Formulation of L_1 Norm Minimization in Gauss-Markov Models

AliReza Amiri-Simkooei[1]

Abstract: L_1 norm minimization adjustment is a technique to detect outlier observations in geodetic networks. The usual method for implementation of L_1 norm adjustment leads to the solving of a linear programming problem. In this paper, the formulation of the L_1 norm minimization for a rank deficient Gauss-Markov model will be presented. The results have been tested on both linear and non-linear models, which confirm the efficiency of the suggested formulation.

DOI: 10.1061/(ASCE)0733-9453(2003)129:1(37)

CE Database keywords: Tendons; Creep rupture; Relaxation, mechanics; Fiber reinforced polymers; Models; Linear programming; Measurement; Geodetic surveys.

Introduction

The method of least squares, which gives a best solution by minimizing the sum of squares of weighted discrepancies between measurements, is one of the most frequently used methods to obtain unique estimates for unknown parameters from a set of redundant measurements. It is important to note that the implementation of the method of least squares does not require the knowledge of the distribution from which the observations are drawn for the purpose of parameters estimation. It can be shown, however, that if the weight matrix is chosen to be the inverse of the covariance matrix of the observations, the least squares estimate is an *unbiased* and *minimum variance* estimate. If, in addition, the observation errors are normally distributed, the least squares method will give an identical solution vector to those from the maximum likelihood method. The weighted least squares (L_2 norm minimization) method states that the sum of the squares of the weighted residuals $\mathbf{v}$ should be minimal (Mikhail 1976)

$$\mathbf{v}^{\mathbf{T}}\mathbf{P}\mathbf{v}=\sum_{i=1}^{n} p_i v_i^2 \rightarrow \min \tag{1}$$

where $\mathbf{P}$ = weight matrix of observations.

As mentioned, the least squares technique is a powerful mathematical tool for the estimation of unknown parameters. The only assumption that this technique requires to correctly interpret the results from statistical and physical points of view is that the observation errors should be random and preferably normally distributed. If these assumptions are violated, that is, if the observations are affected by gross errors, the properties of least squares are no longer valid. In such cases, the gross error of observations should be detected and eliminated by using robust techniques, and then, the least squares adjustment should be applied. An extensive discussion of robust parameters estimation can be found in Koch (1999). One of these robust techniques involves the minimization of the L_1 norm of the weighted residuals, that is

$$\mathbf{p}^T|\mathbf{v}|=\sum_{i=1}^{n} p_i|v_i| \rightarrow \min \tag{2}$$

where $\mathbf{p}=n\times 1$ vector, which contains the diagonal elements of the matrix $\mathbf{P}$. It should be noted that the L_1 norm adjustment is an unbiased estimate like least squares, but other advantages of the least squares such as minimum variance and maximum likelihood are no longer valid. The advantage of L_1 norm minimization compared to the least squares is its robustness, which means that it is less sensitive to outliers. In the following section, the formulation of the L_1 norm minimization in a rank deficient Gauss-Markov model will be presented.

Formulation of L_1 Norm Minimization

In the classical Gauss-Markov model, the unknown parameters $\mathbf{x}$ for a linear (linearized) parametric adjustment are determined based on the following functional and stochastic models:

$$\begin{gathered}\mathbf{l}+\mathbf{v}=\mathbf{A}\mathbf{x}\\ \mathbf{D}^T\mathbf{x}=\mathbf{0}\\ \mathbf{P}=\mathbf{Q}_l^{-1}=\sigma_0^2\mathbf{C}_l^{-1}\end{gathered} \tag{3}$$

where $\mathbf{v}_{n\times 1}$ as before is the vector of residuals; $\mathbf{l}_{n\times 1}$ = vector of observations; $\mathbf{A}_{n\times u}$ = rank deficient design matrix; $\mathbf{P}_{n\times n}$ = weight matrix of observations; $\mathbf{D}_{u\times d}$ = datum matrix of the network added to complete the rank deficiency of the design matrix; $\mathbf{0}_{d\times 1}$ = zero vector; $\mathbf{C}_{l(n\times n)}$ = covariance matrix of observations; $\mathbf{Q}_{l(n\times n)}$ = cofactor matrix; and σ_0^2 = a priori variance factor. As mentioned, the definition of L_1 norm minimization is the estimation of parameters, which minimize the weighted sum of the absolute residuals. The work presented here originates from the Marshall and Bethel (1996) paper in which they developed the basic concepts of L_1 norm adjustment and its implementation through linear programming and the simplex method. The main

