REMARKS ON SIMPLIFYING HYPOTHESES IN LEAST SQUARE ADJUSTMENTS

V. Casella

DIET – Università degli Studi di Pavia, Pavia, Italy
email: casella@unipv.it

ISPRS Commission VI, Working Group 3

KEYWORDS: Algorithm, Least squares, Coordinate transformations

ABSTRACT

When parameters of a coordinate transformation are estimated by means of some double points, the simplifying hypothesis of considering the randomness of only one kind of coordinates is often used. This choice produces an error that is negligible in certain cases, but not always. Because of this problem is not always well treated in textbooks, a detailed exposition is given, first of all; then it is described a simulation that has been conducted to point out when the simplified method is admissible.

1. INTRODUCTION

Let's consider for instance a parameter estimation problem for a conformal coordinate transformation

$$\begin{pmatrix} u \\ v \end{pmatrix} = \lambda \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} T_1 \\ T_2 \end{pmatrix} \qquad (1)$$

by means of some double points. This could be, but not necessarily, a typical interior orientation problem where (UV) are the instrumental coordinates of the fiducial marks, measured on each image, and (XY) are the image coordinates of the same points *read* on the calibration certificate, but *measured* once and for all by the camera manufacturer. We could also consider, equivalently, another well known planar coordinate transformations such as the similarity transformation

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$$

the special affine transformation

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$$

and the affine transformation

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$$

In a rigorous LS approach both (XY) and (UV) should be considered random variables (rv from now): under this perspective the relationship (1) is only strictly true for coordinate mean values. Nevertheless, it often happens that (XY) are treated as constants; this is not too far from truth if (XY) are a little widespread with respect to (UV); this is wrong instead if their spread is comparable or even bigger than the (UV) one. As a consequence, we could expect that in the former case a small, negligible error is introduced in the results, while, in the latter case, results are afflicted by a consistent error.

In practice the approximated approach is applied almost always. This paper's aim is to clearly explain how the parameters of a conformal coordinate transformation are estimated by LS, mostly detailing on the deterministic model structure, since this is difficult to find even in textbooks. Besides, it will be described an empirical investigation, based on simulations, that has been conducted to outline the admissibility limits of the approximated way of solving adjustments.

2. THEORETICAL BACKGROUND

The paper will take into account only the parametric model of LS. We'll consider an m-dimensional rv $\mathbf{Y}$ and an n-dimensional rv $\mathbf{X}$ ($m \geq n$) whose mean values are related in the following way

$$\overline{\mathbf{Y}} = \mathbf{A}\overline{\mathbf{X}} + \mathbf{a} \qquad (2)$$

where $\mathbf{A}$ is m x n matrix and $\mathbf{a}$ is an m vector; they must be constant, that is, **they can't have a stochastic nature**. The goal is to give an estimation of the mean values of $\mathbf{X}$ and $\mathbf{Y}$ (that will be called $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$), provided an extraction $\mathbf{Y}_0$ (a measurement) of the $\mathbf{Y}$ rv is known, so that

- the estimations $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$ satisfy the (2) condition:

$$\hat{\mathbf{Y}} = \mathbf{A}\hat{\mathbf{X}} + \mathbf{a}$$

- the observable mean value estimation is as close as possible to the observations $\mathbf{Y}_0$, in the LS sense.

In practice we will try to satisfy, though not exactly, the observation equations

$$\mathbf{Y}_0 = \mathbf{A}\hat{\mathbf{X}} + \mathbf{a} \tag{3}$$

in the best possible way, and this means, in the LS approach, to choose $\hat{\mathbf{X}}$ as the solution of the following minimum problem

$$\min_{\mathbf{X}} (\mathbf{Y}_0 - \mathbf{A}\mathbf{X} - \mathbf{a})' \mathbf{Q}^{-1} (\mathbf{Y}_0 - \mathbf{A}\mathbf{X} - \mathbf{a}) \tag{4}$$

