Journal of Structural Geology, Vol 13, No 7, pp 813 to 821, 1991
Printed in Great Britain

0191-8141/91 $03 00+0 00

A robust approach to the calculation of paleostress fields from fault plane data

THOMAS M. WILL and ROGER POWELL

Department of Geology, The University of Melbourne, Parkville, Victoria, 3052, Australia

(*Received* 21 *August* 1990; *accepted in revised form* 6 *February* 1991)

Abstract—Algebraic methods combined with robust regression techniques are used to calculate paleostress tensors from field observations on faults. Previously, such calculations have involved least-squares regression; however such regression estimates are likely to break down and produce meaningless results if data are included that are inconsistent with the main body of the data. Such inconsistent data are called outliers, i.e. measurements that are discrepant with respect to the majority of the observations. In two dimensions, the trend of the main body of the data, and their outliers, can be found by plotting the data and examining them visually. Least-squares regression can then be safely applied to the data-set once the outliers have been manually removed. However, the paleostress problem possesses a four-dimensional parameter space, and, as a consequence, this approach cannot be used. To overcome this difficulty, a robust regression estimator, involving the least *median* of squares (LMS), is applied to the estimation of paleostress tensors from fault plane data; not only can the parameters of the tensor be estimated but also the quality of the data-set assessed. For a data-set that is composed of data from several stress fields the dominant reduced stress tensor will be found by the LMS estimator. A computer program, *PSALMS*, that calculates paleostress directions using this robust estimator is presented.

INTRODUCTION

THE ESTIMATION of principal stress directions from fault plane orientations, their associated slickenline orientations and the sense-of-shear on the given fault plane is commonly referred to as paleostress analysis. Since the mid-1970s several algorithms have been developed for determining–either graphically or algebraically—paleostress directions from field measurements (e.g. Carey & Brunier 1974, Angelier 1975, 1979, 1984, Angelier & Mechler 1977, Angelier & Goguel 1979, Etchecopar *et al.* 1981, Angelier *et al.* 1982, Armijo *et al.* 1982, Lisle 1987, 1988).

For a given stress field operating in the vicinity of a fault, the slip orientation on the fault is determined by the orientation of the resolved shear stress acting on this fault plane. Clearly, the orientation of the resolved shear stress (and, hence, the orientation of fault plane striae) depends on the orientation of the fault plane with respect to the stress field. The *basic assumption* common to all studies, including the present one, is that a given tectonic event is characterized by one regional homogeneous stress field (Wallace 1951, Bott 1959). The implication is that the slip direction on a fault plane is determined by a single stress deviator and that all faults which have slipped during one tectonic event moved independently but in a way consistent with this single stress deviator.

In this study, a new approach has been followed in calculating paleostress directions from field observations. The formulation of the problem is identical to the one suggested by Angelier *et al.* (1982); the statistical treatment of the data, however, is different and is, in contrast to other studies (e.g. Etchecopar *et al.* 1981, Angelier *et al.* 1982), capable of calculating paleostress directions from data which include outliers, which would otherwise have a deleterious effect on the calculations. This means that even in the presence of data belonging to different stress fields the statistical treatment employed here enables the dominant principal stress directions to be found. The corresponding computer program, *PSALMS*, and an application are briefly described in the Appendix.

FORMULATION OF THE PROBLEM

Any stress tensor, **S**, is composed of an isotropic component, the mean stress $\beta\mathbf{I}$ with **I** being the identity matrix, and a deviatoric component, the stress deviator **D**:

$$\mathbf{S} = \alpha\mathbf{D} + \beta\mathbf{I} \qquad (1)$$

(e.g. Etchecopar *et al.* 1981) in which α and β are a material constant and the lithostatic pressure, respectively, and α is always positive. Vectors and tensors will be denoted by boldface throughout. In the general case, **S** involves six parameters, the three main-diagonal elements and the off-diagonal elements which are symmetric with respect to the main diagonal:

$$\mathbf{S} = \begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{bmatrix},$$

with $S_{12} = S_{21}$, $S_{13} = S_{31}$ and $S_{23} = S_{32}$.