[1]Dept. of Surveying Engineering, Faculty of Engineering, The Univ. of Isfahan, 81744 Isfahan, Iran. E-mail: amiri@eng.ui.ac.ir; also, Dept. of Mathematical Geodesy and Positioning, Faculty of Civil Engineering and Geosciences, TU Delft, Thijsseweg 11, 2629 JA, Delft, The Netherlands. E-mail: a.r.amiri-simkooei@geo.tudelft.nl

Note. Discussion open until July 1, 2003. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on November 1, 2001; approved on May 28, 2002. This paper is part of the ***Journal of Surveying Engineering***, Vol. 129, No. 1, February 1, 2003. ©ASCE, ISSN 0733-9453/2003/1-37–43/$18.00.

differences of Marshall and Bethel (1996) compared to the present work are mainly based on the following two aspects: In Marshall and Bethel (1996), the formulation of the L_1 norm minimization are presented only for two special examples, and there is no full formulation for a general Gauss-Markov model. The second advantage of the present paper is the consideration of the datum problem in the suggested formulation.

For modification of the objective function [Eq. (2)] and the constraints [Eq. (3)], the usual derivation of the least squares (L_2) estimates will not work for L_1 estimates, that is, an analog to the well-known normal equations is not possible for L_1 estimation. To transform Eqs. (2) and (3) into something we can work with, the usual strategy is to borrow a trick from linear programming and introduce slack variables that guarantee nonnegativity, and this allow us to write the objective function without absolute value signs (Marshall and Bethel 1996).

As mentioned, setting up the L_1 estimation problem by a linear programming solution requires us to formulate a mathematical model where all variables, both parameters and residuals, are nonnegative. The development begins with the familiar parametric Eq. (3) and is then transformed into an L_1 estimation problem by adding slack variables. To convert these equations into a form where there are nonnegative parameters and nonnegative residuals, we introduce two slack vectors, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, for the parameters, and two slack vectors, $\mathbf{u}$ and $\mathbf{w}$, for residuals. The parameters as well as the residuals may be positive or negative, so we replace these unknowns and residuals vectors with Marshall and Bethel (1996)

$$\mathbf{v}=\mathbf{u}-\mathbf{w}, \quad \mathbf{u},\mathbf{w}\geqslant\mathbf{0}$$

$$\mathbf{x}=\boldsymbol{\alpha}-\boldsymbol{\beta}, \quad \boldsymbol{\alpha},\boldsymbol{\beta}\geqslant\mathbf{0} \tag{4}$$

Rewriting the original parametric equations and datum constraints [Eq. (3)] and the objective function [Eq. (2)] in terms of slack variables yields

$$z=\mathbf{p}^T|\mathbf{v}|=\mathbf{p}^T|\mathbf{u}-\mathbf{w}|=\mathbf{p}^T(\mathbf{u}+\mathbf{w})\rightarrow\min$$

where

$$u_i=0 \quad \text{or} \quad w_i=0 \tag{5}$$

subject to

$$\mathbf{l}+\mathbf{u}-\mathbf{w}=\mathbf{A}(\boldsymbol{\alpha}-\boldsymbol{\beta})$$

$$\mathbf{D}^T(\boldsymbol{\alpha}-\boldsymbol{\beta})=\mathbf{0} \tag{6}$$

and

$$\mathbf{u},\mathbf{w},\boldsymbol{\alpha},\boldsymbol{\beta}\geqslant\mathbf{0}$$

or equivalently

$$\mathbf{z}=[\mathbf{0}^T \quad \mathbf{0}^T \quad \mathbf{p}^T \quad \mathbf{p}^T]\begin{bmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\beta}\\ \mathbf{w}\\ \mathbf{u}\end{bmatrix}\rightarrow\min \tag{7}$$

subject to

$$\begin{bmatrix}\mathbf{A} & -\mathbf{A} & \mathbf{I} & -\mathbf{I}\\ \mathbf{D}^T & -\mathbf{D}^T & \mathbf{Z} & \mathbf{Z}\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\beta}\\ \mathbf{w}\\ \mathbf{u}\end{bmatrix}=\begin{bmatrix}\mathbf{l}\\ \mathbf{0}\end{bmatrix}, \quad \boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{w},\mathbf{u}\geqslant\mathbf{0} \tag{8}$$