where $\mathbf{Q}$ is the structure of the variance-covariance matrix of the $\mathbf{Y}$ rv. The word *structure* recalls that, writing the variance-covariance matrix in the form $\mathbf{C}_{YY} = \sigma_0^2 \mathbf{Q}$, the matrix $\mathbf{Q}$ has to be known beforehand, while the coefficient σ_0 will be estimated by LS solution. Under all these conditions, the solution is well known

$$\hat{\mathbf{X}} = (\mathbf{A}'\mathbf{Q}\mathbf{A})^{-1} \mathbf{A}'\mathbf{Q}^{-1} (\mathbf{Y}_0 - \mathbf{a}) \tag{5}$$

Moreover, it is also possible to calculate the estimated mean value of $\mathbf{Y}$, the observations rv

$$\hat{\mathbf{Y}} = \mathbf{A}\mathbf{X} + \mathbf{a}$$

the overall quality parameter

$$\sigma_0 = \sqrt{\frac{(\hat{\mathbf{Y}} - \mathbf{Y}_0)' \mathbf{Q}^{-1} (\hat{\mathbf{Y}} - \mathbf{Y}_0)}{m - n}}$$

and the variance-covariance matrixes for both estimated parameters and observations, respectively

$$\mathbf{C}_{XX} = \sigma_0^2 (\mathbf{A}'\mathbf{Q}\mathbf{A})^{-1}$$
$$\mathbf{C}_{YY} = \mathbf{A}\mathbf{C}_{XX}\mathbf{A}'$$

If the relationship between the mean values of the measured quantities and the unknowns is not linear

$$\overline{\mathbf{Y}} = \mathrm{f}(\overline{\mathbf{X}}) \tag{6}$$

the solution is reached iteratively by a sequence of linear approximated problems; indeed if $\mathbf{X}^*$ is an approximation of $\overline{\mathbf{X}}$, the (6) relation can be Taylor-linearized

$$\overline{\mathbf{Y}} \overset{1^\circ}{=} \mathbf{J}(\mathrm{f})_{\mathbf{X}^*} (\overline{\mathbf{X}} - \mathbf{X}^*) + \mathrm{f}(\mathbf{X}^*)$$

So the approximated observation equations, analogous to the (3), have the form

$$\mathbf{Y}_0 = \mathbf{J}(\mathrm{f})_{\mathbf{X}^*} (\hat{\mathbf{X}} - \mathbf{X}^*) + \mathrm{f}(\mathbf{X}^*)$$

The Jacobian matrix $\mathbf{J}(\mathrm{f})_{\mathbf{X}^*}$ is obtained differentiating all the components of the f function with respect to all the components of the unknowns vector $\overline{\mathbf{X}}$. The Jacobian matrix $\mathbf{J}(\mathrm{f})_{\mathbf{X}^*}$ plays the role of the matrix $\mathbf{A}$ and, being evaluated on a fixed point, it is a constant, from a stochastic point a view. The quantity $\mathrm{f}(\mathbf{X}^*)$ plays the role of the $\mathbf{a}$ vector and, again, it is a constant, being the value of a function on a fixed point.

After the first solution has been calculated, it is used as the approximated one for the next iteration and so on. The procedure is stopped when the solution is stable, that is, when further iterations give no meaningful gain. It is important to underline that, while solution (5) is rigorously valid (every LS problem of the type (2) has a unique solution of that kind), the iterative solution for not-linear problems (6) is not guaranteed: depending on the form of the f function and also on the starting solution $\mathbf{X}^*$, it could also happen that the iterations give wrong results or no results.

