Improving absolute gravity estimates by the $L_p$-norm approximation of the ballistic trajectory

V D Nagornyi, S Svitlov and A Araya

1 Metromatix, Inc., 111B Baden Pl, Staten Island, NY 10306, USA
2 Institut für Erdmessung, Leibniz Universität Hannover, Schneiderberg 50, D-30167 Hanover, Germany
3 Earthquake Research Institute, University of Tokyo, 1-1-1, Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan

E-mail: vn2@member.ams.org, svitlov@ife.uni-hannover.de and araya@eri.u-tokyo.ac.jp

Received 25 August 2015, revised 18 December 2015
Accepted for publication 6 January 2016
Published 2 February 2016

Abstract
Iteratively re-weighted least squares (IRLS) were used to simulate the $L_p$-norm approximation of the ballistic trajectory in absolute gravimeters. Two iterations of the IRLS delivered sufficient accuracy of the approximation without a significant bias. The simulations were performed on different samplings and perturbations of the trajectory. For the platykurtic distributions of the perturbations, the $L_p$-approximation with $3 < p < 4$ was found to yield several times more precise gravity estimates compared to the standard least-squares. The simulation results were confirmed by processing real gravity observations performed at the excessive noise conditions.

Keywords: absolute gravimeter, $L_p$-norm approximation, iteratively re-weighted least squares, antikurtosis

(Some figures may appear in colour only in the online journal)

1. Introduction

Absolute ballistic gravimeters measure gravity acceleration by tracking the free motion of the test mass in the gravity field. Having the distances $SS_i$, ...,$SS_N$ covered by the test mass over the time intervals $TT_i$, ...,$TT_N$, the acceleration is found as a parameter of some trajectory model fitted to the data pairs $(T_i, S_i)$ by methods of the regression analysis. One of the most commonly used models describes the unperturbed vertical motion of the test mass in the gravity field

$$z_i = z_0 + V_0 T_i + g T_i^2/2. \quad (1)$$

The expression (1) represents a linear regression model, because it is defined by a linear combination of the estimated parameters $z_0, V_0, g$. The estimates are usually found by minimising the sum of the squared discrepancies between the measured data and the model:

$$(z_0, V_0, g) : \sum \epsilon_i^2 \to \min, \quad \text{where} \quad \epsilon_i = S_i - z_i. \quad (2)$$

Minimised here is the square norm ($L_2$-norm) of the vector $\epsilon = \{ \epsilon_i \}$, which is a special case of the more general $L_p$-norm

$$\| \epsilon \|_p = \left( \sum |\epsilon|^p \right)^{1/p}, \quad p \geq 1. \quad (3)$$

The tradition of minimising the sum of squares (i.e. finding the least squares) has a long and rich history dating back to the beginning of the 19th century. At that time the main reason for using the least squares was their computational simplicity, as for $p = 2$ the fitting of a linear model is reduced to finding a unique solution for a full-rank system of linear equations. At any other $p$ the equations become non-linear in the estimated parameters and require involved computations to find the solution.

Another reason for using the least squares relates to a different kind of linearity: linearity of the estimates with respect to the dependent variable ($S_i$ in our case), which implies no bias of the estimates. Moreover, according to the Gauss–Markov theorem, of all the linear/unbiased estimates, the least-squares deliver the most precise one, in case
the errors in the observations are uncorrelated, have zero mean and constant variance. This is especially important in metrology, where the bias may lead to incorrect estimates of the measurand and its uncertainty.

Nowadays, when advanced computing is no longer a problem, the \( L_p \)-approximation with \( p = 2 \) is widely used \([1–7]\). Many researchers have noticed that the value of \( p \) delivering the lowest variance to the \( L_p \)-estimates decreases as the kurtosis of the error distribution grows; there are formulae suggested for this dependence \([8, 9]\). Most results have been obtained for the values \( 1 < p < 2 \), which provide better estimates when the noises have heavily-tailed distributions. In trajectory tracking, however, more prevalent are limited-band noises, for which the properties of the \( L_p \)-approximation are less known.