where $\mathbf{Z}$ is a zero matrix. Denoting

$$\begin{bmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\beta}\\ \mathbf{w}\\ \mathbf{u}\end{bmatrix}=\underline{\mathbf{x}}; \quad \begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ \mathbf{p}\\ \mathbf{p}\end{bmatrix}=\underline{\mathbf{c}} \tag{9}$$

and

$$\begin{bmatrix}\mathbf{A} & -\mathbf{A} & \mathbf{I} & -\mathbf{I}\\ \mathbf{D}^T & -\mathbf{D}^T & \mathbf{Z} & \mathbf{Z}\end{bmatrix}=\underline{\mathbf{A}}; \quad \begin{bmatrix}\mathbf{l}\\ \mathbf{0}\end{bmatrix}=\underline{\mathbf{b}} \tag{10}$$

One gets

$$z=\underline{\mathbf{c}}^T\underline{\mathbf{x}}\rightarrow\min \tag{11}$$

subject to

$$\underline{\mathbf{A}}\underline{\mathbf{x}}=\underline{\mathbf{b}}; \quad \underline{\mathbf{x}}\geqslant\mathbf{0} \tag{12}$$

This is a special operations research problem that can be solved by linear programming. The simplex method is a technique designed to solve a linear programming problem systematically, moving from corner to corner in the allowed solution space, i.e., Eq. (12), which is usually underdetermined, until the objective function reaches a minimum. For more information refer to the Appendix.

Solving for vector $\underline{\mathbf{x}}$ will yield the vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\mathbf{w}$, and $\mathbf{u}$, consequently, the solution vector $\mathbf{x}$ and the residual vector $\mathbf{v}$ can be obtained. These operations should be iterated for non-linear (linearized) models until the solution vector $\mathbf{x}$ converges to zeros.

Numerical Results on Simulated and Real Networks

For verification of the above formulation, three examples are presented. The first example is a simulated linear model; the second one is a simulated nonlinear model, and the third one is a real nonlinear network. The results presented above were also tested on other simulated and real networks by this author. All the results obtained have been calculated using a computer program in MATLAB software in which the subroutine ''linprog.m'' has been used.

Example 1

In the first example, a leveling network is assumed. The network consists of four points (P1, P2, P3, and P4) with six height difference observations. The datum of the network is provided by minimum constraints; point P1 has been considered as a fixed station, with a height of $H_{P1}=100$ m. The degree of freedom of the network is $df=3$. Two kinds of observations have been considered for this network. Table 1 shows a list of these observations and their observed values (columns 4 and 7). In the case of the second observations (column 7), it is assumed that the observed height difference P3 to P1 is erroneous by 5 cm. As seen from the table, the accuracy of these observations is 10 mm.

Table 1 also shows the results of adjustment by L_1 and L_2 norm minimization of the residuals. In this simple example, $\underline{A}$ in Eq. (12) is a 7×20 matrix. Because the model is linear, both techniques have been done without iteration. L_1 norm results have been obtained according to the suggested formulation. The adjusted residuals of the first group of observations by L_1 and L_2 norm are given in columns 5 and 6 of Table 1. The maximum absolute values of the residuals for L_1 and L_2 norms are 21.1 and 19.0 mm, respectively. It may imply that the L_1 norm minimization is more sensitive to outliers than L_2. The adjusted residuals of the second observations by L_1 and L_2 norms are given in

Table 1. Two Types of Observations of Leveling Network (Δh_i and Δh_i) with Their L_1 and L_2 Norm Adjusted Residuals

Observations				Residuals			Residuals	
From	To	σ (mm)	$\Delta h_i(m)$ (**1**)	L_1 (mm)	L_2 (mm)	$\Delta h_i(m)$ (**2**)	L_1 (mm)	L_2 (mm)
P1	P2	10	1.0066	0.00	−0.50	1.0031	−7.20	−19.6
P2	P3	10	0.9845	19.0	18.5	1.0142	−18.2	−22.4
P3	P4	10	0.9697	14.8	17.4	1.0033	0.00	6.9
P4	P1	10	−2.9946	0.00	−1.60	−2.9952	0.00	9.7
P3	P1	10	−2.0101	0.00	1.10	−1.9960+**0.05**	**−46.0**	**−29.3**
P2	P4	10	2.0091	**−21.1**	**−19.0**	1.9993	0.00	2.8
Absolute sum:		—	—	**54.9**	**58.0**	—	**71.3**	**90.7**

columns 8 and 9 of Table 1. The maximum absolute values of the residuals for L_1 and L_2 norms are 46.0 and 29.3 mm, respectively. It is obvious that the 5-cm gross error has been detected clearly by the L_1 norm rather than L_2. This means, again, that the L_1 norm minimization is a powerful technique to detect outlier observations.