3. RIGOROUS AND APPROXIMATED APPROACH FOR THE CONFORMAL TRANSFORMATION

Let's start considering the **rigorous approach:** we'll think to have N double points. For each of them, there are four observables (u,v,x,y) so that the LS adjustment has 4N observations. The unknown set is made of the parameters $T_1, T_2, \alpha, \lambda$ together with the x, y coordinates of each point, so there are 4+2N unknowns

$$\mathbf{X} = (T_1, T_2, \alpha, \lambda, x_1, y_1, \ldots, x_N, y_N)^t$$

The deterministic model function f will have 4N components: in particular those related with the *i*-th point have the form

$$\overline{\mathbf{Y}} = \begin{pmatrix} \vdots \\ \overline{u}_i \\ \overline{v}_i \\ \overline{x}_i \\ \overline{y}_i \\ \vdots \end{pmatrix} = f(\overline{\mathbf{X}}) = \begin{pmatrix} \vdots \\ \overline{\lambda}(\overline{x}_i \cos\overline{\alpha} - \overline{y}_i \sin\overline{\alpha}) + \overline{T}_1 \\ \overline{\lambda}(\overline{x}_i \sin\overline{\alpha} + \overline{y}_i \cos\overline{\alpha}) + \overline{T}_2 \\ \overline{x}_i \\ \overline{y}_i \\ \vdots \end{pmatrix} \quad (7)$$

The third and forth relations only apparently are strange: they simply mean that, among the quantities that are to be estimated, some are directly measured; in any case these quantities are estimated not only on the base of their direct measurements, because they are also involved in the other two equations, so also u and v measurements contribute to x and y estimation. It could be useful and explanatory to consider also the observation equations, analogous to the (3)

$$\begin{pmatrix} \vdots \\ u_{0i} \\ v_{0i} \\ x_{0i} \\ y_{0i} \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ \hat{\lambda}(\hat{x}_i \cos\hat{\alpha} - \hat{y}_i \sin\hat{\alpha}) + \hat{T}_1 \\ \hat{\lambda}(\hat{x}_i \sin\hat{\alpha} + \hat{y}_i \cos\hat{\alpha}) + \hat{T}_2 \\ \hat{x}_i \\ \hat{y}_i \\ \vdots \end{pmatrix}$$

so that it is clear that the $\hat{x}_i$ and $\hat{y}_i$ values we are looking for are those that fit in the best way with both x_0, y_0 and u_0, v_0 measurements.

The f function considered in (7) is not linear, so it is necessary to follow the approximated iterative way; therefore it is necessary to evaluate the Jacobian of f evaluated on an arbitrary fixed point $\mathbf{X}^*$. Remembering the rule $\mathbf{J}_{ij} = \frac{\partial f_i}{\partial X_j}$, (9) shows that part of the Jacobian which is related to the *i*-th point.

$\mathbf{0}(1, 2(i-1))$ indicates, as an example, the matrix made of zeros, with 1 row and $2(i-1)$ columns. The so obtained Jacobian perfectly fits the requirements for LS adjustments because there isn't any randomness; it must be noted that this approach doesn't mean to give the same weight to all the observations, because the **Q** matrix allows u,v and x,y measurements to have different weights.

Let's now consider the **approximated approach**. Let's suppose, as an example, that the double points are the fiducial marks of an aerial image and that the (XY) coordinates are given by the camera's constructor, so that they are very precise; let's suppose again that (UV) coordinates have been measured on a 300 dpi image. The (XY) coordinates are much more precise than the (UV) ones, so we could regard the former as constants, even if they aren't, strictly speaking; in this case there are only two measured quantities (u,v) for each point so that the LS adjustment has 2N observations. The unknown set is only made of the parameters $T_1, T_2, \alpha, \lambda$, so there are 4 unknowns

$$\mathbf{X} = (T_1, T_2, \alpha, \lambda)^t$$

This way of proceeding has consistent computational advantages: equation number is halved and the unknown number is fixed to four, while it increases with point number in the rigorous approach. In the approximated approach, relations (7) take the form, for the *i*-th point

