Research Article

Stefano S. Suraci*, Leonardo C. Oliveira, Ivandro Klein, and Ronaldo R. Goldschmidt

Introducing covariances of observations in the minimum L1-norm, is it needed?

https://doi.org/10.1515/jogs-2022-0135
received May 26, 2022; accepted September 8, 2022

Abstract: The most common approaches for assigning weights to observations in minimum L1-norm (ML1) is to introduce weights of p or $\sqrt{p}$, p being the weights vector of observations given by the inverse of variances. Hence, they do not take covariances into consideration, being appropriated only to independent observations. To work around this limitation, methods for decorrelation/unit-weight reduction of observations originally developed in the context of least squares (LS) have been applied for ML1, although this adaptation still requires further investigations. In this article, we presented a deeper investigation into the mentioned adaptation and proposed the new ML1 expressions that introduce weights for both independent and correlated observations; and compared their results with the usual approaches that ignore covariances. Experiments were performed in a leveling network geometry by means of Monte Carlo simulations considering three different scenarios: independent observations, observations with "weak" correlations, and observations with "strong" correlations. The main conclusions are: (1) in ML1 adjustment of independent observations, adaptation of LS techniques introduces weights proportional to $\sqrt{p}$ (but not p); (2) proposed formulations allowed covariances to influence parameters estimation, which is unfeasible with usual ML1 formulations; (3) introducing weighs of p provided the closest ML1 parameters estimation compared to that of LS in networks free of outliers; (4) weighs of $\sqrt{p}$ provided the highest successful rate in outlier identification with ML1. Conclusions (3) and (4) imply that introducing covariances in ML1 may adversely affect its performance in these two practical applications.

*** Corresponding author: Stefano S. Suraci,** Defense Engineering Program, Military Institute of Engineering, Rio de Janeiro, Brazil, e-mail: stefano@ime.eb.br
Leonardo C. Oliveira, Ronaldo R. Goldschmidt: Defense Engineering Program, Military Institute of Engineering, Rio de Janeiro, Brazil
Ivandro Klein: Graduate Program in Geodetic Sciences, Federal University of Paraná, Curitiba, Brazil

Keywords: adjustment computations, Cholesky factorization, decorrelation of observations, leveling network, minimum L1-norm, Monte Carlo simulation, weights of observations

1 Introduction

To relate the parameters to the observations, the linear (or linearized) model of Gauss–Markov is most commonly used in geodetic networks (Klein et al. 2019). With m being the number of observations, n the number of parameters (unknown coordinates of network points), $A_{m\times n}$ the design matrix (Jacobian), $L_{m\times 1}$ the vector of the observed values, $v_{m\times 1}$ the vector of the residuals of the observations (previously unknown), Σ_{Lmxm} the covariance matrix of the observations, σ_0^2 the priori variance factor, and $P_{m\times m}$ the (symmetric positive-definite) matrix of the weights of observations, the parameter vector $x_{n\times 1}$ is determined based on the following mathematical model (Gauss 1809):

$$Ax = L + v;\ P = \sigma_0^2 \sum_L^{-1}. \tag{1}$$

Thus, the smaller the variance (the better the precision) of an observation, its weight tends to be higher in the adjustment of observations. The least squares (LS) method is the most well-established for adjustment computations in geodetic networks. Its objective function imposes the minimization of the sum of the squares of the residuals weighted by the weight matrix of observations P, that is

$$\text{LS: } v^T P v = \min. \tag{2}$$

The first formal reports on LS go back to Legendre (1805) and Gauss (1809). LS is the best linear unbiased estimator (BLUE) for the parameters (Teunissen 2018) and the maximum likelihood estimator in case of normally distributed observational errors. Its solution for x is given by

$$x = (A^T \mathrm{PA})^{-1} A^T P L. \tag{3}$$

Also widely explored in the geodetic literature, the estimator that minimizes the L1-norm of residuals (equation (4)), herein called minimum L1-norm (ML1), aims at the minimization of the sum of absolute residuals weighted by p (Amiri-Simkooei 2003), p being the vector of weights p_i of independent observations, elements of the main diagonal of P (a diagonal matrix in this case, as observations are independent). ML1 is usually applied for outlier identification in geodetic networks (Suraci et al. 2021). After this intermediate step, final LS estimation can be performed in the set of geodetic observations free of outliers.

$$\text{ML1}(p): \ p^T|v| = \min. \tag{4}$$

Inal et al. (2018), Amiri-Simkooei (2018), and Klein et al. (2021) have also adopted the vector p to weigh independent observations in ML1. Some other authors, on the other hand, have considered the vector $\sqrt{p}$ in this task, $\sqrt{p}$ being the vector of the square root of elements of p – e.g., Junhuan (2005) and Marshall (2002). For this latter case, ML1 objective function of equation (4) is then replaced by

$$\text{ML1}(\sqrt{p}): \ \sqrt{p}^T|v| = \min. \tag{5}$$

Note, however, that such formulations do not take covariances into consideration in the adjustment of observations. Therefore, as mentioned, equations (4) and (5) are appropriated only for independent observations. Nevertheless, correlated observations are common in surveying engineering, which motivated the use of decorrelation methods of observations originally developed for LS in the context of ML1, as in Yetkin and Inal (2011) and Baselga et al. (2020).

