



Overdetermined Systems of Linear Equations

OSLE

MUSTAFA Ç. PİNAR
Bilkent University, Ankara, Turkey

MSC2000: 65K05, 65D10

Article Outline

[Keywords](#)
[See also](#)
[References](#)

Keywords

Parameter estimation; Optimization; Approximation

Consider a set of experimental observations represented by a vector $b \in \mathbf{R}^n$. The goal is to estimate a set of parameters $x \in \mathbf{R}^m$ with the help of a matrix of independent input conditions represented by $A \in \mathbf{R}^{n \times m}$. In other words, one wishes to express b in terms of A . However, one may have a larger number of experimental observations than parameters to be estimated, i. e., it may be the case that $n \gg m$. The problem described above is a typical estimation problem which gives rise to an overdetermined system of linear equations:

$$Ax = b. \quad (1)$$

In general one cannot expect to obtain a vector x which satisfies (1) even if A has m linearly independent columns. This feature of the problem leads to the search for a vector x which makes Ax as close as possible to b . The closeness is measured in some suitable norm which is usually either the 2-norm, or the 1-norm, or the ∞ -norm. The most common is the 2-norm which yields the well-known linear *least squares problem*:

$$\min_x \|Ax - b\|_2 = \sqrt{(Ax - b)^\top (Ax - b)}. \quad (2)$$

The linear least squares approach is usually preferred because it leads to a simpler problem. More precisely, it admits a closed-form solution which can be obtained by solving the linear system of equations:

$$A^\top Ax = 2A^\top b. \quad (3)$$

Since $A^\top A$ is a symmetric positive (semi)definite matrix (it is positive definite when A has m linearly independent columns, in which case the solution is unique) it can be decomposed in the form of LDL^\top (or, Choleski factorization) where L is unit lower triangular, and D is diagonal. The factored form can then be used to solve (3) which has always a solution. However, this method is only reliable when A is a well-conditioned matrix. A more numerically stable way to solve (3) is to use an orthogonal factorization (e. g., QR) combined with a pivoting strategy. A detailed treatment of the linear least squares problem can be found in [8].

In some instances, the set of observations includes gross inaccuracies or wild points. In such cases, it may be preferable to use the 1-norm which leads to the following estimation problem

$$\min_x \|Ax - b\|_1 = \sum_{i=1}^n |(Ax - b)_i|, \quad (4)$$

where $(Ax - b)_i$ is used to represent the i th component of $Ax - b$. The function in (4) is not differentiable at those points where $(Ax - b)_i = 0$ for some $i \in \{1, \dots, n\}$. The problem is commonly referred to as the ℓ_1 *estimation problem*. The parameter values obtained from the minimization problem (4) will not be as adversely affected by the presence of wild points as the estimates obtained using (3). On the other hand, in contrast to the linear least squares problem (4) is a combinatorial optimization problem because it can be shown that a minimizing point x has the property that some of the components of the *residual vector* $Ax - b$ are equal to zero, some are positive and some are negative (this property is what makes this approach immune to wild points). Hence, if one had access to the information as to which components are zero, positive, and negative, respectively, one could find a minimizing point x by solving the following linear program:

$$\begin{cases} \min_x & \sum_{i \in \mathcal{A}^c} s_i^* (Ax - b)_i \\ \text{s.t.} & (Ax - b)_i = 0, \quad \forall i \in \mathcal{A}, \end{cases}$$

where \mathcal{A} is the set of indices corresponding to zero components of $Ax - b$, \mathcal{A}^c is its complement with respect to $\{1, \dots, n\}$, and s_i^* is the *sign function* which assumes the value $+1$ for positive residuals, and -1 for negative residuals, respectively.