If **D**, α and β are known, the stress field is completely determined and the directions of the principal stresses *and* their magnitudes are specified. The magnitudes of the principal stresses, however, cannot be derived alone from a knowledge of fault plane and striation orientations and the sense-of-shear on the fault planes. Additional information in terms of the rheological properties of rocks with respect to their rupture and friction behavior and the paleo-depth (or lithostatic load) at the time of faulting are needed (e.g. Angelier 1989). This leads directly to the concept of the "reduced stress tensor", for which the constants α and β that define the magnitudes of the principal stresses, are chosen arbitrarily. The orientations of the principal stresses are not affected by this arbitrary choice because neither the addition of the mean stress (i.e. $\beta\mathbf{I}$) to the reduced stress tensor nor the multiplication of $\mathbf{D}$ by a positive scalar (i.e. α) will change the orientation and the sense of the principal stresses (e.g. Angelier *et al.* 1982). Thus, the orientations of the principal stress axes are solely determined by $\mathbf{D}$. Following Angelier *et al.* (1982), the colinearity between the resolved shear stress and the striation on that plane is expressed as:

$$\mathbf{s}\cdot\mathbf{T}\cdot\mathbf{n} = +\sqrt{(\mathbf{T}\cdot\mathbf{n})\cdot(\mathbf{T}\cdot\mathbf{n}) - (\mathbf{n}\cdot\mathbf{T}\cdot\mathbf{n})^2}, \qquad (2)$$

where $\mathbf{T}$ is the reduced stress tensor, $\mathbf{n}$ and $\mathbf{s}$ are the unit normal and the unit striation on the fault plane, and '$\cdot$' denotes a dot product. Equation (2) is identical to equation (5) of Angelier *et al.* (1982). This is the key equation used in solving for the reduced stress tensor, $\mathbf{T}$.

Remembering that the full stress tensor, $\mathbf{S}$, has six degrees of freedom an infinite number of different stress tensors, $\mathbf{S}$, can be obtained from the reduced stress tensor, $\mathbf{T}$, using different values of α and β. Following Angelier & Goguel (1979), α and β are fixed arbitrarily so that:

$$T_{11} + T_{22} + T_{33} = 0 \quad \text{and} \quad T_{11}^2 + T_{22}^2 + T_{33}^2 = \tfrac{3}{2}. \qquad (3)$$

A solution to these equations (e.g. Angelier *et al.* 1982) is:

$$T_{11} = \cos[x], \qquad T_{22} = \cos\left[x + \frac{2\pi}{3}\right] \qquad \text{and}$$

$$T_{33} = \cos\left[x + \frac{4\pi}{3}\right].$$

Clearly, x is modulo 2π. Therefore, by choosing α and β so that equation (3) holds, the stress tensor is reduced from six to four unknowns and $\mathbf{T}$ has the form:

$$\mathbf{T} = \begin{bmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{bmatrix} = \begin{bmatrix} \cos[x] & a & b \\ a & \cos\left[x + \frac{2\pi}{3}\right] & c \\ b & c & \cos\left[x + \frac{4\pi}{3}\right] \end{bmatrix}. \qquad (4)$$

Thus, the unknowns to be solved for—using the field observations—are the parameters of the reduced stress tensor, x, a, b and c.

The orientations of the fault planes, their striations and the sense-of-shear on these planes are completely described by three angles: the dip direction of the plane, d, the dip of the plane, p, and the slip angle, ι (Fig. 1). The slip angle is the pitch of the lineation with the sense-of-shear taken into account (i.e. the angle is measured clockwise from the horizontal to the 'head' of the slip vector and, therefore, ranges from 0° to 360°). The slip vector is taken to lie on the footwall block. A set of these three angles constitutes *one* field observation.

ROBUST REGRESSION AND THE INVERSE PROBLEM

Given a reduced stress tensor, $\mathbf{T}$, it is straightforward to calculate the orientation and direction of a slip vector on a given fault plane: this is the 'direct problem'. To estimate the stress tensor from a knowledge of fault and striation orientations is more difficult: this is the 'inverse problem'.