1.1 Decorrelation methods originally developed for LS already applied in the context of ML1

The goal of the decorrelation of observations is to obtain an equivalent mathematical model (with the same solution for the parameters), but with all observations decorrelated (independent, without covariances). Moreover, as a result of the decorrelation process, usually performed based on the Cholesky factorization of the matrix P, all the weights of observations in the new model become equal to 1, which justifies the use of the term unit-weight reduction. Actually, in order to obtain a model with all observations of unit weight, the decorrelation process can be applied even to independent observations (Suraci et al. 2019).

Let $D \in R^{n\times n}$ be a positive definite matrix. By the Cholesky factorization, D can be expressed by $D = W^TW$, where $W \in R^{n\times n}$ is upper triangular with only positive elements on the main diagonal (Higham 2009). Methods for computing the respective (unique) matrix W are presented in Golub and Loan (1996). In geodetic applications, as presented in Strang and Borre (1997), the matrix of the weights of observations P can then be decomposed by the Cholesky method into

$$P = W^TW. \tag{6}$$

Continuing the approach of Strang and Borre (1997), multiplying the mathematical model of equation (1) by W, and adopting the identity matrix $I_{m\times m}$ as the matrix of the weights P' of observations, we have

$$\begin{aligned} &A'x = L' + v', \ \text{with } A' = \text{WA}, \ L' = WL \\ &\text{and } v' = Wv; \ P' = I. \end{aligned} \tag{7}$$

Therefore, it is possible to demonstrate that the LS objective function (equation (2)) of this mathematical model of equation (7) (with unit weights) remains the same as in the model of equation (1) (with the original weights of P):

$$\begin{aligned} \min({v'}^TP'v') &= \min({v'}^TIv') = \min((Wv)^TI(Wv)) \\ &= \min(v^TW^TWv) = \min(v^TPv). \end{aligned} \tag{8}$$

As a consequence, as shown by Suraci et al. (2019), the LS solution for x in the model of equation (7) remains the same as in the model of equation (1) as well:

$$\begin{aligned} x &= ((\text{WA})^TIWA)^{-1}(\text{WA})^TIWL \\ &= (A^TW^TIWA)^{-1}A^TW^TIWL = (A^TPA)^{-1}A^TPL. \end{aligned} \tag{9}$$

Hence, this is a proper procedure to introduce the weights of P via decorrelation of observations in LS adjustment computations. From equation (8), it is further noted that W could be any matrix such that $W^TW = P$, but the matrix from Cholesky factorization is probably the most viable option, as it has an established procedure to be computed in the literature (Golub and Loan 1996).

Still in the LS context, Ingram (1911) demonstrates that unit-weight reduction can also be obtained by multiplying the equation of every original observation by respective $\sqrt{p_i}$. However, this technique has the disadvantage of necessarily starting from independent observations, i.e., it is not suitable for the case of correlated observations. This same technique is called homogenization in LS by Koch (1999). Although useful for applications that are outside the scope of this article, this procedure does not provide uncorrelated

observations if applied to correlated observations (Prószyński 2010).

Actually, the homogenization is the particular case of the unit-weight reduction process of Strang and Borre (1997) where all observations are independent. Let $\sqrt{P}$ be the diagonal matrix (hence also upper triangular) with $\sqrt{p_i}$ elements on the main diagonal. Thus, multiplying all observations by their respective $\sqrt{p_i}$, in the terms of Ingram (1911), but in matrix format, we have

$$\sqrt{P}Ax = \sqrt{P}L + \sqrt{P}v;\ P' = I. \tag{10}$$

Moreover, since $\sqrt{P}$ is a diagonal matrix

$$P = \sqrt{P}^T\sqrt{P}. \tag{11}$$

Hence, comparing equations (6) and (7) with equations (11) and (10), respectively, we can see that, for this particular case of independent observations, $\sqrt{P}$ is the (unique) upper triangular matrix with only positive elements on the main diagonal of the Cholesky factorization of P, i.e., $W = \sqrt{P}$.

ML1, however, has no direct analytical solution (Suraci et al. 2021), which makes it difficult to algebraically guarantee that ML1 adjustment of the model of equation (6) (with unit weights) has the same solution of the model of equation (1) (with the original weights). In this context, Suraci et al. (2019) showed by means of Monte Carlo (MC) simulations that the approach of Strang and Borre (1997), properly developed to introduce weights of P in LS adjustments, needs further investigation in ML1 as it presented results different from ML1(p) (equation (4)) even in scenarios with only independent observations.

Although demonstrated only for the LS method, the approach of Strang and Borre (1997) has already been applied for unit-weight reduction in the ML1 adjustment of geodetic networks by Yetkin and Inal (2011). The approach of Ingram (1911), a particular case of the approach of Strang and Borre (1997), was also applied in the context of ML1 by Baselga et al. (2020).

