NONOPTIMALITY LEVELS IN NUMERICAL IMPLEMENTATION OF THE LEAST ABSOLUTE DEVIATIONS METHOD

Alexander Matasov
Faculty of Mechanics and Mathematics
Moscow Lomonosov State University
Russia
alexander.matasov@gmail.com

Pavel Akimov
Faculty of Mechanics and Mathematics
Moscow Lomonosov State University
Russia
akmpavel@rambler.ru

Abstract
This article is devoted to the variational problem of the least absolute deviations method. The main goal is to construct a reliable and effective nonoptimality level for a current iteration. The duality theory for convex variational problems are applied to the least absolute deviations method to obtain upper bounds for the nonoptimality levels.

Key words
Least absolute deviations method, estimation, duality theory, nonoptimality levels, satellite navigation.

1 Introduction

Anomalous measurement errors (outliers) are frequently occur in the processing of measurement data. In this case, the least absolute deviations (LAD) method is an effective estimation method. However, various numerical algorithms for the implementation of this method are iterative and the question of convergence rate is not always clear. The present paper is devoted to constructing the nonoptimality levels for the current iteration. These levels allow us to guarantee the approximation accuracy and in so doing to give a reliable stopping criterion for the iteration process.

2 Least Absolute Deviations

Consider the measurements of the form

$$z = H^T q + r, \tag{1}$$

where $q \in \mathbf{R}^n$ is an unknown vector to be estimated, $z \in \mathbf{R}^N$ is the measurement vector, $H = (H_1, \ldots, H_N)$ is a specified matrix of dimension $n \times N$ $(N \geq n)$, $r \in \mathbf{R}^N$ is the vector of measurement errors.

In accordance with the least absolute deviations method the estimation of q reduces to the solution of the following variational problem:

$$I_0 = \min_q I(q), \qquad I(q) = \sum_{i=1}^{N} |z_i - H_i^T q|. \tag{2}$$

In other words, the l_1-norm of the residual vector is minimized.

The LAD method permits to decrease the influence of the outliers in the desired estimate. This is the distinction of the LAD method from the classical least squares method that has a good averaging property but is not robust to anomalous measurement errors.

There are several ways for solving the variational problem (2). Among them one can cite the reduction to linear programming [Boyd, Vandenberghe, 2004; Ekeland, Temam, 1976]. The linear programming problem can be solved by the well-known simplex method as well as by the more modern interior-point method [Boyd, Vandenberghe, 2004]. Another method for solving the problem (2) is the so-called Weiszfeld method [Weiszfeld, 1937; Weiszfeld, Plastria, 2008], which is briefly described in the next section. All these methods are iterative. So, there appears the necessity to obtain the stopping criterion, particularly for processing the large number of measurements ($N \gg 1$).

3 Weiszfeld Method

The Weiszfeld method (it is also called the variationally-weighted quadratic approximations method) is popular in engineering literature [Mudrov, Kushko, 1983]. It involves the iterations each defined by the minimization of the specially constructed quadratic form in q [Weiszfeld, 1937; Weiszfeld, Plastria, 2008] (instead of the initial nonsmooth function $I(q)$):

$$q^k = \operatorname{argmin}_q \left(\sum_{i=1}^N W_i^{k-1}|z_i - H_i^T q|^2\right), \quad k = 0, 1, \ldots. \tag{3}$$

Here k is the iteration number, q^{k-1} is the estimate vector obtained at the previous iteration, $W_i^{k-1} = \left(|z_i - H_i^T q^{k-1}|\right)^{-1}$ is a weight coefficient that corresponds to the i-th entry of the residual vector at the previous step. The initial vector q^0 (with $k = 0$) is chosen from a priori (maybe rough) information on the unknown parameter. Thus the minimization of the quadratic form can be considered as the weighted least squares (WLS) method with the specified weight coefficient matrix $W = \operatorname{diag}(W_1^{k-1}, \ldots, W_N^{k-1})$:

$$J_0 = \min_q \left((z - H^T q)^T W (z - H^T q)\right). \tag{4}$$

The chosen method of constructing the quadratic form is confronted by numerical difficulties under small values of the residual vector entries $z - H^T q^{k-1}$. Whereas one of these entries becomes zero, it is impossible to calculate the coefficient $W_i^{k-1} = \left(|z_i - H_i^T q^{k-1}|\right)^{-1}$ at a given step. In order to resolve this difficulty, the regularization is applied (the details are omitted). Obviously, the Weiszfeld method is very attractive from the computation point of view. Since there is no mathematical guarantee that this algorithm always converges, it is important to evaluate how much the vector q^k at a step k is close (in cost function) to the optimal one; in other words, how much the current value $I(q^k)$ is greater than the unknown optimal value $I_0 = I(q^*)$ where q^* is the exact solution of the problem (2).

We shall call the value

$$\Delta = \frac{I(q^k)}{I(q^*)} = = \frac{I(q^k)}{I_0} \tag{5}$$

the nonoptimality level [Matasov, 1998] of the k-th iteration of the algorithm (e.g., of the variationally-weighted quadratic approximation method).

The exact value of the nonoptimality level is unknown since the cost optimal value I_0 is unknown. However if we succeed in the evaluation of Δ from above: $\Delta \leq \Delta_0$, and Δ_0 turned out to be close to 1, then it is possible to assert that $I(q^k)$ is little different from I_0. Then q^k can be chosen as a successful approximation for the problem (2) solution. So, the calculation of the upper bound for the nonoptimality level of the current iteration provides the reliable stopping criterion for the calculation of the sequence $\{q^k\}$. The desired accuracy for Δ_0 can vary for different classes of problems. For an example below an appropriate stopping condition is $\Delta_0 \leq 1 + 10^{-6}$. In what follows we shall consider two approaches that permit to evaluate Δ quite accurately; one of them can be used not only for the Weiszfeld method but for other iteration algorithms for solving the problem (2) as well.

4 Levels of Nonoptimality

In the basis of the both approaches the duality theory for convex variational problems lies. With this theory we can estimate from above the optimal cost I_0 for the LAD method (2). Introduce the notation $\|\cdot\|_1$ and $\|\cdot\|_\infty$ for the l_1- and l_∞- norm respectively.

For the variational problems under consideration the so-called dual problems can be put into correspondence [Boyd, Vandenberghe, 2004; Ekeland, Temam, 1976].

THEOREM 1. *The problem dual to* (2) *has the form:*

$$I^0 = \max_\lambda z^T \lambda, \qquad \lambda \in \mathbf{R}^N, \tag{6}$$

under the constraints

$$H\lambda = 0, \qquad \|\lambda\|_\infty \leq 1.$$

Moreover, the duality relation holds: $I_0 = I^0$.

Note that the problem (6) is the maximization problem; it allows us to propose the following estimate for the nonoptimality level [Boyd, Vandenberghe, 2004; Matasov, 1998]:

$$\Delta = \frac{\|z - H^T q^k\|_1}{I_0} = \frac{\|z - H^T q^k\|_1}{I^0} \leq \\ \leq \frac{\|z - H^T q^k\|_1 \cdot \|\tilde{\lambda}\|_\infty}{|z^T \tilde{\lambda}|}, \tag{7}$$

where $\tilde{\lambda} \in \mathbf{R}^N$ ia a vector that satisfies the condition $H\tilde{\lambda} = 0$. One possibility of this idea application is presented in the following theorem.

THEOREM 2. *Let q^k be the solution of the weighted least squares method* (4). *Then for the nonoptimality level Δ the following inequality holds:*

$$\Delta \leq \Delta_0^{(1)} \quad \textit{where} \\ \Delta_0^{(1)} = \frac{\|z - H^T q^k\|_1 \, \|W(z - H^T q^k)\|_\infty}{(z - H^T q^k)^T W (z - H^T q^k)}. \tag{8}$$

At the derivation of inequality (9) the solution of the problem dual to the weighted least squares problem (4) was chosen as $\tilde{\lambda}$.

Another version of the upper bound for Δ is also based on the choice of an appropriate vector $\tilde{\lambda}$ that can obtained from the following assertions.