Estimating $\mathbf{T}$ involves relating the orientation and direction of the theoretical resolved shear stresses on a set of fault planes to the observed slip directions on each of those planes. According to the basic assumption, the movement on the faults is solely governed by one single regional stress deviator. Generally, equation (2) will not be exactly obeyed for each of the measured faults in a set of fault planes for any single stress tensor, $\mathbf{T}$ (e.g. Angelier *et al.* 1982). Instead, a tensor $\mathbf{T}$ is sought which minimizes the angular misfit between the observed slip directions, given by the fault plane lineations, and the slip directions calculated using the estimated $\mathbf{T}$. This is the same logic that has been applied in existing algorithms (e.g. Etchecopar *et al.* 1981, Angelier *et al.* 1982).

The function to be minimized in order to calculate a

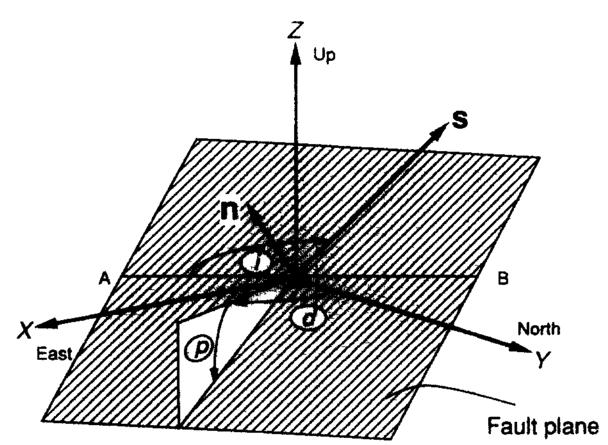

Fig. 1. Cartesian reference system used for recording field observations. **n** and **s** are the unit normal and the unit striation vectors. d is the dip direction, p the dip of the fault and ι is the slip angle measured clockwise from the horizontal AB towards the positive end of **s**. In the figure, **s** is assumed to lie on the footwall-side and since it points upwards, the hangingwall moved down with respect to the footwall shown. This corresponds to a normal fault. If the direction of **s** is changed (i.e. a reverse instead of normal fault) the slip angle ι is increased by 180°.

best paleostress tensor, **T**, varies slightly between workers (e.g. Etchecopar *et al.* 1981, Angelier *et al.* 1982) but they all have one aspect in common: they use the classical least-squares (LS) estimator in order to solve for a best-fit tensor. In the LS estimator, the *sum* of the least squares of the residuals, $\mathbf{e}_i$, are minimized:

$$\text{minimize} \sum_{i=1}^{J} \mathbf{e}_i^2 \qquad (5)$$

in which $\mathbf{e}_i = \mathbf{y}_i - \hat{\mathbf{y}}_i$ are the differences between the observed (or measured), $\mathbf{y}_i$, and the predicted, $\hat{\mathbf{y}}_i$, values. In the formulation above, the residuals are the differences between the measured values of *d*, *p* and *i* and the ones calculated using **T**.

LS techniques give optimum regressions of data (in a statistical sense) if the data to be analysed are Gaussian distributed and, more importantly, are free of outliers. An outlier is simply a measurement that is discrepant with respect to the trend of the majority of the data, for example, as a consequence of belonging to a different stress field or of erroneous measurement, especially the wrong sense-of-shear; incorrect geological interpretation, recording or keypunch errors, etc. As soon as any data-set contains outliers LS techniques may give spurious answers, and, hence, are usually inappropriate for real data (e.g. Rousseeuw & Leroy 1987). In order to process data-sets containing outliers, alternative statistical methods which can cope with the presence of outliers should be used. Statisticians call such methods 'robust'; they have been under development since the early 1960s (e.g. Huber 1964) but they have not yet been applied in paleostress analyses. In order to solve the inverse problem, a robust approach is followed here.

Robustness can be appreciated from the concept of the breakdown point, the smallest fraction of arbitrary contamination a data-set can sustain before a regression estimator will produce spurious results from a given data-set. The LS technique has an asymptotic breakdown point of 0% because *one* single outlier can influence the best-fit estimate, causing an arbitrary result. The breakdown point is increased up to 50% if the least *median* of squares (LMS) estimator is used (a proof is given in section 3.4. of Rousseeuw & Leroy 1987). The LMS estimator was first proposed by Rousseeuw (1984) and provides an optimally robust measure of the scatter of the residuals. As Rousseeuw & Leroy (1987) point out, a breakdown of 50% is the best that can be achieved because for larger amounts of contamination (i.e. more than 50%) it becomes impossible to distinguish between 'good' and 'bad' data. The LMS estimator has the following form:

$$\text{minimize median } (\mathbf{e}_i^2). \qquad (6)$$

The breakdown point governs the maximum number of outliers that can be found. In the case of the paleostress problem, $(n - 4)/2$ outliers can be detected, where *n* is the number of fault planes.