Example 2

In the second example, a simulated trilateration network is assumed. The network consists of 6 points with 30 distance observations. Fig. 1 illustrates the position of the stations as well as the distance observations. Table 2 shows the list of observations and their observed values. The degree of freedom of the network is $df=21$. The accuracy of the observations is 10 mm, and the datum of the network is provided by inner constraints. Two kinds of observations have been considered for this network. These observations are the same except that the first observation is erroneous by 10 cm.

Each adjustment has been calculated using two techniques, i.e., the L_1 norm minimization that was formulated in this paper and the famous L_2 norm minimization. In this example, $\underset{\sim}{\mathbf{A}}$ in Eq. (12) is a 33×84 matrix. Because the Taylor series expansion was used, the final solution is obtained through iterations. The adjusted coordinates of the stations are given in Tables 3 and 4 after convergence of the corrections to zeros (after 3 iterations for both techniques). The initial coordinates for both techniques are the same and are given in Tables 3 and 4. Table 2 also shows the results of adjustment by L_1 and L_2 norm minimization of the residuals. The adjusted residuals of the first observations by L_1 and L_2 norms are given in columns 5 and 6. In the results of the L_1 norm, nine residuals (equivalent to the minimum observations for solving the problem) are zeros. The maximum absolute values of the residuals for L_1 and L_2 norms are 86.82 and 68.65 mm, respectively, which are related to the sixth observation (distance 2-1). As seen, both techniques resulted in a gross error in this observation.

The residuals of the second adjustment by L_1 and L_2 norms are given in columns 7 and 8. The maximum absolute values of the residuals for L_1 and L_2 norms are 108.50 and 63.52 mm, respectively, which are related to the first observation. As is obvious, a 10-cm gross error has been detected clearly by the L_1 norm rather than L_2. Another point that is clear from these results is that the residual of observation 6 in the L_1 norm adjustment has not been changed from its initial value (−86.82) in the first adjustment. This means that the error can be detected in the network just as it had been detected previously. But such a situation does not exist for the L_2 norm residual of the sixth observation. It has been reduced from −68.65 to −41.84 mm. This means that this residual may not detect the error of its observation. The above results may be summarized as follows: Implementation of the L_1 norm estimation as a powerful procedure for outliers detection proves to be very effective. The efficiency of this technique is more straightforward when more than one observation have errors.

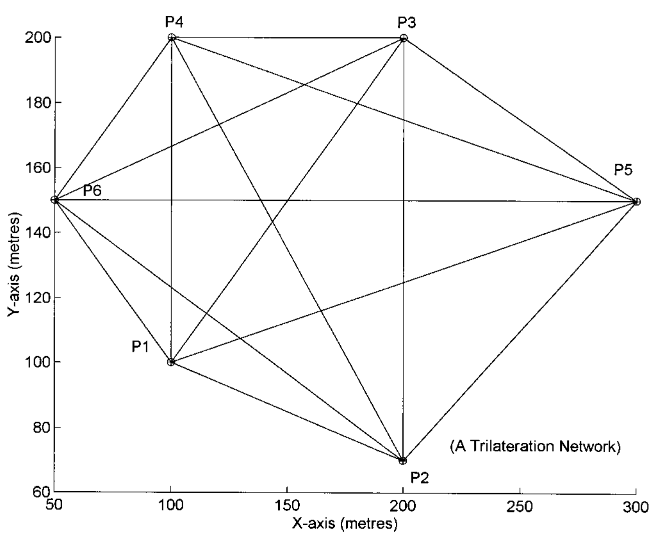

Fig. 1. Configuration of the trilateration network in Example 2

Example 3

In the third example, a real 2D-trilateration network was obtained from the Department of Surveying Engineering (1993) for an area (Baghbahadoran) in Isfahan Province in Iran. This network was established for educational purposes. The network consists of 8 points with 28 distance observations with nominal standard deviation of 3 mm+2 ppm using EDM DI1600 Wild. Fig. 2 illustrates the position of the stations as well as the distance observations. Table 5 gives a list of observations and their observed values. Degree of freedom of the network is $df=15$. The datum of the network is provided by inner constraints.