In this sense, this article provides new results that lead to a better understanding of why these approaches originally for unit-weight reduction in LS should not be unrestrictedly applied in the ML1 context, presents an attempt to formulate ML1 expressions that introduces weights for both independent and correlated observations, and analyzes the optimal ways of introducing weights in ML1 in order to have the best performance in outlier identification and to provide parameters estimation closer to LS in outlier-free networks.

2 Formulations for both independent and correlated observations in ML1

For independent observations, ML1(p) expression of equation (4) may be rewritten as the minimization of:

$$\begin{aligned} p^T|v| &= p_1|v_1| + p_2|v_2| + \ldots + p_m|v_m| \\ &= \mathrm{sum}\left(\begin{bmatrix} p_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & p_m \end{bmatrix}\begin{bmatrix} |v_1| \\ \vdots \\ |v_m| \end{bmatrix}\right) \\ &= \mathrm{sum}(P|v|), \end{aligned} \tag{12}$$

where sum(.) refers to the sum of elements of a vector. On the other hand, if we try to assume the mathematical model given by the approach of Strang and Borre (1997) (equation (7)), as done in equation (8), but now to ML1 adjustment (instead of LS), it will provide the following objective function:

$$\min(\mathrm{sum}(|v'|)) = \min(\mathrm{sum}(|Wv|)). \tag{13}$$

Since $W = \sqrt{P}$ for independent observations and $\sqrt{P}$ is a diagonal matrix with only positive elements on the main diagonal, as mentioned, equation (13) can be rewritten as

$$\begin{aligned} \min(\mathrm{sum}(|v'|)) &= \min(\mathrm{sum}(|Wv|)) \\ &= \min(\mathrm{sum}(|\sqrt{P}v|)) \\ &= \min(\mathrm{sum}(\sqrt{P}|v|)) \\ &= \min(\sqrt{p}^T|v|). \end{aligned} \tag{14}$$

As equation (14) is equal to equation (5), it shows that the approach of Strang and Borre (1997), the way it is applied in LS without changing its objective function (equation (8)), introduces weights proportional to $\sqrt{p}$ (but not to p) in ML1 adjustment of independent observations. Hence, it must not be applied to introduce weights of P in ML1. Actually, applying the approach of Strang and Borre (1997) in the context of ML1 with independent observations results in a model with weights of $\sqrt{p}$. At this point, it is essential to emphasize that we are not stating that the adaptation to ML1 of the approach of Strang and Borre (1997) is inappropriate, but that it forces weights proportional to $\sqrt{p}$ (but not to p), also something usual in the ML1 context, as mentioned.

However, note that, differently from equation (4), in the (new) formulation of equation (12), P does not need to be a diagonal matrix. As a consequence, besides being equivalent to ML1(p) in case of independent observations,

it also allows covariances to be considered in the ML1 adjustment computation. Therefore, it is an alternative in order to have a general expression for introducing weights of P in ML1 for both the correlated and independent observations. Hence, we now define ML1(P), equivalent to ML1(p) in case of independent observations

$$\text{ML1}(P):\ \text{sum}(P|v|) = \min. \tag{15}$$

Considering the general case of correlated observations in equation (15), we have the minimization of

$$\begin{aligned}
\text{sum}(P|v|) &= \text{sum}\left(\begin{bmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & \ddots & \vdots \\ p_{m1} & \cdots & p_{mm} \end{bmatrix}\begin{bmatrix} |v_1| \\ \vdots \\ |v_m| \end{bmatrix}\right) \\
&= (p_{11}|v_1| + p_{12}|v_2| + \ldots + p_{1m}|v_m|) \\
&\quad + (p_{21}|v_1| + p_{22}|v_2| + \ldots + p_{2m}|v_m|) \\
&\quad + \ldots + (p_{m1}|v_1| + p_{m2}|v_2| + \ldots + p_{mm}|v_m|) \\
&= \underbrace{(p_{11} + p_{21} + \ldots + p_{m1})}_{k_1}|v_1| \\
&\quad + \underbrace{(p_{12} + p_{22} + \ldots + p_{m2})}_{k_2}|v_2| \\
&\quad + \ldots + \underbrace{(p_{1m} + p_{2m} + \ldots + p_{mm})}_{k_m}|v_m| \\
&= k_1|v_1| + k_2|v_2| + \ldots + k_m|v_m| \\
&= k^T|v|,\ k_i > 0.
\end{aligned} \tag{16}$$

We included the restriction $k_i > 0$, which implies that $k^T|v| \geq 0$, as a norm function must always be positive or equal to zero. Besides, $k_i \leq 0$ could cause numerical instability to ML1 solution, because respective $|v_i|$ could tend to grow higher than expected as an attempt to minimize $k^T|v|$ (if $k_i < 0$) or it would not be considered in the objective function (if $k_i = 0$). However, note that p_{ii} is always positive with absolute value usually much higher than other terms p_{ji} that add up to k_i; some p_{ji} are also positive; and, many p_{ji} are zero even in networks with correlated observations – refer to many examples of these points in Ghilani (2010). Therefore, all k_i tend to be positive, and so such restriction is generally satisfied for most geodetic networks. The less likely case where the sum of the absolute values of the terms p_{ji} with minus sign is higher or equal to the sum of p_{ii} with the terms p_{ji} with positive sign may be a flaw of this formulation and was not addressed in this article.