Unfortunately, one has a priori no idea about s_i^* and A . An alternative way to pose (4) leads to the following problem:

$$\begin{cases} \min_{\substack{c \geq 0 \\ d \geq 0 \\ u \geq 0 \\ v \geq 0}} & \sum_{i=1}^n (u_i + v_i) \\ \text{s.t.} & u - v + A(c - d) = b, \end{cases}$$

with $(Ax - b)_i = u_i + v_i$ and $x_j = c_j - d_j$. The equivalence of (4) and the above linear program is discussed in [17]. The most successful attempts at solving (4) were based on the above reformulation and its dual problem. Notably, I. Barrodale and F.D.K. Roberts [4] specialized the simplex algorithm of linear programming to the above formulation by taking advantage of the complementarity between the u_j and v_j variables in the pivoting process. R.D. Armstrong, E.L. Frome and D.S. Kung [1] developed a revised simplex algorithm for the linear programming formulation of the problem. A different algorithm which aims at minimizing the nondifferentiable 1-norm function (4) was given in [6].

A more recent idea for solving the ℓ_1 estimation problem was given in [12]. This idea is quite different from those mentioned above in that it replaces the original function with a once continuously differentiable function, and leads to the following problem:

$$\min_x \sum_{i=1}^n \rho((Ax - b)_i), \tag{5}$$

where

$$\rho(t) = \begin{cases} \frac{t^2}{2\gamma} & \text{if } |t| \leq \gamma, \\ |t| - \frac{\gamma}{2} & \text{if } |t| > \gamma, \end{cases} \tag{6}$$

with t being a knock-off variable, and γ a positive scalar. This function is known as *Huber's M-estimator function* ([9]) in the statistics literature as it was introduced by P.J. Huber as a *robust estimator* in the face of *inaccuracies in the observations*. K. Madsen and H.B. Nielsen observed that they can obtain a solution (4) by repeatedly solving (5) for decreasing values γ tending to zero. They were also able to avoid the potentially ill-conditioning effects of driving γ to zero.

As far as obtaining a set of parameters 'immune' to grossly inaccurate observations, one has the option to

use the 1-norm or (4), or the Huber problem (5). It is interesting that Huber's problem was used as a subproblem to solve (4). The relationship between problems (4) and (5) were further explored in [13] and [11].

Another popular choice for the solution of overdetermined systems of linear equations is to compute a solution to minimize the ∞ -norm of the residual vector. This approach yields the problem

$$\min_x \max_i |(Ax - b)_i| . \tag{7}$$

The problem is commonly known as the *Chebyshev problem*. Here, one faces again a problem of a combinatorial nature as it can be proved that a solution to (7) has certain residual values equal to the maximum in absolute value, and others smaller than this value in modulus, respectively. This partition of the residuals at a minimizing point is obviously unknown. Hence, one must resort to some algorithm to compute a solution to (7) much the same way as in the case of (4). Here, again there exist approaches based on minimizing the nondifferentiable function in (7) (nondifferentiable at points where at least two residuals attain the maximum value in modulus). The most notable of such methods are that of Bartels–Golub [7], Bartels–Conn–Charalambous [5]. There exist also methods based on the linear programming formulation which is given as follows in [17]:

$$\begin{cases} \min_{x,z} & z \\ \text{s.t.} & -z \leq (Ax - b)_i \leq z, \quad \forall i = 1, \dots, n. \end{cases}$$

Some of the approaches based on linear programming favored the above primal formulation for use in a penalty function algorithm [10,15]. Some others used the dual formulation:

$$\begin{cases} \min_{\substack{v \geq 0 \\ w \geq 0}} & (v - w)^T b \\ \text{s.t.} & A^T (v - w) = 0 \\ & e^T (v + w) = 1, \end{cases}$$

where e represents a vector of all ones. Among these approaches, the most successful is the simplex adaptation of [2].

A survey of the use of the 2-norm, 1-norm and ∞ -norm criteria in linear regression in statistics is given in [14], but contains only developments until 1981.