Each adjustment has been calculated, again, using two techniques, i.e., L_1 norm minimization and the famous L_2 norm minimization. In this network, $\underset{\sim}{\mathbf{A}}$ in Eq. (12) is a 31×88 matrix. Because Taylor series expansion was used, the final solution is obtained through iterations. The adjusted coordinates of the stations are given in Table 6 after convergence of the corrections to zeros (after 3 iterations for both techniques). The initial coordinates for both techniques are the same and are given in Table 6. Table 5 also shows the results of adjustment by L_1 and L_2 norm minimization of the residuals. The adjusted residuals of the first

Table 2. Two Types of Observations of Trilateration Network with Their L_1 and L_2 Norm Adjusted Residuals

	Distances			Residuals		Residuals	
Observation Number	From	To	Observed values (m)	L_1 (mm)	L_2 (mm)	L_1 (mm)	L_2 (mm)
1	1	2	104.4226	−8.50	9.66	**−108.50**	**−63.52**
2	1	3	141.4264	0.46	4.37	0.46	8.35
3	1	4	100.0186	−21.01	−23.39	−21.01	−28.12
4	1	5	206.1519	21.76	20.53	21.76	32.61
5	1	6	70.6993	0.00	−6.42	0.00	−17.28
6	2	1	104.5010	**−86.82**	**−68.65**	**−86.82**	**−41.84**
7	2	3	130.0119	−22.90	−13.95	−22.90	−14.91
8	2	4	164.0010	7.95	19.24	7.95	27.02
1	2	5	128.0688	0.00	−8.68	0.00	−16.23
10	2	6	169.9940	0.00	11.70	0.00	24.82
11	3	1	141.4269	0.00	3.90	0.00	7.88
12	3	2	129.9890	0.00	8.95	0.00	7.98
13	3	4	100.0009	4.19	3.41	4.19	0.92
14	3	5	111.7834	14.32	9.14	14.32	6.80
15	3	6	158.1090	−2.11	−1.96	−2.11	−4.52
16	4	1	100.0046	−6.99	−9.37	−6.99	−14.09
17	4	2	164.0090	0.00	11.29	0.00	19.07
18	4	3	100.0124	−7.31	−8.08	−7.31	−10.58
19	4	5	206.1490	10.10	2.83	10.10	−0.21
20	4	6	70.6874	13.07	13.35	13.07	12.36
21	5	1	206.1430	30.68	29.45	30.68	41.53
22	5	2	128.0730	−4.20	−12.88	−4.20	−20.43
23	5	3	111.8023	−4.58	−9.76	−4.58	−12.10
24	5	4	206.1591	0.00	−7.27	0.00	−10.32
25	5	6	250.0094	−11.80	−19.03	−11.80	−19.82
26	6	1	70.6895	9.80	3.38	9.80	−7.48
27	6	2	169.9936	0.43	12.14	0.43	25.25
28	6	3	158.1068	0.00	0.14	0.00	−2.40
29	6	4	70.7005	0.00	0.28	0.00	−0.70
30	6	5	249.9982	−0.53	−7.76	−0.53	−8.54

observations by L_1 and L_2 norms are given in columns 6 and 7 of Table 5. In the results of the L_1 norm, 13 residuals (equivalent to the minimum observations for solving the problem) are zeros. The maximum absolute values of the residuals for the L_1 norm are 17.16, 16.01, and 10.87 mm. The maximum absolute values of the residuals for the L_2 norm are 8.85, 9.95, and 6.57 mm. They all, both L_1 and L_2 residuals, are related to the same observations, i.e., distances 2–4, 3–5, and 4–6. This means that both techniques resulted in gross errors in these observations. One point which is, again, obvious from these results is that the L_1 norm residuals of the errors are more than the L_2 norm residuals. This means that the L_2 norm criterion tends to distribute the errors into good quality observations. This can be implied by looking at the residual of the 23rd observation, which equals 6.38 mm.

Conclusions

L_1 norm adjustment is a technique used in geodetic networks to detect errors in observations. In this paper, the formulation and implementation of L_1 norm minimization were presented for rank deficient Gauss-Markov models, which leads to the solving a linear programming problem. For verification of the suggested formulation, three examples were presented on linear and nonlinear

Table 3. L_1 and L_2 Norm Adjusted Coordinates for First Observations in Example 2

	Initial Coordinates		Adjusted Coordinates (L_1 norm)		Adjusted Coordinates (L_2 norm)	
Point Number	X	Y	X	Y	X	Y
1	100	100	99.9875	99.9993	99.9816	100.0024
2	200	70	200.0000	70.0025	200.0101	69.9958
3	200	200	200.0031	199.9915	200.0035	199.9938
4	100	200	99.9981	199.9970	99.9992	199.9976
5	300	150	300.0045	150.0071	299.9980	150.0072
6	50	150	50.0068	150.0026	50.0076	150.0032