Similar to equation (15), the general formulation proposed for both independent and correlated observations that is equivalent to ML1($\sqrt{p}$) (equation (5)) in case of independent observations is given by equation (17). A detailed expression for the general case of correlated observations of equation (17) is obtained by substituting every element p_{ij} of P with respective w_{ij} of W in equation (16).

$$\text{ML1}(W):\ \text{sum}(W|v|) = \min. \tag{17}$$

3 Numerical results

Experiments were conducted in the leveling network geometry of Figure 1. It consists of one control station with fixed coordinate $h_A = 10.0$ mm, $m = 6$ observations (height differences), and $n = 3$ unknowns (station heights). Correlated observations may occur in leveling networks – e.g., Knight, Wang and Rizos (2010) and Schaffrin (1997) – due to the use of the same equipment in different leveling lines (observations).

In each experiment, $M = 200{,}000$ MC trials were run with random errors of each observation e_i, $i = (1,2,\ldots,m = 6)$, generated according to a multivariate normal distribution $e \sim N(0, \Sigma_L)$. ML1 adjustments were computed by the *simplex* method of linear programming (Dantzig 1963). Programming codes were developed under *Octave* software, version 6.3.0. The reader can contact the first author to obtain the *Octave* codes of the experiments.

The MC method enables one to obtain a large number of independent realizations of the same experiment in a controlled environment, which is difficult to guarantee with real data (Yang et al. 2021). From the statistical distributions of the input variables, repeated samples of them are generated, allowing an accurate characterization of the results of the analyzed model. The larger the number of simulations M, the greater the accuracy in characterizing the results. This is a clear advantage over simple tests with only one geodetic network, which can lead to wrong conclusions because they are more subject to random results that are not very representative of the phenomenon. The quantity of $M = 200{,}000$ was suggested by Rofatto et al. (2020) in the context of geodetic networks.

For the investigation of the points discussed, the network was adjusted considering the following five approaches for introducing weights in ML1 formulation:

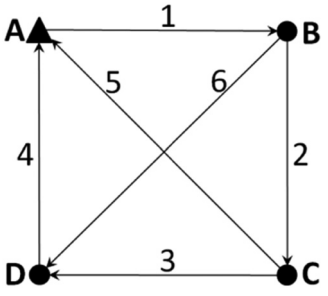

Figure 1: Leveling network geometry.

1) ML1(p) : $p^T|v|$ = min.
2) ML1(P) : sum($P|v|$) = min.
3) ML1($\sqrt{p}$) : $\sqrt{p}^T|v|$ = min.
4) ML1(W) : sum($W|v|$) = min.
5) ML1(decorrelation_LS) : sum($|v'|$) = min.

As presented, note that the approaches 1–4 correspond to the objective functions of equations (4), (5), (15), and (17), respectively. Approach 5 refers to the approach of Strang and Borre (1997) for unit-weight reduction of observations in LS (considering the mathematical model of equation (7)), but applied to ML1.

In Experiments 1–3, we also computed the same 200,000 MC scenarios of each experiment, but with adjustment of observations by LS (instead of ML1). Our goals were to use LS parameters estimation as a reference, and to measure the effect of correlations in LS, also to use as a reference for assessment of the effect in ML1. For this purpose, LS parameters were calculated by equation (3), using the LS default matrix of weights P (considering covariances); then, for comparison, we assumed the weights of the diagonal matrix P_d (with the same main diagonal of P, but with no covariances) in LS parameters estimation.

4 Experiment 1: independent observations

In the first experiment, observations were assumed to be independent (Σ_L and P were diagonal matrices). For the MC simulations, the variances of observations σ_{ii}^2, elements of the main diagonal of Σ_L, adopted were [7.5 10 12.5 15 17.5 20] mm^2. Table 1 shows the mean value of the parameters estimated (considering all 200,000 MC scenarios) in each adjustment computation.

5 Experiment 2: correlated observations ("weak" correlations)

In the second experiment, observations were simulated with "weak" correlations, i.e., with variances relatively much higher than the magnitude of the covariances. For the MC simulations, variances adopted were the same as in Experiment 1, but now covariances were not zero, being randomly selected from a uniform distribution of values between –1.0 and 1.0 mm^2. Hence, the maximum absolute covariance was more than 7.5 times smaller than the minimum variance (which we call "weak" correlations) in Experiment 2. For the respective ML1 adjustment computations, the elements of p and $\sqrt{p}$ were computed by the inverse of the variances and the square root of the variances, respectively, without taking covariances into account. Table 2 shows the mean value of the parameters estimated (considering all 200,000 MC scenarios) in each adjustment computation.