$$\begin{pmatrix} \vdots \\ \overline{u}_i \\ \overline{v}_i \\ \vdots \end{pmatrix} = f(\overline{\mathbf{X}}) = \begin{pmatrix} \vdots \\ \overline{\lambda}(\overline{x}_i \cos\overline{\alpha} - \overline{y}_i \sin\overline{\alpha}) + \overline{T}_1 \\ \overline{\lambda}(\overline{x}_i \sin\overline{\alpha} + \overline{y}_i \cos\overline{\alpha}) + \overline{T}_2 \\ \vdots \end{pmatrix} \quad (8)$$

and the Jacobian of the f function is obtained differentiating only with respect to $T_1, T_2, \alpha, \lambda$, so it has the form indicated in (10).

$$\mathbf{J}=\begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \lambda^*(-x_i^*\sin\alpha^*-y_i^*\cos\alpha^*) & x_i^*\cos\alpha^*-y_i^*\sin\alpha^* & \mathbf{0}(1,2(i-1)) & \lambda^*\cos\alpha^* & -\lambda^*\sin\alpha^* & \mathbf{0}(1,N-2i) \\ 0 & 1 & \lambda^*(x_i^*\cos\alpha^*-y_i^*\sin\alpha^*) & x_i^*\sin\alpha^*+y_i^*\cos\alpha^* & \mathbf{0}(1,2(i-1)) & \lambda^*\sin\alpha^* & \lambda^*\cos\alpha^* & \mathbf{0}(1,N-2i) \\ 0 & 0 & 0 & 0 & \mathbf{0}(1,2(i-1)) & 1 & 0 & \mathbf{0}(1,N-2i) \\ 0 & 0 & 0 & 0 & \mathbf{0}(1,2(i-1)) & 0 & 1 & \mathbf{0}(1,N-2i) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \tag{9}$$

$$\mathbf{J}=\begin{pmatrix} \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \lambda^*(-x_{0i}\sin\alpha^*-y_{0i}\cos\alpha^*) & x_{0i}\cos\alpha^*-y_{0i}\sin\alpha^* \\ 0 & 1 & \lambda^*(x_{0i}\cos\alpha^*-y_{0i}\sin\alpha^*) & x_{0i}\sin\alpha^*+y_{0i}\cos\alpha^* \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \tag{10}$$

It is noticeable in (10) that starred quantities are fixed arbitrarily, that is, stochastically constant, while x_{0i}, y_{0i} are measurements (regarded as constants, of coarse), that is stochastic quantities. Therefore the matrix **J** contains a certain quantity of randomness and this violates one important validity condition of the LS theory: only if randomness is small compared to that contained in *u*, *v*, we could expect that the approximated approach gives good results (containing a small negligible error), though non strictly correct. This is a necessary condition for the simplified approach, that should be checked every time before applying it.

4. THE INTERPOLATING LINE: THEORETICAL SCHEME AND EMPIRICAL INVESTIGATIONS

The conceptual scheme of approximated and rigorous solutions for LS adjustments has been exposed; we have chosen to illustrate it with an example based on conformal coordinate transformations because this is a very well known subject. Next question is to practically investigate the size of the error introduced by the approximated approach and, mostly, to find the conditions under which this error is negligible or not.

We have decided to work out the empirical part of the investigation on the interpolating line problem rather than on the conformal coordinate transformation: the former is certainly the simplest LS problem, but this hasn,t been the reason of the choice. Actually, for planar coordinates transformations both rigorous and approximated problems are in general not-linear, so their LS solution is an approximated one, reached iteratively; on the contrary, the simplified approach of the interpolating line is linear, as it will be next shown, so it has an exact LS solution: if such a solution has an error, this can be undoubtedly ascribed to the only existing source of error: the hypothesis on which the approximation is based.