The purpose of this work is to investigate the feasibility and possible advantages of approximating the gravimeter’s test mass trajectory in different \( L_p \)-norms.

The paper has the following structure. Section 2 discusses the implementation of the iteratively re-weighted least squares (IRLS) for the \( L_p \)-norm approximation. Section 3 describes the Monte-Carlo simulation of gravity estimates in the \( L_p \)-norm. A real-life example of improving the estimates by the \( L_p \)-approximation is discussed in section 4. We summarise our study in section 5.

2. The \( L_p \)-norm approximation by the iteratively re-weighted least-squares

We implement the \( L_p \)-norm approximation by performing the weighted least-squares (WLS) approximation, in which the weights are defined in a special way \([10]\). For model (1), the WLS solution is determined by

\[
\mathbf{\xi} = A^* \mathbf{S},
\]

where \( \mathbf{\xi} = (\zeta_0, \mathbf{v}_0, g)^T \) is the vector of the unknown parameters, \( \mathbf{S} = (S_1, S_2, ..., S_N)^T \) is the vector of the observations, \( A^* \) is the Moore–Penrose pseudo-inverse matrix calculated as

\[
A^* = (A^T \mathbf{V} A)^{-1} A^T \mathbf{V},
\]

where

\[
A = \begin{pmatrix}
1 & T_1 & T_1^2 / 2 \\
1 & T_2 & T_2^2 / 2 \\
: & : & : \\
1 & T_N & T_N^2 / 2
\end{pmatrix}
\]

is the experiment design matrix,

\[
\mathbf{V} = \text{diag}(v_1, v_2, ..., v_N)
\]

is the diagonal matrix of weights. The approximation in the \( L_p \)-norm is achieved when the weights \( v_i \) are related to the discrepancies \( \epsilon_i \) like \([10]\)

\[
v_i = |\epsilon_i|^{p - 2}, \quad \text{for } p \geq 2,
\]

For \( p = 2 \), the weights (8) turn into units, yielding the standard least-squares solution. Formula (8) cannot be used directly to find the weights \( v_i \) because the weights need to be known before the approximation, while the residuals \( \epsilon_i \) are known only after the approximation is done. Because of that, the WLS is usually applied several times leading to the iteratively re-weighted least squares. The iterations start with the weights \( v_i = 1 \), i.e. with the regular least-squares. The residuals \( \epsilon_i \) are then used to build the new weights

\[
V = \text{diag}(|\epsilon_1|^{p - 2}, |\epsilon_2|^{p - 2}, ..., |\epsilon_N|^{p - 2}),
\]

which are then applied to approximate the original data. The iterations can be repeated, but the process is known to sometimes have a problematic convergence. Although different approaches exist to improve the convergence \([10]\), the simulations found that two iterations always produce an approximation sufficiently accurate for our purposes, with the bias not discernible in random noise. For better computational stability we normalised the weights (8), so they would range from 0 to 1:

\[
v_i = \left( \frac{|\epsilon_i|}{M} \right)^{p - 2}, \quad p \geq 2,
\]

where

\[
M = \max|\epsilon_i|.
\]

If \( p < 2 \), the powers in (8) become negative, and to prevent zero division errors another modification of the weights was necessary:

\[
v_i = \left( \max(|\epsilon_i|, rM) / rM \right)^{p - 2}, \quad p < 2,
\]

where \( r = 0.001 \) is a regularisation parameter establishing the threshold of smallness as fraction \( r \) of the maximum residual \( M \). According to (12), the residuals below the threshold are replaced with the threshold value \( rM \). A similar way of regularisation is implemented in \([7]\).

3. Modelling the properties of the absolute gravity estimates in the \( L_p \)-norm

3.1. Organisation of the simulation experiment

For the simulations, we added random perturbations to the parabolic trajectory (1) with known acceleration \( g = 9.8 \text{m/s}^2 \), and then recovered the acceleration from the perturbed parabola using the \( L_p \)-norm approximation (figure 1). The values of \( p \) ran from 1 to 6 with increments of 0.1, so the standard least-squares solution (\( p = 2 \)) was also available from the simulation and was used for comparison.