The breakdown of LS and the success of LMS can be seen with reference to Fig. 2.

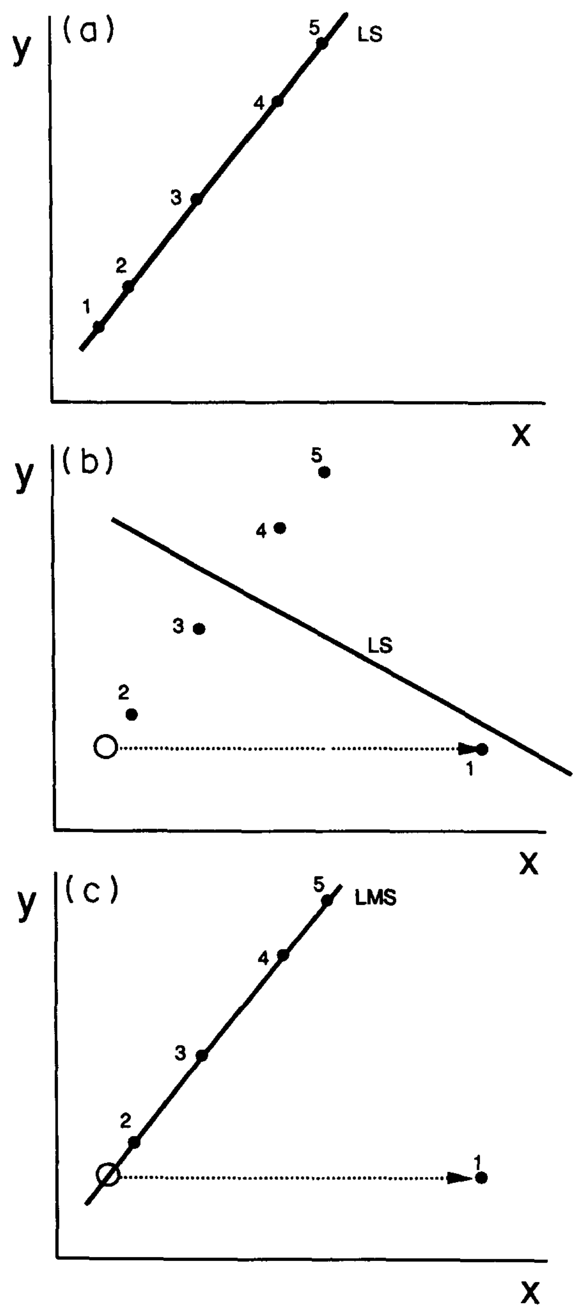

Fig. 2. (a) A data-set with five points and their LS regression line. (b) The same data as in (a) but with an outlier in the *x*-direction. The LS regression line has been drastically influenced by this one leverage point and is now almost orthogonal to its position in (a). (c) In contrast to the LS estimator in (b), the outlier does not influence the LMS best-fit line.

Figure 2(a), contains five data points with a very well-fitted LS line. If, however, only *one* data point is, for example, recorded wrongly, for instance point 1 (Fig. 2b), the LS best-fit line will swing, in this case, by almost 90° away from its original best-fit line, in order to minimize the sum of the squares of the residuals (Fig. 2b). A point like point 1 on Fig. 2(b) is called a leverage point (or an influential outlier) because it affects the LS estimator dramatically. The LS line does not now reflect the main trend of the data because the residual e_1 would have a large value and would dominate the sum in equation (5). Unfortunately, leverage points, like the one in this example, can *not* be detected by looking at LS residuals. On Fig. 2(b), point 1 has tilted the LS line so much that the point is very close to the LS line. Therefore, the residual e_1 is small whereas the residuals e_2 and e_5 which correspond to the data points furthest away from the best-fit line (in Fig. 2b) will have the largest absolute values. Thus, if one follows the logic of deleting points with the largest residuals first, in order to fit the remaining data better, the 'good' data points 2 and 5 would be deleted first. This is the problem that arises in using the regression diagnostics provided in, for example, the paleostress program of Etchecopar *et al.* (1981). In that case, regression diagnostics allow the user to identify data points which have large residuals compared to the best-fit LS line. However, the LS estimator is in *no* position to identify outliers like the ones described above. Therefore, deletion of data (which in itself provides no difficulties) might result in discarding 'good' data and the final estimate of the paleostress directions would be meaningless. More sophisticated regression diagnostics, involving for example normalization of the residuals to amplify their magnitude if the data are influential (e.g. Belsley *et al.* 1980, Powell 1985), may identify individual outliers but cannot cope with numerous outliers particularly when they are grouped. The LMS provides an appropriate alternative.