Let's suppose to have a linear phenomenon, for instance a thermal dilatation, described by the well known law

$$y = px + q$$

If we know the law of the phenomenon, that is p and q, we can predict the value assumed by y at any x. If instead we don't know p and q, but we are able to measure some couple of values (x_{0i}, y_{0i}), we can give an LS estimation of p and q. Let's examine first of all the **rigorous approach**, which treats both x_{0i} and y_{0i} as measures, that is extractions from a rv. For each point, there are two observable rv, x_i, y_i so there are 2N observables if N is the point's number. The unknown set is constituted by p, q, $x_1, \cdots, x_N$, so the unknown vector **X** has 2+N elements structured in the following way:

$$\mathbf{X} = (p, q, x_1, \ldots, x_N)'$$

The deterministic model has the form (once again we'll give explicitly only the components related to the *i*-th point)

$$\begin{pmatrix} \vdots \\ \overline{y}_i \\ \overline{x}_i \\ \vdots \end{pmatrix} = f(\overline{\mathbf{X}}) = \begin{pmatrix} \vdots \\ \overline{px}_i + \overline{q} \\ \overline{x}_i \\ \vdots \end{pmatrix}$$

while the observation equations are

$$\begin{pmatrix} \vdots \\ y_{0i} \\ x_{0i} \\ \vdots \end{pmatrix} = f(\hat{\mathbf{X}}) = \begin{pmatrix} \vdots \\ \hat{p}\hat{x}_i + \hat{q} \\ \hat{x}_i \\ \vdots \end{pmatrix}$$

The model is of coarse not linear, so it is mandatory to use the iterative way, and, in order to carry out it, it is necessary to calculate the f Jacobian matrix on an arbitrary value $\mathbf{X}^*$

$$\mathbf{J} = \begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots \\ x_i^* & 1 & \mathbf{0}(1,i-1) & p^* & \mathbf{0}(1,N-1) \\ 0 & 0 & \mathbf{0}(1,i-1) & 1 & \mathbf{0}(1,N-1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

In the **approximated way** of proceeding, we regard the measured x_{0i} values as constants. There is one only observable for each point, namely y_i, so there are N observations. The unknown set is only constituted of p and q; the unknown vector is $\mathbf{X} = (p,q)^t$.

The deterministic model has the form

$$\overline{\mathbf{Y}} = \begin{pmatrix} \vdots \\ \overline{y}_i \\ \vdots \end{pmatrix} = f(\overline{\mathbf{X}}) = \begin{pmatrix} \vdots \\ \overline{p}\overline{x}_i + \overline{q} \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots & \vdots \\ \overline{x}_{0i} & 1 \\ \vdots & \vdots \end{pmatrix} \begin{pmatrix} \overline{p} \\ \overline{q} \end{pmatrix}$$

that assumes the general form (2), giving the **A** matrix the expression

$$\mathbf{A} = \begin{pmatrix} \vdots & \vdots \\ \overline{x}_{0i} & 1 \\ \vdots & \vdots \end{pmatrix}$$

and thinking that **a** is in this case the N-dim null vector. The observation equations are

$$\mathbf{Y}_0 = \begin{pmatrix} \vdots \\ y_{0i} \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots & \vdots \\ x_{0i} & 1 \\ \vdots & \vdots \end{pmatrix} \begin{pmatrix} \hat{p} \\ \hat{q} \end{pmatrix} = \mathbf{A}\hat{\mathbf{X}}$$

In this case it is possible to apply directly LS techniques, of coarse insofar as $x_{01},\ldots,x_{0N}$ can be treated as constants.

5. THE SIMULATION

To quantify the error introduced by the approximated approach, a simulation has been worked out with the following scheme.

- True values have been chosen: $\overline{p}=1$, $\overline{q}=1000$, $\overline{x}=1000,\ldots,2000$ (10 equally spaced values).
- The true values $\overline{x}_i$ and $\overline{y}_i$ have been increased with a normal noise characterised respectively by σ_x and σ_y, to have the *measured* values x_{0i} and y_{0i}. Different values of σ_x and σ_y have been taken into account.
- LS estimation of $\overline{p}$ and $\overline{q}$ has been worked out both with the rigorous method and with the approximated one.
- The procedure has been repeated 400 times; results have been recorded; at the end of the iterations, mean values and rmse have been calculated.
- Differences between the true values and the estimated ones have been analysed.