To build the parabola, 700 data points \([T_1, ..., T_{700}]\) representing 699 time intervals with the common start at \( T_1 \) were used, spanning the total time \( T_{700} = 0.22 \text{s} \). We considered three distributions of the \( T_i \)-points (experiment designs) found

\[5\] The other parameters were \( \zeta_0 = 1 \text{nm} \) and \( V_0 = 0 \text{m/s}^{-1} \). Zero initial velocity was chosen to maximise the difference between the experiment designs.
in different types of absolute gravimeters (figure 2). The EST (equally spaced in time) design has the uniform distribution of the \(T_i\)-points and is used in both free-fall and rise-and-fall gravimeters. The ESD\(_{FF}\) design has the minimal density at the start, linearly increasing towards the end of the interval. The ESD\(_{RF}\) design has the minimal density at the centre, linearly increasing towards the start and the end of the interval. These two designs are found in free-fall (FF) and rise-and-fall (RF) gravimeters with the levels equally spaced in distance (ESD).

For every type of perturbation we simulated 3000 drops and, after approximating the trajectory in different \(L_p\)-norms, analysed the results. Every realisation of the simulated noise was used in all three experiment designs by assigning either EST, ESD\(_{FF}\), or ESD\(_{RF}\) realisations of the vector \(T = [T_1, ..., T_{100}]\) to the same noise vector \(\epsilon = [\epsilon_1, ..., \epsilon_{100}]\).

The simulations and data analysis were performed with MATLAB\textsuperscript{®} software.

3.2. Simulation results for the random uncorrelated perturbations

The trajectory noise in absolute gravimeters comes from a number of sources, such as ground vibrations, laser stabilisation, digital signal acquisition, fringe counting etc, so the combined noise most often is of the Gaussian type. Sometimes, however, individual sources can dominate, creating other noise distributions. We did the simulations for several noise types, all having the standard deviation of 1 nm.

For all noise distributions, the distribution of the \(g\)-estimates was normal for any \(p\). For every noise distribution there
existed an optimal value of $p$ delivering minimal variance of the estimate (figure 3(a)). The optimal $p$ ranged from 1.4 to 3.3 depending on the noise distribution. This result deviates from the theory [11] predicting no optimal $p$ for at least the Uniform distribution.

For the numeric characteristic of the distribution shape, we used antikurtosis $\chi$, because it can only assume values from 0 to 1 for any distribution. We found it more convenient than the traditionally used kurtosis, which can range from 1 to infinity. The optimal $p$ increased with antikurtosis (table 1), but not always the variance at the optimal $p$ improved substantially. We visualised the improvement by plotting the variance for every $p$ (figure 3(b)) divided by the variance of the standard least-squares. The factor of the variance improvement is shown in the plot of the inverse ratio (figure 3(c)), which is equivalent to the relative efficiency of the $L_p$-estimate compared to the standard least-squares.

Table 1. Results of the simulation of the absolute gravity estimates by approximating the trajectory in different $L_p$-norms.

| Noise Distribution | Anti-kurtosis $\chi$ | Best $p$ | Rel. eff.\(^a\) for the best $p$ |
|--------------------|----------------------|---------|-----------------------------------|
| Laplace            | 0.41                 | 1.4     | 1.3                               |
| Normal             | 0.58                 | 2.0     | 1.0                               |
| Triangle           | 0.65                 | 2.5     | 1.1                               |
| Uniform            | 0.75                 | 3.3     | 2.8                               |
| Arcsin             | 0.82                 | 3.3     | 7.0                               |

Note: The case of the random uncorrelated perturbations of the trajectory.

\(^a\) Determined as the ratio of the variance of the $L_2$-estimate to the variance of the $L_p$-estimate for the same experiment design.

For the numeric characteristic of the distribution shape, we used antikurtosis $\chi$, because it can only assume values from 0 to 1 for any distribution. We found it more convenient than the traditionally used kurtosis, which can range from 1 to infinity. The optimal $p$ increased with antikurtosis (table 1), but not always the variance at the optimal $p$ improved substantially. We visualised the improvement by plotting the variance for every $p$ (figure 3(b)) divided by the variance of the standard least-squares. The factor of the variance improvement is shown in the plot of the inverse ratio (figure 3(c)), which is equivalent to the relative efficiency of the $L_p$-estimate compared to the standard least-squares.