Table 4. L_1 and L_2 Norm Adjusted Coordinates for Second Observations in Example 2

Point Number	Initial Coordinates		Adjusted Coordinates (L_1 norm)		Adjusted Coordinates (L_2 norm)	
	X	Y	X	Y	X	Y
1	100	100	99.9875	99.9993	99.9682	100.0059
2	200	70	200.0000	70.0025	200.0229	69.9933
3	200	200	200.0031	199.9915	200.0028	199.9902
4	100	200	99.9981	199.9970	100.0010	199.9964
5	300	150	300.0045	150.0071	299.9974	150.0092
6	50	150	50.0068	150.0026	50.0078	150.0050

(both real and simulated) models. The results showed that L_1 norm minimization is more sensitive than least squares for outlier detection. L_1 norm minimization is more efficient than the L_2 norm when there are more errors in observations.

It seems that the lack of attention paid to L_1 norm minimization in geodetic applications has been mainly due to the relative complexity of its implementation compared to least squares, but this complexity is not important in the presence of modern computing techniques.

Appendix: Linear Programming Problem

Linear programming (LP) is a branch of operations research. This method deals with the problem of minimizing or maximizing (optimizing) a linear function in the presence of linear equality and/or inequality constraints. This appendix describes the problem of LP. The main concepts of this optimization problem are presented. To solve the problem, some effective solution algorithms will be introduced. For details of these solution methods, some references are suggested. The discussion will be started by formulating a particular type (standard format) of linear programming. Any general linear programming may be manipulated into this form.

Standard Format of LP

The subject of linear programming, sometimes called linear optimization, in standard form is concerned with the following problem: For n independent variables $x_1, x_2, \ldots, x_n$, minimize the function

$$z = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n \quad (13)$$

subject to the primary constraints

$$x_1 \geqslant 0, x_2 \geqslant 0, \ldots, x_n \geqslant 0 \quad (14)$$

and simultaneously subject to m additional constraints of the form

$$a_{i1} x_1 + a_{i2} x_2 + \ldots + a_{in} x_n = b_i; \quad i = 1,2,\ldots,m \quad (15)$$

Here, z = the objective function to be minimized. The coefficients $c_1, c_2, \ldots, c_n$ are the (known) cost coefficients, and $x_1, x_2, \ldots, x_n$ are the decision variables to be determined. The coefficients a_{ij} for $i = 1,2, \ldots, m$ and $j = 1,2, \ldots, n$ are called the technological coefficients. The constraints in Eq. (14) are the nonnegativity constraints. A set of variables $x_1, x_2, \ldots, x_n$ satisfying all the constraints are called a feasible point or a feasible vector. The set of all such points constitutes the feasible region or the feasible

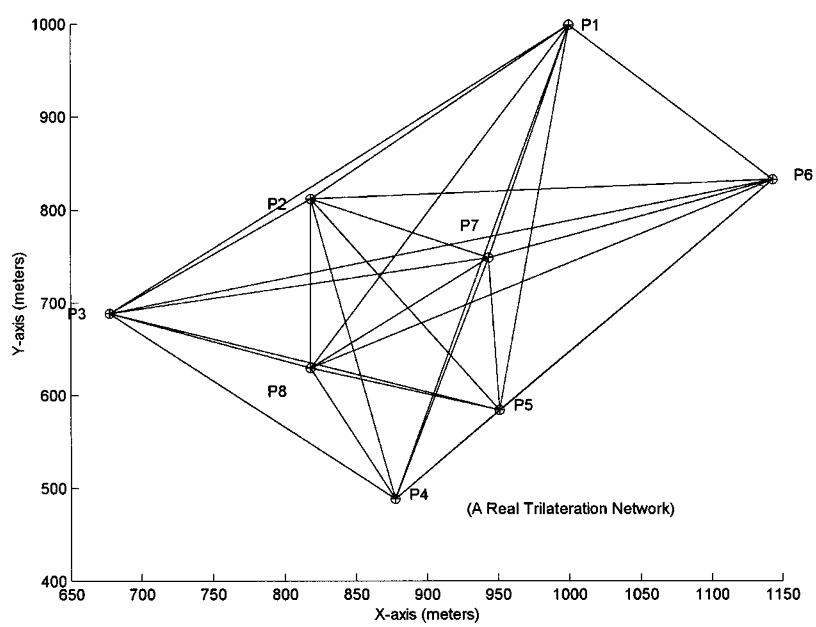

Fig. 2. Configuration of the real trilateration network in Example 3

Table 5. List of Observations of Real Geodetic Network with Their L_1 and L_2 Norm Adjusted Residuals