Table 1: Mean value of the parameters estimated (mm) (Experiment 1)

	h_B	h_C	h_D
ML1(p)	10.0038	10.0074	10.0139
ML1(P)	10.0038	10.0074	10.0139
ML1($\sqrt{p}$)	10.0040	10.0076	10.0141
ML1(W)	10.0040	10.0076	10.0141
ML1(decorrelation_LS)	10.0040	10.0076	10.0141
LS(P)	9.9993	9.9928	10.0058
LS(P_d)	9.9993	9.9928	10.0058

Note that even though we had the same variances in Experiment 1, since covariances were different, the values of the random errors of observations simulated were also different. Hence, adjustment results of ML1(p) in Experiments 1 and 2 were not equal for the three parameters, for example. Similarly, this statement can be extended to all adjustment approaches in all experiments.

6 Experiment 3: correlated observations ("strong" correlations)

In the third experiment, observations were simulated with "strong" correlations, i.e., covariances with values

Table 2: Mean value of the parameters estimated (mm) (Experiment 2)

	h_B	h_C	h_D
ML1(p)	10.0038	10.0057	10.0123
ML1(P)	10.0038	10.0057	10.0123
ML1($\sqrt{p}$)	10.0052	10.0072	10.0138
ML1(W)	10.0052	10.0072	10.0138
ML1(decorrelation_LS)	10.0049	10.0060	10.0131
LS(P)	9.9996	9.9925	10.0058
LS(P_d)	9.9995	9.9927	10.0057

relatively closer to those of the variances. For the MC simulations, variances adopted were the same as in Experiments 1 and 2, but now covariances were randomly selected from a uniform distribution of values between −5.0 and 5.0 mm^2. Hence, the maximum absolute covariance was approximately 2/3 of the minimum variance (which we call "strong" correlations) in Experiment 3. As in Experiment 2, for the respective adjustment computations, the elements of p and $\sqrt{p}$ were computed by the inverse of the variances and the square root of the variances, respectively, without taking covariances into account. Table 3 shows the mean value of the parameters estimated (considering all 200,000 MC scenarios) in each adjustment computation.

7 Experiment 4: outlier identification

Since outlier identification is the most widely explored application of ML1 in geodetic networks, the last experiment was designed to verify if correlations in ML1 affect this process. In Experiment 4, we also computed MC scenarios of both Experiments 2 and 3, but now we also purposely added one gross error with random sign to a random selected observation in each scenario. In each case, we simulated 50,000 scenarios with one gross error with uniform distribution in the interval magnitude $3–6\sigma_i$, and 50,000 in the interval $6–9\sigma_i$, σ_i being the standard deviation of respective observation, totaling 100,000 scenarios with "weak" correlations and 100,000 of "strong" correlations, adding up to 200,000 scenarios.

The test statistic for outlier identification was the ratio between residuals in ML1 adjustment and respective observation standard deviation, as reported by Klein et al. (2021) and Hekimoglu and Erenoglu (2007). Outlier identification was computed with all five approaches for introducing weights in ML1 formulation. In order to provide a fair comparison among the 5 approaches, we considered the same false positive rate $\alpha = 5\%$ for all of them. To make this possible, the procedure for calculating the critical value for identifying outliers (the value of the test statistic from which the observation is classified as an outlier) by ML1 proposed by Suraci et al. (2021) was applied. Tables 4 and 5 present the results for the 100,000 scenarios of Experiment 2 ("weak" correlations) and Experiment 3 ("strong" correlations), respectively.

Table 3: Mean value of the parameters estimated (mm) (Experiment 3)

	h_B	h_C	h_D
ML1(p)	10.0038	9.9995	10.0034
ML1(P)	10.0038	10.0038	10.0132
ML1($\sqrt{p}$)	10.0051	10.0008	10.0048
ML1(W)	10.0038	10.0038	10.0082
ML1(decorrelation_LS)	10.0036	10.0005	10.0063
LS(P)	10.0010	9.9952	10.0055
LS(P_d)	10.0004	9.9925	10.0055

Table 4: Successful rate in outlier identification (%) – "weak" correlations

	1 outlier $3–6\sigma_i$ (%)	1 outlier $6–9\sigma_i$ (%)	Average rate (%)
ML1(p)	49.97	74.28	62.13
ML1(P)	49.97	74.28	62.13
ML1($\sqrt{p}$)	55.24	85.89	70.57
ML1(W)	55.24	85.89	70.57
ML1(decorrelation_LS)	54.57	85.03	69.80

8 Discussions

8.1 Discussion 1: comparison of parameters estimation of the five ML1 approaches

In Experiment 1, since observations were independent, ML1(p) and ML1($\sqrt{p}$) were useful to validate the results of other ML1 formulations. ML1(P) had the same results of ML1(p), as they are equivalent for independent observations (equation (12)). Similarly, ML1(W) and ML1($\sqrt{p}$) also had the same results. Besides, as expected, proposed formulations ML1(P) and ML1(W) had different results between them, because their adjustments are based on different stochastic models. The most remarkable point to be noted in

Table 5: Successful rate in outlier identification (%) – "strong" correlations

	1 outlier $3–6\sigma_i$ (%)	1 outlier $6–9\sigma_i$ (%)	Average rate (%)
ML1(p)	48.64	73.57	61.11
ML1(P)	30.50	47.05	38.78
ML1($\sqrt{p}$)	53.71	85.06	69.39
ML1(W)	24.49	45.78	35.14
ML1(decorrelation_LS)	35.03	35.67	35.35

Experiment 1, however, is that ML1(decorrelation_LS) had the same results of ML1($\sqrt{p}$) and ML1(W), not ML1(p) and ML1(P), which is in accordance with the theoretical discussion presented. This further confirms that the approaches of Ingram (1911) and Strang and Borre (1997) for unit-weight reduction should not be performed in ML1 if one wants to introduce the weights of P. Actually, for independent observations, the results of unit-weight reduction in ML1 are equivalent to those of the weights of $\sqrt{P} = W$.

