greater or equal to zero, and therefore

$$\sum_{i\in R}\left(\sum_{j=1}^{m}-a_{ij}u_j\right)\left|d_i-\sum_{j=1}^{m}a_{ij}x_{jp^+}\right|^{p-1}\operatorname{sgn}\left(d_i-\sum_{j=1}^{m}a_{ij}x_{jp^+}\right)\geq 0. \qquad (11)$$

Thus $\mathbf{x}_{p^+}$ is also a solution using the norm l_p. By the same argument, we can show that $\mathbf{x}_{p^-}$ is also a solution. We conclude that

$$\mathbf{x}_{p^+}=\mathbf{x}_{p^-}=\mathbf{x}_p, \qquad (12)$$

because of the uniqueness of the solution. By Cauchy's convergence condition, $\mathbf{x_p}$ has a limit as $p\to 1^+$. So the limiting solution exists and is unique.

Where is the limiting solution?

We will show that the limiting solution is in the region where the absolute value norm $E_1(\mathbf{e})$ reaches its minimum value. That is, the limiting solution is also a solution for the l_1 norm.

The limiting solution satisfies the inequality (9.b) as $p\to 1^+$,

$$\sum_{i\in R}\left(\sum_{j=1}^{m}-a_{ij}u_j\right)\operatorname{sgn}\left(d_i-\sum_{j=1}^{m}a_{ij}x_{j1^+}\right)\geq 0, \qquad (13)$$

while

$$\begin{aligned}
&\sum_{i=1}^{n}\left(\sum_{j=1}^{m}-a_{ij}u_j\right)\lim_{h\to 0^+}\operatorname{sgn}\left(d_i-\sum_{j=1}^{m}a_{ij}\left(x_{j1^+}+hu_j\right)\right)\\
=\;&\sum_{i\in R}\left(\sum_{j=1}^{m}-a_{ij}u_j\right)\operatorname{sgn}\left(d_i-\sum_{j=1}^{m}a_{ij}x_{j1^+}\right)+\sum_{i\notin R}\left|\sum_{j=1}^{m}-a_{ij}u_j\right|\\
\geq\;&0. \qquad (14)
\end{aligned}$$

This shows that $\mathbf{x}_{1^+}$ satisfies inequality (9.a) and thus that it is a solution for the l_1 norm.

A conceptual algorithm

The limiting solution is somewhere in the convex region B where the l_1 norm achieves its minimum value. Generally B is in an $\hat{m}$-dimensional subspace of an m-dimensional space. Within B, $\mathbf{x}$ has only $\hat{m}$ independent components. The rest of the components are linear combinations of these components. Without loss of generality, we assume the first $\hat{m}$ components of $\mathbf{x}$ are independent in B. Then we have

$$\begin{bmatrix} x_{\hat{m}+1}\\ x_{\hat{m}+2}\\ \cdot\\ \cdot\\ \cdot\\ x_m \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1\hat{m}}\\ c_{21} & c_{22} & \dots & c_{2\hat{m}}\\ \cdot & \cdot & \dots & \cdot\\ \cdot & \cdot & \dots & \cdot\\ \cdot & \cdot & \dots & \cdot\\ c_{n1} & c_{n2} & \dots & c_{n\hat{m}} \end{bmatrix} \begin{bmatrix} x_1\\ x_2\\ \cdot\\ \cdot\\ \cdot\\ x_{\hat{m}} \end{bmatrix} + \begin{bmatrix} h_1\\ h_2\\ \cdot\\ \cdot\\ \cdot\\ h_{\hat{m}} \end{bmatrix}, \qquad (15)$$

where c_{ij} and h_j are constants.

Substituting these relations into equation (7), we have

$$E_p\left(\mathbf{e}\right)=\left(\sum_{i=1}^{n}\left|\hat{d}_i-\sum_{j=1}^{\hat{m}}\hat{a}_{ij}x_j\right|^p\right)^{\frac{1}{p}}, \qquad (16)$$

where

$$\hat{d}_i=d_i-\sum_{k=1}^{m-\hat{m}}a_{i\hat{m}+k}h_k \quad\text{and}\quad \hat{a}_{ij}=a_{ij}-\sum_{k=1}^{m-\hat{m}}a_{i\hat{m}+k}c_{kj}.$$

So the limiting solution of equation (1) is also the limiting solution of the new linear equation set in the convex region B.