THEOREM 3. *Let* $\operatorname{rank} H = n$. *Then there exists a solution q^* of the problem* (2) *such that n entries of the* *residual vector $z_i - H_i^T q^*$ equal zero. In these residuals, the associated vectors H_i are linear independent.*

This solution q^* is closely linked to the solution of the dual problem (6).

THEOREM 4. *Let q^* be a solution of the LAD problem* (2). *Vector $\hat{\lambda}$ is the solution of the problem* (6) *if and only if the following relations for the entries $\hat{\lambda}_i$ hold:*

$$\hat{\lambda}_i = \mathrm{sign}(z_i - H_i^T q^*) \qquad \textit{if} \quad z_i - H_i^T q^* \neq 0, \\ H\hat{\lambda} = 0, \qquad \|\hat{\lambda}\|_\infty \leq 1. \tag{9}$$

This theorem permits to propose the following algorithm to construct the proper vector $\tilde{\lambda}$ for the estimate (7).

STEP 1. Put in increasing order the absolute values of the current residual vector entries $e_i = z_i - H_i^T q^k$:

$$|e_{j_1}| \leq \ldots \leq |e_{j_n}| \leq |e_{j_{n+1}}| \leq \ldots \leq |e_{j_N}|.$$

Consider the first n entries in this chain, i.e., the n least in absolute value entries of the current residual vector. Denote by $\mathcal{K}$ the set of these entries indices: $\mathcal{K} = \{j_1, \ldots, j_n\}$.

STEP 2. Define $N - n$ entries of the vector $\tilde{\lambda}$ that are related to 'not least' residuals by the following relations:

$$\tilde{\lambda}_i = \mathrm{sign}(z_i - H_i^T q^k) \quad \text{for } i \notin \mathcal{K}. \tag{10}$$

STEP 3. Find the other entries of $\tilde{\lambda}_j$ for $j \in \mathcal{K}$ from the system of linear algebraic equations:

$$\sum_{j \in \mathcal{K}} H_j \tilde{\lambda}_j = -\sum_{i \notin \mathcal{K}} H_i \tilde{\lambda}_i, \tag{11}$$

where the right-hand side and the columns H_j are known. If the system matrix $(H_{j_1}, \ldots, H_{j_n})$ is nonsingular, then the vector $\tilde{\lambda}$ can be uniquely determined. Thus the equality $H\tilde{\lambda} = 0$ holds by construction.

If the current vector q^k is sufficiently close to the optimal one, then it is quite reasonable to expect that the signs of the entries of $e_{j_{n+1}}, \ldots, e_{j_N}$ coincide with the signs of the corresponding entries of the optimal residual vector. Then it follows from Theorem 4 that $\tilde{\lambda}$ is the solution of (6). Thus, for sufficiently 'late' iterations, the second estimate for the nonoptimality level allows us to find the exact value $I_0 = I^0$. Note that the second estimate is applicable to any iterative algorithm as well since it exploits the structures of the primal and dual problems (2) and (6) only.

5 Examples

The approach described above was tested at various estimation problems. For example, we investigated a model problem of the Doppler data processing in the satellite navigation system (GPS) for determining the object velocity. A large array of observations ($N = 9886$) was considered in a certain time interval. The unknown parameter was three-dimensional ($n = 3$). Both LAD method and the classical least squares method were applied. For the numerical implementation of the LAD method, the Weiszfeld algorithm and the interior-point method (the latter is incorporated in the software package MathLab) were applied. With the number of measurements $N \sim 10000$ the interior-point method requires a significant storage capacity. In this regard, the Weiszfeld method was found to be more efficient. In addition, the construction of the nonoptimality level at each iteration (not only at 'late') was useful.

Both levels $\Delta_0^{(1)}$ and $\Delta_0^{(2)}$ were calculated. The stopping criterion was chosen as: min $\{\Delta_0^{(1)}, \Delta_0^{(2)}\} \leq 1 + 10^{-7}$. Figure 1 below shows the evolution of the levels.