Results referring to three different simulations are shown in the last page; they are related to the q estimations: p estimations show the same behaviour. The left part of each graph shows the behaviour of the approximated method: the square shaped markers show the estimated value of q (obtained averaging over all the repetitions) as a function of σ_x and σ_y; the right part shows the results obtained from the rigorous method: markers are in this case diamond shaped. In each graph, the black line represents the true value for the estimated parameter; the error bars corresponds to the 95% confidence interval.

In all the simulations the rigorous method shows the correct behaviour. Let's now examine the approximated method's results. In the first simulation (Figure 1) σ_x and σ_y have the same values and range from 50 to 5: in the high value interval there is clearly a bias in the (approximated) estimations, which tends to vanish when the relative error become smaller. This could explain why in our environment, in which relative errors are small in general, people sometimes use the wrong procedure without having too big consequences.

The other two simulations want to underline that, in any case, the necessary condition regards the ratio σ_X / σ_Y : if this is small, the simplified approach is acceptable. Let's consider the values $\sigma_X = \sigma_Y = 5$ for which, according to the previous results, there is no bias. Let's fix $\sigma_Y = 5$ and let σ_X range from 5 to 50, so that the ratio σ_X / σ_Y ranges from 1 to 10: the bias shows itself as soon as the ratio σ_X / σ_Y becomes greater than 1 and increases rapidly with σ_X (Figure 2).

Let's consider the opposite situation, that is $\sigma_X = \sigma_Y = 50$, which shows a big bias, according to the first simulation. Let's fix $\sigma_Y = 50$ and let σ_X range from 50 to 5: the bias decreases with σ_X and it vanishes when the ratio σ_X / σ_Y is lower than 1/2 (Figure 3).

6. CONCLUSIONS

We have tried to follow a didactic style, giving explicitly the shape of the deterministic models of the various problems taken into account. Simulations results could also be effectively used for teaching LS adjustments and to illustrate related problems. The aproximated/rigorous approach behaviour can be briefly synthesised as follows.

- The rigorous approach always gives good results.
- The simplified approach isn't always good, as the theory tells: it is valid when the ratio σ_X / σ_Y is small.
- It seems that when X and Y are equally widespread and their relative dispersion is small, the bias vanishes. This requires some more research work.
- Further analysis is necessary so to examine not only unknown estimation, but also their variance-covariance estimation.

7. REFERENCES

Benciolini, B., 1989. Modelli di Compensazione. In: Progettazione e Ottimizzazione del Rilievo Topografico e Fotogrammetrico di Controllo, CISM, Udine, Italy.

Benciolini, B., 1990. Modelli Analitici di Base della Fotogrammetria. In: Dall'Analitico al Digitale; Nuovi Sviluppi della Fotogrammetria applicata all'Ingegneria, CISM, Udine, Italy.

Kraus, K., 1994. Fotogrammetria: Teoria e Applicazioni, Vol I, trad. Sergio Dequal, Torino. Libreria Universitaria Levrotto & Bella, Torino, Italy.

Mikhail, E.M., 1976. Observations and Least Squares. IEP, New York.

Mussio, L., 1984. Il calcolo delle Trasformazioni Piane Elementari. In: Ricerche di Geodesia Topografia e Fotogrammetria, CLUP, Milano, Italy, Vol IV.

Sansò, F., 1989. Il Trattamento Statistico dei Dati. Città Studi, Milano, Italy.