$^a$The antikurtosis $\chi$ relates to the kurtosis $\beta_2$ as $\chi = 1/\sqrt{\beta_2}$ [12].
According to table 1, the $L_p$-estimates with $p > 2$ can provide a significant increase in efficiency when the errors are distributed by the Uniform or Arcsine law. In trajectory tracking such errors are often caused by harmonic perturbations of the reference frame. We modelled the perturbations by the sum of a sinusoid and Gaussian noise. In every simulated drop, a sinusoid with a given frequency $f$ and amplitude of 1.41 nm was assigned an initial phase randomly taken from the uniform distribution on $0, 2\pi$, and then sampled according to the experiment design. The sinusoid samples were added by the Gaussian noise, one noise realisation for all three experiment designs. The signal-to-noise ratio (SNR) is defined as the quotient of the standard deviations of the sinusoid and the noise, so the infinite SNR means no random noise.

The results of the simulation for the frequencies of 17 Hz, 35 Hz, 55 Hz and the SNR values of $\infty$, 100, 10, 2 are shown in table 2. The antikurtosis $\chi$ of the perturbation was from 0.70 to 0.82, typical for the range occupied by the Uniform and Arcsine distributions. The resulting $g$-estimates passed the normality test (Lilliefors, 5%) only for some low SNR values. The optimal $p$ and relative efficiencies for which normality was observed are shaded in table 2. Like for the uncorrelated noises (table 1), there always existed an optimal $p$ delivering the lowest value to the variance of the $L_p$-estimate. The optimal $p$ was between 2.8 and 3.8, yielding the variance decrease of the estimate between 1.2 and 30.6 times. The values of $p$ and the efficiency gain have a complicated dependence on the signal frequency, the experiment design and the SNR, so the values have to be evaluated experimentally in every practical case.

### 4. Example of using the $L_p$-estimates on real data

The field measurement of absolute gravity are often performed under harsh geophysical conditions that deteriorate the performance of the instruments. Often it is not possible...
to repeat the measurements or to improve the conditions, so the researchers have to make the best possible gravity estimates using the available data. As an example, we consider a series of measurements taken by the TAG-1 absolute gravimeter developed by the Earthquake Research Institute at the University of Tokyo. The instrument features a laser interferometer with a built-in accelerometer for the independent measurement and correction of the ground vibrations, and a homodyne quadrature fringe signal detector [13, 14]. The free-fall trajectory of about 12 cm is sampled with the frequency of 10 MHz (EST-design). A typical drop-to-drop scatter of the g-values at the Esashi station is about 20 μGal, so the standard deviation of the mean below 1 μGal is reached in several hundreds drops. The bandwidth of the built-in accelerometer is up to 4 Hz, so the increased intensity of the high-frequency seismic noise could significantly disturb the measurements. Figure 5(a) shows one such noisy set of gravity estimates taken on December 23 2011 at the ‘Esashi’ gravity station, after applying standard tidal, atmospheric and polar motion corrections. To obtain the \( L_p \)-estimates, the trajectory was decimated to 20 kHz, the standard speed-of-light and gradient corrections were applied to the time and distance coordinates, respectively, then the gravity was found using the standard least-squares approximation. The set fails the normality test, making the standard ‘three-sigma’ rejection too restrictive for the obtained heavily-tailed distribution of the \( L_2 \)-estimates. However, neither four, nor five sigmas made much difference to the mean \( g \)-value or its standard deviation, which remained at about 3 μGal (table 3).