In Experiments 2 and 3, we had no reference (ground truth) to validate the results, because: (a) ML1(p) and ML1($\sqrt{p}$) do not consider covariances; (b) ML1(decorrelation_LS), as mentioned, needs further investigations; and, (c) ML1(P) and ML1(W) are being proposed (and still under examination) in this study. In fact, there is no formulation in the literature that is guaranteed to provide a ground truth for ML1 adjustment of correlated observations. Nevertheless, it is possible to identify interesting points from the results of Experiments 2 and 3.

In Experiment 2, ML1(P) and ML1(p) had the same results again, even though only ML1(P) takes covariances into consideration in the adjustment computation. The same occurred between ML1(W) and ML1($\sqrt{p}$). These results suggest that with variances relatively much higher than the magnitude of the covariances, the covariances do not influence the ML1 parameters estimation for the network geometry analyzed, considering the formulation proposed in this study. Since ML1 provides the parameters estimation by finding a special non redundant subset of observations to solve the functional model with no residuals (Amiri-Simkooei, 2018), it is not unexpected that too "weak" correlations may not be able to modify such subset. One might guess that ML1(decorrelation_LS) introduced weights proportional to $\sqrt{p}$ (as in Experiment 1), but its result was not exactly equal to ML1($\sqrt{p}$) because it also captured the "small" covariances. However, since there is no ground truth in Experiment 2, we have no elements to make this or any further claim about ML1(*decorrelation_LS*).

In Experiment 3, we can verify that now covariances did influence the results of ML1(P) and ML1(W), since they were no longer the same of ML1(p) and ML1($\sqrt{p}$) results, respectively. This occurred because the magnitudes of covariances were much closer to the variances than in Experiment 2. Nevertheless, since there is no ground truth in Experiment 3 either, we have no elements to state if ML1(P) have captured covariances in relation to ML1(p) in the most proper way, and if ML1(W) or ML1(*decorrelation_LS*) captured covariances more appropriately in relation to ML1($\sqrt{p}$).

As a partial conclusion of Discussion 1, we first recall that ML1(p) and ML1($\sqrt{p}$) do not take covariances into consideration, and unit-weight reduction techniques originally designed for LS introduce weights of $\sqrt{p}$ (but not to p) even in scenarios with only independent observations, being not appropriate to introduce weights of P. In this context, even though there is no guarantee that our approach is the most appropriate for introducing weights in ML1, the proposed formulations ML1(P) and ML1(W) presented results that seem reasonable: they were equal to ML1(p) and ML1($\sqrt{p}$), respectively, for both cases of only independent observations and of "weak" correlations among observations; and, they allowed covariances to influence parameters estimation in the case of "strong" correlations, which was not feasible with usual formulations ML1(p) and ML1($\sqrt{p}$).

8.2 Discussion 2: comparison of covariances effect between ML1 with proposed formulations and LS

From Table 3, we can also verify the "low" effect (of small magnitude) of covariances in the mean value of parameters estimated by ML1 (considering 200,000 MC simulations). The maximum difference between a parameter computed by ML1(P) or ML1(W), and respective formulation that do not consider covariances ML1(p) or ML1($\sqrt{p}$), respectively, was less than 0.01 mm, being even less than 0.005 mm for all other parameters. In this sense, LS results in Experiments 2 and 3 were important to clarify that, considering the same analysis with 200,000 MC scenarios, this alleged "low" effect also occurs in LS context (with its well-established formulation for introducing weights/covariances). In scenarios of "weak" correlations (Experiment 2), the maximum difference between the mean value of a parameter computed by LS(P) and LS (P_d) was 0.0002 mm. Even in scenarios of "strong" correlations (Experiment 3), the maximum difference was less than 0.003 mm.

Hence, comparing correlation effects in LS and ML1 (the latter with proposed formulations), they both tend to increase as correlations increase, as expected. However, while "weak" correlations had no effect in ML1, "strong" correlations influenced ML1 results more than LS ones. It suggests that even though "weak" correlations tend to have more effect in LS, the influence on ML1 results may grow even more than in LS as correlations become "stronger."