We know that throughout B, $E_1(\mathbf{e})$ is equal to its minimum value, therefore

$$\frac{\partial E_1(\mathbf{e})}{\partial x_j}=\sum_{i=1}^{n}\left(-\hat{a}_{ik}\right)\operatorname{sgn}\left(\hat{d}_i-\sum_{j=1}^{\hat{m}}\hat{a}_{ij}x_j\right)\equiv 0 \qquad (17)$$

for each $k=1,2,\ldots,\hat{m}$. Clearly for each i the sign function is constant within B. Therefore $E_p(\mathbf{e})$ has continuous partial derivatives for each $p\geq 1$. We conclude that

$\mathbf{x}_p$ is the solution of the l_p norm in B if and only if all these partial derivatives are equal to zero. Thus

$$\sum_{i=1}^{n}(-\hat{a}_{ik})\left|\hat{d}_i - \sum_{j=1}^{\hat{m}} \hat{a}_{ij}x_j\right|^{p-1} \operatorname{sgn}\left(\hat{d}_i - \sum_{j=1}^{\hat{m}} \hat{a}_{ij}x_j\right) = 0. \tag{18}$$

As Dellinger did for finding the median of any set of numbers, we can show that the limiting solution $\mathbf{x}_{1^+}$ satisfies the nonlinear equations

$$\prod_{i\in R^+}\left(\hat{d}_i - \sum_{j=1}^{\hat{m}} \hat{a}_{ij}x_{j1^+}\right)^{\hat{a}_{ik}} = \prod_{i\in R^-}\left(\hat{d}_i - \sum_{j=1}^{\hat{m}} \hat{a}_{ij}x_{j1^+}\right)^{\hat{a}_{ik}}, \tag{19}$$

for $k = 1, 2, \ldots, \hat{m}$; where

$$R^+ = \left\{\; i \;\middle|\; \hat{d}_i > \sum_{j=1}^{\hat{m}} \hat{a}_{ij}x_{j1^+} \right\},$$

and

$$R^- = \left\{\; i \;\middle|\; \hat{d}_i < \sum_{j=1}^{\hat{m}} \hat{a}_{ij}x_{j1^+} \right\}.$$

Now we have a conceptual algorithm:

1. Find the convex region B within which there are $\hat{m}$ independent components. Express the dependent components as linear combinations of these independent components.

2. Substitute these relations into equations (16) to obtain a new linear equation set.

3. Solve the nonlinear equation set (19) to obtain $\hat{m}$ independent components.

4. Use equation (15) to determine the rest of the components.

EXAMPLES

Let us look at two simple examples for the 2D case.

Example 1

We have equations

$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

The convex region B is defined by

$$B = \left\{ (x,y) \;\middle|\; x = \frac{1}{2}, \frac{1}{4} < y < \frac{1}{2} \right\}.$$

Substituting $x = 1/2$ into the equation we obtain

$$\begin{bmatrix} 1 \\ 1 \\ 2 \\ 0 \end{bmatrix} [\, y \,] = \begin{bmatrix} 1 \\ \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{bmatrix}.$$

Equation (19) gives

$$\left(2y - \frac{1}{2}\right)^2 = \left(\frac{1}{2} - y\right)(1 - y).$$

Solving this equation, we obtain two possible roots,

$$y = \frac{1 \pm \sqrt{13}}{12},$$

but only the positive one is in the region B. So we have the limiting solution

$$\begin{aligned} x &= \frac{1}{2} \\ y &= \frac{1 + \sqrt{13}}{12}. \end{aligned}$$

Figure 1 shows contours of $E_1(\mathrm{e})$ and the limiting solution for this example.

Example 2

We have equations

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & -1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

The convex region B is defined by

$$B = \left\{ (x,y) \;\middle|\; \frac{y}{2} < x < \min\{2y, 1\} \text{ and } \frac{x}{2} < y < \min\{2x, 1\} \right\}.$$

Equation (19) gives

$$\begin{aligned} (2x - y)^2 &= (2y - x)(1 - x) \\ (2x - y)^{-1} &= (2y - x)^{-2}(1 - y). \end{aligned}$$

Solving these equations, we obtain the limiting solution,

$$\begin{aligned} x &= \frac{1}{2} \\ y &= \frac{1}{2}. \end{aligned}$$

FIG. 1. Contours of the objective function $E_1(\mathbf{e})$ from example 1. The fat line segment shows the convex region B. The small square indicates the location of the limiting solution within B.