We then approximated the trajectory in the \( L_p \)-norm for \( p \) ranging from 1 to 6 (see section 2). For each \( p \) we obtained the variance of the set gravity estimate and plotted the relative efficiency (figure 6). According to the plot (figure 6(a), curve 1), the variance at \( p = 3.5 \) decreased several times w.r.t. \( p = 2 \), reducing the standard deviation of the mean to about 1 μGal (table 3). Based on the simulations (section 3), the peak of the efficiency at \( p > 2 \) is suggestive of the residuals with non-Gaussian distributions. We investigated the residuals and found the values of antikurtosis ranging from 0.487 to 0.813, which corresponds to the wide range of distributions, from Laplace to Arcsine (figure 7). A likely reason for the non-Gaussian errors was the resonant oscillations of the reference mirror caused by excessive seismic noise. The most prevalent in the set were residuals with the Gaussian and Arcsine distributions. According to the simulations (section 3), the most efficient \( L_p \)-approximations for the Arcsine-noised trajectory were achieved for the values of \( p \) of between three and four, which explains the decrease in uncertainty. However, for the Gaussian noises the approximation with \( p \) in that range can be several times less efficient than the standard

\[
1 \text{ μGal} = 10^{-8} \text{ ms}^{-2}.
\]
least-squares (figure 4, second from the top). This led us to the idea of splitting the set into two parts with predominantly Gaussian and Arcsine noises, and independently optimising the $L_p$-norm on each set. We split the groups at the antikurtosis $\chi = 0.65$, approximately in the middle of two peaks in figure 7. The gravity estimates in each group passed the normality test (Lilliefors, 5%), so further processing was carried out with the $3\sigma$ rejection, each group with its own $\sigma$. The splitting increased the efficiency of the estimate for the Arcsine group with $p = 3.5$ from 10.0 to 12.5 (figure 6 (a), 2). For the Gaussian group we used the standard least-squares with $p = 2.0$, which helped to avoid the decrease in efficiency at $p = 3.5$ for that group (figure 5(b)). Combining the estimates for both groups reduced the standard deviation of the weighted mean to 0.76 $\mu$Gal (figure 5(c)). The gravity estimates obtained with different approximations are summarised in table 3.

5. Conclusions

We simulated the approximation of the ballistic trajectory in different $L_p$-norms, in view of using it to improve the absolute gravity estimates. To perform the approximation, the iteratively re-weighted least squares method was implemented. The estimates were simulated for three different plans of the experiment corresponding to two types of ballistic trajectory (free-fall, rise-and-fall) and two schemes of data location (equally spaced in time or in distance). Each plan was applied to several perturbation types, including the cases of random uncorrelated noises and a harmonic in the Gaussian noise. As an application example, the simulation results were used to improve the absolute gravity estimates obtained at the ‘Esashi’ gravity station by a TAG-1 gravimeter.

The following conclusions can be drawn from the study.

(i) The iteratively re-weighted least-squares with two iterations approximate the ballistic trajectory in the $L_p$-norm with an accuracy sufficient for the purposes of absolute gravimetry. No statistically significant bias of the estimate was found for the wide range of trajectory noises.

(ii) The $L_p$-norm approximation of the ballistic trajectory with $p > 2$ estimates gravity more precisely than the standard least-squares ($p = 2$) if the trajectory perturbations deviate from the Gaussian towards more platykurtic distributions, i.e. those with a flattish or concave shape.

(iii) The observed efficiency maximum for certain values of $p$ on platykurtic distributions disagrees with the prediction of the work [11], where the monotonous increase in efficiency with norm power $p$ is obtained for the Uniform distribution. This could possibly be explained by a limited iterations of our IRLS procedure, resulting in a ‘quasi’ rather than the ‘exact’ $L_p$-approximation.

(iv) For the same measurement interval $T$ and fixed number of data points $N$, the EST sampling of the free-fall trajectory makes the gravity estimates about twice as efficient as those with the data equally spaced in distance for any $p$. This points to an obvious way of improving the modern free-fall absolute gravimeters, which predominantly use the ESD scheme. For the symmetric rise-and-fall trajectory, the efficiency of the EST and ESD sampling is about the same.
The present study is a first attempt to obtain gravity estimates by approximating the ballistic trajectory in the \( L_p \)-norm with \( p \neq 2 \). Many important questions are yet to be answered by future works. The questions include, but are not limited to:

(i) The scope of the applicability of the \( L_p \)-norm approximation. For the purposes of absolute gravimetry we investigated only the estimates of the quadratic coefficient of the second-order polynomial model. The results can be applied to other models to build more efficient estimates in a wider range of applications.