At the last iteration the nonoptimality levels took on the values

$$\Delta_0^{(1)} = 1 + 2.84\ 10^{-3} \quad \text{(solid curve)}$$

$$\Delta_0^{(2)} = 1 + 7.93\ 10^{-9} \quad \text{(dash curve)}.$$

Thus the guaranteed high accuracy was achieved: the cost value turned out to be very close to the optimal value. The second estimate at 'last' iterations was much more accurate than the first one. The employment of the second estimate permitted to stop the calculations earlier, thus to avoid deliberately unnecessary iterations. However, in practice, the first version of the estimate is useful as well: $\Delta_0^{(1)}$ also attains the desired accuracy but in a larger number of steps (one and half or twice larger). In this case, the nonoptimality estimate requires considerably less computation.

When solving the high dimension problems the number of required iterations M can be large (for example, in the previous example $M = 153$); in the problems of lesser dimension the desired level of accuracy attains faster. To illustrate the latter case we considered the problem of the object velocity determination with the help of phase GPS measurements at a given instant. The observation number $N = 9$; $n = 3$ as in the previous example. In this case, the extremely high accuracy was achieved in 24 iterations: $\Delta_0^{(1)} = 1 + 1.43\ 10^{-7}$ and $\Delta_0^{(2)} = 1 + 4.21\ 10^{-11}$. This example shows that the first version for the estimate can be quite accurate as well.

In both examples the testing was performed for the data with anomalous errors and without outliers as well. It was found out that without outliers the LAD method and the classical least squares method yield the

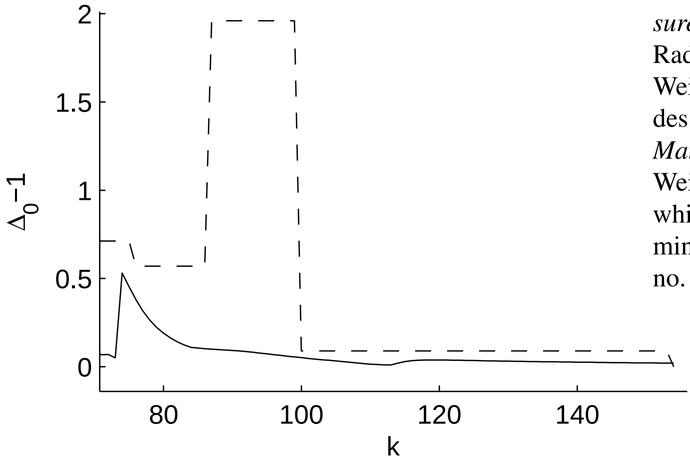

Figure 1. The evolution of the nonoptimality levels.

same accuracy: the divergence in the estimates of q did not exceed 0.01 percent. However, with the presence of anomalous errors, the least squares method strongly distorts the correct result, it did not give us even an idea about the quantities of the object coordinates. In contrast, the estimates for q obtained by the LAD method vary less than by 0.05 percent.

6 Conclusion

In the paper, some improvements of the existing solving algorithms for the LAD method were proposed. Two versions of the upper bound for the current iteration nonoptimality level were obtained. The corresponding derivation is based on the duality theory for convex variational problems. The proposed approach is quite universal and is applicable for various classes of practical problems. The numerical features of the approach were investigated at various data processing problems. The results demonstrated that the proposed nonoptimality levels realize the guaranteed control for the calculation accuracy. Thus these levels yield a useful tool for the analysis of the estimation problems with anomalous measurement errors.

Acknowledgements

The work was supported by the RFBR Grant no. 08-08-00904.

References

Boyd S., Vandenberghe L. (2004). *Convex Optimization.* Cambridge University Press.

Ekeland I., Temam R. (1976). *Convex Analysis and Variational Problems.* North-Holland Publishing Company; American Elsevier Publishing Company .

Matasov A.I. (1998). *Estimators for Uncertain Dynamic Systems.* Kluwer Academic Publishers.

Mudrov V. I., Kushko V. L. (1983). *Methods for Measurement Processing: Quasi-Likelyhood Estimates.* Radio and Communication (in Russian).

Weiszfeld E. (1937). Sur le point par lequel la somme des distances de n points donnés est minimum. *Tohoku Mathematics Journal,* vol. 43, pp. 355-386.

Weiszfeld E., Plastria F. (2008). On the point for which the sum of the distances to n given points is minimum. *Annals of Operations Research,* vol. 167, no. 1, pp. 7-41.