	Distances				Residuals	
Observation Number	From	To	Observed value	Accuracy (mm)	L_1 norm (mm)	L_2 norm (mm)
1	1	2	261.298	3.05	4.66	0.99
2	1	3	448.088	3.13	0	−2.04
3	1	4	526.093	3.18	0	3.6
4	1	5	418.081	3.11	−3.94	−3.72
5	1	6	219.709	3.03	0	−0.87
6	1	7	257.271	3.04	3.69	3.88
7	1	8	412.409	3.11	−0.87	−2.54
8	2	3	187.167	3.02	2.84	4.4
9	2	4	329.064	3.07	**−16.01**	**−9.95**
10	2	5	263.253	3.05	0	5.02
11	2	6	325.765	3.07	0	1.23
12	2	7	140.336	3.01	−2.03	0.76
13	2	8	182.207	3.02	0	2.75
14	3	4	283.135	3.05	0	1.96
15	3	5	292.817	3.06	**−10.87**	**−6.57**
16	3	6	487.913	3.15	−5.64	−3.22
17	3	7	272.913	3.05	0	1.12
18	3	8	152.423	3.02	2.99	4.81
19	4	5	121.382	3.01	0	5.73
20	4	6	435.856	3.12	**−17.16**	**−8.85**
21	4	7	268.871	3.05	1.68	5.17
22	4	8	153.454	3.02	0	2.16
23	5	6	314.454	3.07	3.82	6.38
24	5	7	164.339	3.02	0	0.03
25	5	8	140.562	3.01	1.08	3.55
26	6	7	216.929	3.03	0	1.96
27	6	8	383.912	3.10	0	2.06
28	7	8	173.118	3.02	0.26	−0.52

space. Using this terminology, the linear programming problem can be stated as follows: Among all feasible vectors, find one that minimizes the objective function.

A linear programming problem can be stated in a more convenient form using a matrix notation. Consider the following column vectors **x**, **c**, and **b**, and the $m \times n$ matrix **A**

$$\mathbf{x}=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{c}=\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}, \quad \mathbf{b}=\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

$$\mathbf{A}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \tag{16}$$

Then, the standard form of the LP problem can be written as follows:

$$\text{Minimize} \quad \mathbf{c}^T\mathbf{x}, \tag{17}$$

$$\text{subject to} \quad \mathbf{A}\mathbf{x}=\mathbf{b}, \tag{18}$$

Table 6. L_1 and L_2 Norm Adjusted Coordinates of Real Geodetic Network in Example 3

	Initial Coordinates		Adjusted Coordinates (L_1 Norm)		Adjusted Coordinates (L_2 Norm)	
Point Number	X (m)	Y (m)	X (m)	Y (m)	X (m)	Y (m)
1	1,000	1,000	1,000.000	1,000.001	999.999	1,000.001
2	818.516	812.010	818.511	812.010	818.511	812.013
3	677.672	688.732	677.674	688.732	677.673	688.734
4	877.679	488.325	877.680	488.326	877.677	488.322
5	951.401	584.753	951.398	584.759	951.401	584.758
6	1,143.558	833.680	1,143.555	833.676	1,143.556	833.679
7	943.863	748.931	943.866	748.925	943.867	748.924
8	818.243	629.803	818.248	629.804	818.248	629.804

$$\mathbf{x} \geqslant \mathbf{0}$$

Another form of the LP problem is called a canonical form. A minimization problem is in canonical form if all variables are nonnegative, and all the constraints in Eqs. (15) and (18) are of the $\geqslant$ type. By simple manipulations, this problem can be transformed to the standard form. For example, an inequality can be easily transformed into an equation. To illustrate, consider the constraint given by $\Sigma a_{ij} x_j \geqslant b_i$. This constraint can be put in an equation form by subtracting the nonnegative slack variable x_{n+i}, leading to $\Sigma a_{ij} x_j - x_{n+i} = b_i$ and $x_{n+i} \geqslant 0$.