(ii) Auto-correlation of the residuals. The main source of the platykurtic distributions is low-frequency noises. Our simulations of the correlated Arcsine noises and processing of the real data suggest that the efficiency of the \( L_p \)-approximation is defined mostly by the noise distribution and is not deteriorated by the auto-correlation of the residuals. Still, the influence of the auto-correlation calls for a more thorough investigation, including a comparison with other algorithms of low-frequency noise filtering.

(iii) Contaminated distributions within a drop. Further investigations are desirable to better understand the properties of the \( L_p \)-norm approximation for mixed noise distributions, such as harmonics in Gaussian noise.

(iv) Contaminated distributions of the set and set splitting. When a set has drops with different noise types, a single value of \( p \) may not yield the best \( L_p \)-approximations for all drops. In the example (section 4) we split the set in two at the centre of the anti-kurtosis histogram. Although we gained a certain decrease in uncertainty, there may exist better ways of set splitting that are worth investigating.

(v) The reasons for the departure of the simulated results from those predicted in [11] have to be investigated to better understand the properties of the IRLS algorithm of the \( L_p \)-approximation.

Acknowledgment

S Svitlov’s contribution was supported by the German Research Foundation (DFG, Germany) under the project SV 86/1-2.

References

[1] Penneucci F and Callegaro L 2006 Between the mean and the median: the \( L_p \) estimator Metrologia \textbf{43} 213
[2] Willink R 2007 On the \( L_p \) estimation of a quantity from a set of observations Metrologia \textbf{44} 105
[3] Marx C 2013 On resistant \( L_p \)-norm estimation by means of iteratively re-weighted least-squares J. Appl. Geod. \textbf{7} 1
[4] Chen J-M and Chen B-S 2000 System parameter estimation with input/output noisy data and missing measurements IEEE Trans. Signal Process. \textbf{48} 1548–58
[5] Walach E and Widrow B 1984 The least mean fourth (LMF) adaptive algorithm and its family IEEE Trans. Inf. Theory \textbf{30} 275–83
[6] Gui G and Adachi F 2014 Sparse least mean fourth algorithm for adaptive channel estimation in low signal-to-noise ratio region Int. J. Commun. Syst. \textbf{27} 3147–57
[7] Scales J A and Gerszenkorn A 1988 Robust methods in inverse theory Inverse Problems \textbf{4} 71
[8] Money A H, Affleck-Graves J F, Hart M L and Barr G D I 1982 The linear regression model: \( L_p \) norm estimation and the choice of \( p \) Commun. Stat.—Simul. Comput. \textbf{11} 89–109
[9] Sposito V A, Hand M L and Skarpness B 1983 On the efficiency of using the sample kurtosis in selecting optimal \( L_p \) estimators Commun. Stat.—Simul. Comput. \textbf{12} 265–72
[10] Burrus C 2012 Iterative re-weighted least-squares OpenStax-CNX Website http://cnx.org/content/m45285/1.12/
[11] Nyquist H 1983 The optimal \( L_p \) norm estimator in linear regression models Commun. Stat.—Theory Methods \textbf{12} 2511–24
[12] Novitsky P V and Zograf I A 1991 Estimation of Errors of Measurements Results (Leningrad: Energoatomizdat) p 304 (in Russian)
[13] Araya A, Sakai H, Tamura Y, Tsubokawa T and Svitlov S 2014 Development of a compact absolute gravimeter with a built-in accelerometer and a silent drop mechanism Proc. of the Int. Association of Geodesy Symp. on Terrestrial Gravimetry: Static and Mobile Measurements (Saint Petersburg, Russia, 17–20 September 2013) p 98
[14] Svitlov S and Araya A 2014 Homodyne interferometry with quadrature fringe detection for absolute gravimeter Appl. Opt. \textbf{53} 3548