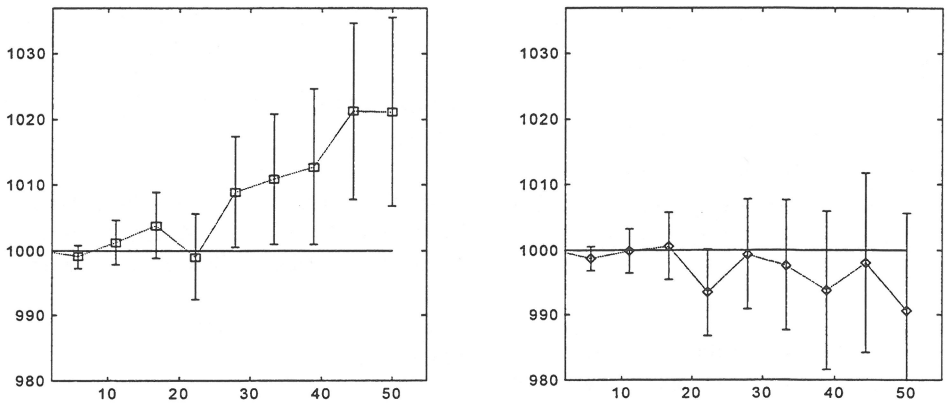

Figure 1: σ_x and σ_y take the same values, between 5 and 50

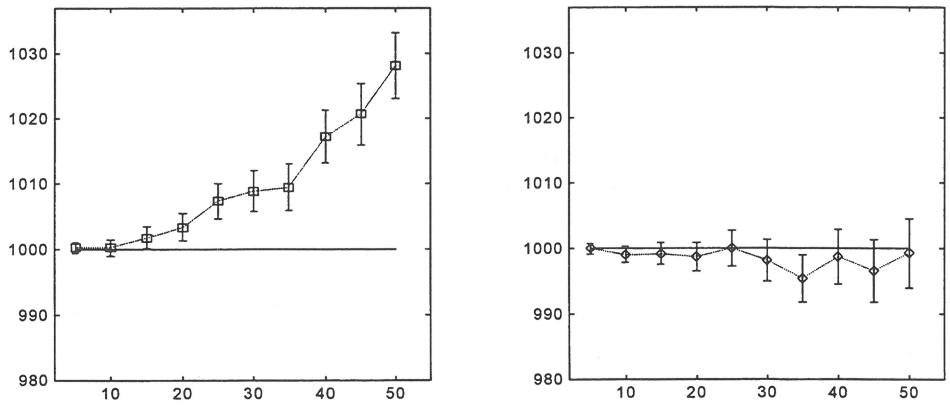

Figure 2: σ_x ranges between 5 and 50; $\sigma_y = 5$

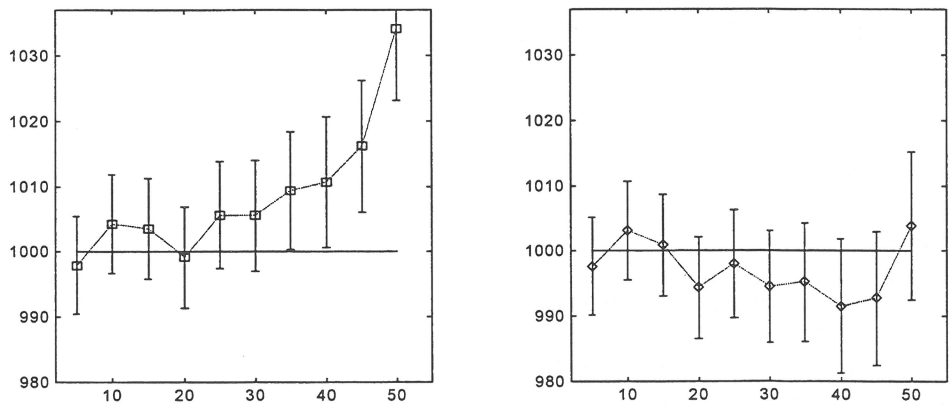

Figure 3: σ_x ranges between 5 and 50; $\sigma_y = 50$