Fundamental Theorems of Linear Programming

Consider the system $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{x} \geqslant \mathbf{0}$, where $\mathbf{A}$ is an $m \times n$ matrix, and $\mathbf{b}$ is an $m \times 1$ vector. Suppose that rank $(\mathbf{A},\mathbf{b}) = \text{rank}(\mathbf{A}) = m$. After possibly rearranging the columns of $\mathbf{A}$, let $\mathbf{A} = [\mathbf{B},\mathbf{N}]$ where $\mathbf{B}$ is an $m \times m$ invertible matrix, and $\mathbf{N}$ is an $m \times (n-m)$ matrix. The solution is $\mathbf{x} = \binom{\mathbf{x_B}}{\mathbf{x_N}}$ to equations $\mathbf{Ax} = \mathbf{b}$, where

$$\mathbf{x}_B = \mathbf{B}^{-1}\mathbf{b}; \quad \text{and} \quad \mathbf{x}_N = \mathbf{0} \tag{19}$$

is called a basic solution of the system. If $\mathbf{x}_B \geqslant \mathbf{0}$, then $\mathbf{x}$ is called a basic feasible solution of the system. Here $\mathbf{B}$ is called the basic matrix (or simply basis), and $\mathbf{N}$ is called the nonbasic matrix. The components of $\mathbf{x}_B$ are called basic variables and the components of $\mathbf{x}_N$ are called nonbasic variables. If $\mathbf{x}_B > \mathbf{0}$, then $\mathbf{x}$ is called a nondegenerate basic feasible solution, and if at least one component of $\mathbf{x}_B$ is zero, then $\mathbf{x}$ is called a degenerate basic feasible solution.

Considering the above definitions, some theorems on linear programming can be summarized as follows:

- Theorem 1. Every basic feasible solution is equivalent to the extreme point of the nonempty feasible region and vice versa.
- Theorem 2. If an optimal solution exists, then an optimal extreme point (or equivalently, an optimal basic feasible solution) exists that is the optimal solution.
- Theorem 3. For every extreme point (basic feasible solution) there corresponds a basis (not necessarily unique), and, conversely, for every basis there corresponds a unique extreme point. Moreover, if an extreme point has more than one basis representing it, then it is degenerate. Conversely, if a degenerate extreme point has more than one basis representing it, then it is degenerate. Conversely, a degenerate extreme point has more than one basis representing it if and only if the system $\mathbf{Ax} = \mathbf{b}$ itself does not imply that the degenerate basic variables corresponding to an associated basis are identically zero.

Simplex Method

Because the extreme points may be enumerated by algebraically enumerating all basic feasible solutions that are bounded by

$$\binom{n}{m}$$

one may think of simply listing all basic feasible solutions and picking the one with the minimal objective value. This is unsatisfactory, however, for a number of reasons. First, the number of basic feasible solutions is bounded by

$$\binom{n}{m}$$

which is large, even for moderate values of m and n. Second, this simple approach does not tell us if the problem has an unbounded solution that may occur if the feasible region is unbounded. Thirdly, if the feasible region is empty, we shall discover that only after all possible ways.

There are different ways to approach the solution of a linear programming problem; the simplex method by Dantzig is recommended first. The simplex method is a clever procedure that moves from an extreme point to another extreme point, with a better (at least not worse) objective. It also discovers whether the feasible region is empty and whether the optimal solution is unbounded. In practice, the method only enumerates a small portion of the extreme points of the feasible region. The interested reader can find more explanations in, Dantzig (1963) and Bazaraa et al. (1990).

Acknowledgments

The writer would like to thank R. C. Burtch, the editor, and the two anonymous reviewers for their valued and constructive comments on this paper.

References

Bazaraa, M. S., Jarvis, J. J., and Sherali, H. D. (1990). *Linear programming and network flows*, 2nd Ed., Wiley, New York.

Dantzig, G. B. (1963). *Linear programming and extensions*, Princeton University Press, Princeton, N.J.

Department of Surveying Engineering (1993). ‘‘Micro-geodesy network for educational purposes in Baghbahadoran.’’ The Univ. of Isfahan, 81744, Isfahan, Iran.

Koch, K. R. (1999). *Parameter estimation and hypothesis testing in linear models*, Springer, Berlin.

Marshall, J., and Bethel, J., (1996). ‘‘Basic concepts of L1 norm minimization for surveying applications.’’ *J. Surv. Eng.*, 122(4), 168–179.

Mikhail, E. M. (1976). *Observations and least squares*, IEP-A dun-Donnelley, Chicago.