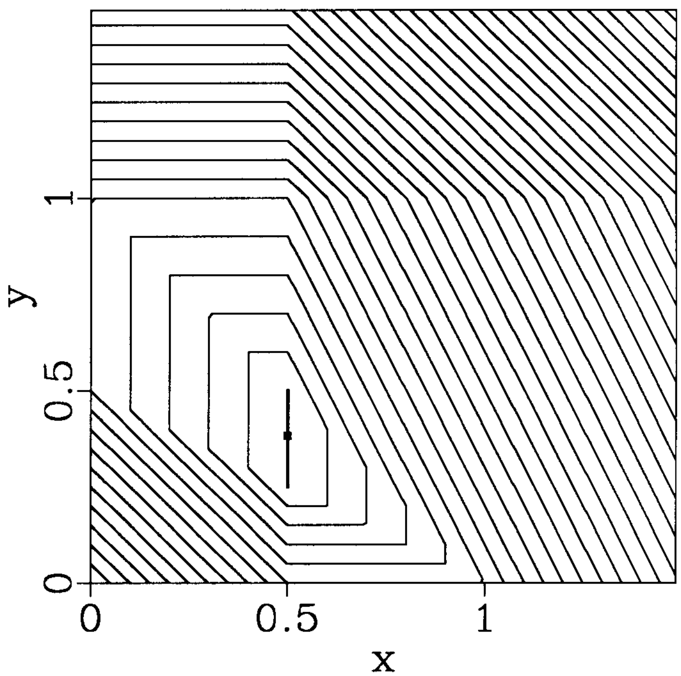

FIG. 2. Contours of the objective function $E_1(\mathbf{e})$ from example 2. The stippled area bordered by a thick line shows the convex region B, and the black square within it indicates the location of the limiting solution.

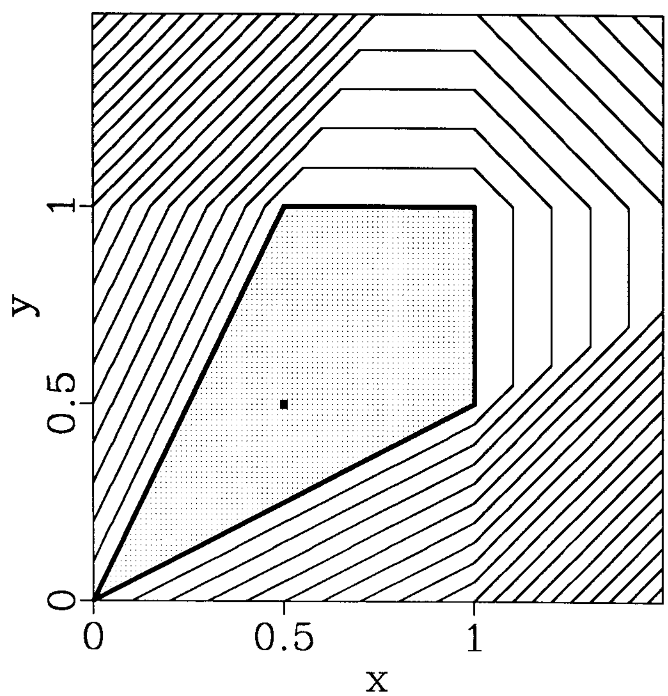

Figure 2 shows contours of $E_1(\mathbf{e})$ and the limiting solution for this example.

SUMMARY

In this paper, we have shown that the limiting solution we defined is a solution for the l_1 optimization problem. It is unique and always exists. We can compute this solution by finding the region where the absolute value norm is minimum and then solving a set of nonlinear equations in this region. The algorithm given is conceptual, so work remains to be done to develop a practical algorithm.

REFERENCES

Claerbout, J.F. and Muir, F., 1973, Robust modeling with erratic data: Geophysics, **38**, 826-844.

Dellinger J., 1984, What is the median of {1,2,3,5} ? : SEP–**41**.

Scales J., Gersztenkorn A. and Treitel S., 1988, Fast l_p solution of large, sparse, linear systems: J. Comp. Phys., **75**, 314–333.