Table 6: RMSE between LS(P) and each of the five ML1 approaches

	Experiment 1	Experiment 2	Experiment 3
ML1(p)	0.0100	0.0088	0.0032
ML1(P)	0.0100	0.0088	0.0069
ML1($\sqrt{p}$)	0.0102	0.0102	0.0040
ML1(W)	0.0102	0.0102	0.0054
ML1(decorrelation_LS)	0.0102	0.0094	0.0034

Moreover, in ML1, those maximum differences for the parameter h_D were always the highest among the three parameters, and were always the smallest for h_B (Table 3). Since station D is connected by observations with lower sum of weights (higher sum of variances) than C, which is connected by observations with lower sum of weights than B, it suggests that the influence of "strong" correlations in ML1 tend to grow more in parameters of the stations connected by observations with lower weights. However, the same is not true for LS. Considering LS adjustments, the highest correlation effects were on the parameter h_c, in both Experiments 2 and 3 (Tables 2 and 3).

8.3 Discussion 3: comparison of parameters estimation of the five ML1 approaches with LS

For a comprehensive analysis of the five ML1 approaches, we must recall that the LS (computed considering the full weight matrix P) is the BLUE for networks free of outliers. Hence, in this case, even though LS and ML1 are two different methods and therefore different results are expected, LS parameters estimation may be considered as a reference and the proximity degree to it may also be a criterion for selecting the optimum ML1 approach. Table 6 shows the root mean squared error (RMSE) between the estimated parameters of LS(P) and each of the five ML1 approaches in Experiments 1, 2, and 3. ML1(p) had the lowest RMSE in all experiments. Therefore, considering the geodetic network analyzed, it is the best choice if a geodesist desires a ML1 parameter estimation closer to the LS one.

8.4 Discussion 4: comparison of outlier identification performance of the five ML1 approaches

In Experiment 4, we tested all five ML1 approaches in outlier identification, the main point in ever applying ML1 in geodetic networks, as mentioned. From Table 4, we can see that ML1(P) and ML1(p) had the same successful rates. Similarly, ML1(W) and ML1($\sqrt{p}$) had the same successful rates between them as well, ML1(*decorrelation_LS*) rate also being very similar. Hence, it is clear that "weak" correlations did not affect outlier identification success rate, and one can safely consider (for simplicity) the observations to be independent, without decreasing ML1 outlier identification performance. From Table 5, however, we can verify that taking "strong" correlations into account adversely affected the outlier identification property of ML1. The average rate of ML1(p) was higher than that of ML1(P) by a significant margin. Similarly, the average rate of ML1($\sqrt{p}$) was higher than that of ML1(W) and of ML1(*decorrelation_LS*) by a significant margin as well. Hence, assuming that observations are independent is the best option for any strength of correlations in the context of outlier identification with ML1. In addition, considering the two approaches ML1(p) and ML1($\sqrt{p}$) that assume observations to be independent, ML1($\sqrt{p}$) had the best performance and is the best choice for the network analyzed.

8.5 Final remarks

This study drew attention to introducing covariances of observations in ML1. It is well-known that the most common approaches in the literature, herein represented by ML1(p) and ML1($\sqrt{p}$), do not take covariances into considerations, being appropriated only to independent observations. Methods for decorrelation/unit-weight reduction of observations originally developed for LS adjustment computations have been previously applied to ML1 in the literature. However, we pointed out that they should not be unrestrictedly performed in the ML1 context, as we have theoretically and experimentally shown that they introduced weights proportional to $\sqrt{p}$ (but not p) even in ML1 adjustment of independent observations.

In this context, we presented an attempt to formulate ML1 expressions that can introduce weights of both independent and correlated observations, they being proportional to p or $\sqrt{p}$. Actually, we proposed two new formulations for

ML1, herein called ML1(P) and ML1(W), extensions of ML1(p) and ML1($\sqrt{p}$), respectively, that can be applied to both independent and correlated observations. Our formulations allowed covariances to influence parameters estimation in ML1, something impossible to accomplish with usual formulations ML1(p) and ML1($\sqrt{p}$).

We compared the effects of covariances in LS and ML1 adjustments (an unprecedented analysis in geodesy). Considering the proposed expressions for ML1 and the network geometry analyzed, "weak" correlations seem to have more effect in LS, but the effect in ML1 tends to be even higher than in LS as the correlations grow.

We also compared the proximity degree to LS of estimated parameters by ML1 with the five approaches tested, and concluded that ML1(p) was the closest to LS. Finally, regarding outlier identification (the usual duty of ML1 in geodetic networks), ML1($\sqrt{p}$) had the best performance. Hence, considering these two practical applications, one can safely assume observations to be independent, but the choice between introducing weights proportional to p or $\sqrt{p}$ is an important issue.

Since ML1 solution may not be unique (Abdelmalek and Malek 2008), we highlight that the conclusions presented refer to ML1 solution by the *simplex* method of linear programming. For future works, one should try to collect further theoretical and experimental evidences of the validity of the new formulations presented, and experiments of this article must be extended to other geodetic networks to further support the points raised. Besides, new approaches for introducing covariances in ML1 more appropriated than the unit-weight reduction processes originally from LS and than our proposed formulations must be investigated.

Acknowledgments: The third author thanks CNPq for the Research Productivity Grant (Process 313699/2021-6).

Funding information: This work was supported by the Department of Science and Technology of the Brazilian Army.