Some of the algorithms mentioned above are available as software packages. In particular, the 1-norm algorithms of Barrodale–Roberts and of Bartels–Conn–Sinclair are available in the NAG (Numerical Algorithms Group) software library. The 1-norm and Huber algorithms of Madsen–Nielsen are available from the authors. The Chebyshev algorithm of Barrodale–Phillips is available in the NAG library, and also in the ACM collection [3]. The Chebyshev algorithm of Pinar–Elhedhli is available from the authors. A copy of the Bartels–Golub algorithm for the Chebyshev problem can be obtained from [16].

See also

- ▶ [ABS Algorithms for Linear Equations and Linear Least Squares](#)
- ▶ [Cholesky Factorization](#)
- ▶ [Interval Linear Systems](#)
- ▶ [Large Scale Trust Region Problems](#)
- ▶ [Large Scale Unconstrained Optimization](#)
- ▶ [Linear Programming](#)
- ▶ [Nonlinear Least Squares: Trust Region Methods](#)
- ▶ [Orthogonal Triangularization](#)
- ▶ [QR Factorization](#)
- ▶ [Solving Large Scale and Sparse Semidefinite Programs](#)
- ▶ [Symmetric Systems of Linear Equations](#)

References

1. Armstrong RD, Frome EL, Kung DS (1979) A revised simplex algorithm for the absolute deviation curve fitting problem. *Comm Statist–Simula Computa* 8:175–190
2. Barrodale I, Phillips C (1974) An improved algorithm for discrete Chebyshev linear approximation. In: Hartnell BL, Williams HC (eds) *Proc. Fourth Manitoba Conf. Numer. Math. Utilitas Math. Pub.*, pp 177–190
3. Barrodale I, Phillips C (1975) Algorithm 495: solution of an overdetermined system of linear equations in the Chebyshev norm. *ACM Trans Math Softw* 1:264–270
4. Barrodale I, Roberts FDK (1973) An improved algorithm for discrete linear ℓ_1 approximation. *SIAM J Numer Anal* 10:839–848
5. Bartels RH, Conn AR, Charalambous C (1978) On Cline’s direct method for solving overdetermined linear systems in the ℓ_∞ sense. *SIAM J Numer Anal* 15:255–270
6. Bartels RH, Conn AR, Sinclair JW (1978) Minimisation techniques for piecewise differentiable functions: the ℓ_1 solution to an overdetermined linear system. *SIAM J Numer Anal* 15:224–241
7. Bartels RH, Golub GH (1968) Stable numerical methods for obtaining the Chebyshev solution to an overdetermined system of equations. *Comm ACM* 11:401–406
8. Björk A (1996) *Numerical methods for least squares problems*. SIAM, Philadelphia
9. Huber PJ (1981) *Robust statistics*. Wiley, New York
10. Joe B, Bartels RH (1983) An exact penalty method for constrained, discrete, linear ℓ_∞ data fitting. *SIAM J Sci Comput* 4:69–84
11. Li W, Swetits J (1998) Linear ℓ_1 estimator and Huber M-estimator. *SIAM J Optim* 8
12. Madsen K, Nielsen HB (1993) A finite smoothing algorithm for linear ℓ_1 estimation. *SIAM J Optim* 3:223–235
13. Madsen K, Nielsen HB, Pinar MČ (1994) New characterizations of ℓ_1 solutions to overdetermined systems of linear equations. *Oper Res Lett* 16:159–166
14. Narula SC (1982) Optimization techniques in linear regression: A review. *Optimization in Stat. In: TIMS/Studies Management Sci*, vol 19, pp 11–29
15. Pinar MČ, Elhedhli S (1998) A penalty continuation method for the ℓ_∞ solution of overdetermined linear systems. *BIT* 38:127–150
16. Schryer N (1998) via G.H. Golub, Email: golub@sccm.stanford.edu
17. Watson GA (1980) *Approximation theory and numerical methods*. Wiley, New York