Conflict of interest: Authors state no conflict of interest.

References

Abdelmalek, N. and W. Malek. 2008. *Numerical linear approximation in C*. CRC Press, London.

Amiri-Simkooei, A. 2003, "Formulation of L1 Norm Minimization in Gauss-Markov models." *Journal of Surveying Engineering* 129(1), 37–43.

Amiri-Simkooei, A. 2018, "On the use of two L1 norm minimization methods in geodetic networks." *Earth Observation and Geoinformation Management* 2(1), 1–8.

Baselga, S., I. Klein, S. S. Suraci, L. C. Oliveira, M. T. Matsuoka, and V. F. Rofatto. 2020. "Performance comparison of least squares, iterative and global L1 norm minimization and exhaustive search methods for outlier detection in leveling networks." *Acta Geodynamica et Geomaterialia* 17(4), 425–38.

Dantzig, G. 1963. *Linear programming and extensions*. Princeton University Press, Princeton.

Gauss, C. F. 1809. *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*. Hamburg: Friedrich Perthes and I.H. Besser.

Ghilani, C. D. 2010. *Adjustment computations: spatial data analysis*. 5. ed., John Wiley & Sons, Hoboken, New Jersey.

Golub, G. H. and C. F. V. Loan. 1996. *Matrix computations*. 3 ed., Johns Hopkins University Press, Baltimore and London.

Hekimoglu, S. and R. C. Erenoglu. 2007. "Effect of heteroscedasticity and heterogeneousness on outlier detection for geodetic networks." *Journal of Geodesy* 81, 137–48.

Higham, N. J. 2009. "Cholesky factorization." *Wiley Interdisciplinary Reviews: Computational Statistics* 1(2), 251–4.

Inal, C., M. Yetkin, S. Bulbul, and B. Bilgen. 2018. "Comparison of L1 norm and L2 norm minimisation methods in trigonometric levelling networks." *Technical Journal* 25(1), 216–21.

Ingram, E. L. 1911. *Geodetic surveying and the adjustment of observations* (*method of least squares*). McGraw-Hill Book Company, New York.

Junhuan, P. 2005. "The asymptotic variance–covariance matrix, Baarda test and the reliability of L1-norm estimates." *Journal of Geodesy* 78(11), 668–82.

Klein, I., M. T. Matsuoka, M. P. Guzatto, F. G. Nievinski, M. R. Veronez, and V. F. Rofatto. 2019. "A new relationship between the quality criteria for geodetic networks." *Journal of Geodesy* 93(4), 529–44.

Klein, I., S. S. Suraci, L. C. Oliveira, V. F. Rofatto, M. T. Matsuoka, and S. Baselga. 2021. "An attempt to analyse Iterative Data Snooping and L1-norm based on Monte Carlo simulation in the context of leveling networks." *Survival and Reviews* 54(382), 70–8.

Knight, N. L., J. Wang, and C. Rizos. 2010. "Generalised measures of reliability for multiple outliers." *Journal of Geodesy* 84, 625–35.

Koch, K.-R. 1999. *Parameter estimation and hypothesis testing in linear models*. Springer Science & Business Media, New York.

Legendre, A. 1805. *Nouvelles Méthodes pour la Détermination des Orbites dês Comètes*. Paris: Chez Courcier.

Marshall, J. 2002. "L1-norm pre-analysis measures for geodetic networks." *Journal of Geodesy* 76(6), 334–44.

Prószyński, W. 2010. "Another approach to reliability measures for systems with correlated observations." *Journal of Geodesy* 84, 547–56.

Rofatto, V. F., M. T. Matsuoka, I. Klein, M. R. Veronez, M. L. Bonimani, and R. Lehmann. 2020. "A half-century of Baarda's concept of reliability: A review, new perspectives, and applications." *Survival and Reviews* 52(372), 261–77.

Strang, G. and K. Borre. 1997. *Linear algebra, geodesy, and GPS*. Wellesley-Cambridge Press, Wellesley.

Schaffrin, B. 1997. "Reliability measures for correlated observations." *Journal of Surveying Engineering* 123(3), 126–37.

Suraci, S. S., L. C. Oliveira, and I. Klein. 2019. "Two aspects on L1-norm adjustment of leveling networks." *Revista Brasileira de Cartografia* 71(2), 486–500.

Suraci, S. S., L. C. Oliveira, I. Klein, V. F. Rofatto, M. T. Matsuoka, and S. Baselga. 2021. "Monte Carlo-based covariance matrix of residuals and critical values in minimum L1-Norm." *Mathematical Problems in Engineering* 2021(1):1–9. Doi: 10.1155/2021/8123493.

Teunissen, P. J. G. 2018. "Distributional theory for the DIA method." *Journal of Geodesy* 92, 59–80.

Yang, L., Y. Shen, B. Li, and C. Rizos. 2021. "Simplified algebraic estimation for the quality control of DIA estimator." *Journal of Geodesy* 95, 14.

Yetkin, M. and C. Inal. 2011. "L1 norm minimization in GPS networks." *Survival and Reviews* 43(323), 523–32.