Application of the LMS estimator to the data-set containing the leverage point yields the result shown in Fig. 2(c); the main trend of the data is reflected in the LMS line, and the estimate is not influenced by the outlier. In contrast to LS, there is no analytical solution of equation (6) to provide the LMS estimate, but it can be solved approximately as a combinatorial problem which is straightforward and surprisingly simple, if laborious computationally. For each combination of two data points (of which there are $n(n-1)/2$, where n is the number of data points), the straight line defined by the two points is calculated, and the median of the squares of the residuals of the remaining points to this line is computed. The LMS line is then taken to be the line (out of the 10 possibilities in the case of Fig. 2) that has the smallest median of the squares of the residuals associated with it. The exact LMS line will be very close to this line, particularly for data-sets with more than a few data points.

The application of the LMS estimator to fault plane data has major advantages over the commonly used LS estimator because it allows data-sets including outliers to be reliably processed to give paleostress tensors Application of the LMS estimator to the inverse problem is essentially an extension of the method of Angelier *et al.* (1982).

THE SOLUTION OF THE INVERSE PROBLEM

Solving the inverse problem can be understood in terms of finding the reduced stress tensor, **T**, such that the median of the squares of the residuals is minimized, with the constraint equation, (2), obeyed for each fault plane In contrast to ordinary regression, the number of constraint equations is only a third of the number of measurements since the constraint equation for each fault plane involves three measurements: d, p and i. This precludes the use of conventional LMS as in the program PROGRESS of Rousseeuw & Leroy (1987). The underlying implementation of LMS for problems of this type involves generating general solutions of the constraint equations, in terms of parameters and 'corrected' data, and taking as the LMS solution the parameters corresponding to the 'corrected' data closest to the observations in the LMS sense. The solution provides not only the paleostress tensor, **T**, but also allows identification of the outliers in the data-set, from the observations which are not well approximated by **T**.

The constraint equation

$$\mathbf{y} = \mathbf{s}\cdot\mathbf{T}\cdot\mathbf{n} - \sqrt{(\mathbf{T}\cdot\mathbf{n})\cdot(\mathbf{T}\cdot\mathbf{n}) - (\mathbf{n}\cdot\mathbf{T}\cdot\mathbf{n})^2} = 0$$

is non-linear, solution of a set of such equations is facilitated by linearization. The linearization of equation (2) is done by means of a Taylor expansion, requiring the formulation of the partial derivatives of **y** with respect to each field observation d, p and i (contained in the vectors **n** and **s**), and the four parameters, x, a, b and c (contained in the matrix **T**). The matrix of these partial derivatives is called the Jacobian, **J**; it has n rows and $3n+4$ columns, where n is the number of faults measured The Taylor expansion of equation (2) using only zero and first-order terms is:

$$\mathbf{y} \approx \mathbf{y}_0 + \mathbf{J}\cdot\Delta\mathbf{x} = 0$$

giving the linearized form:

$$\mathbf{J}\cdot\Delta\mathbf{x} = -\mathbf{y}_0 \qquad (7)$$

in which **J** is evaluated at $\mathbf{x}_0$, $\mathbf{y}_0$ is **y** evaluated at $\mathbf{x}_0$, and $\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}_0$. In these, $\mathbf{x}_0$ contains the values around which the Taylor expansion is undertaken; it is of length $3n+4$, where the first $3n$ elements are the observed values of d, p and i, and the last four values are the starting guesses for the parameters, x, a, b and c. In these, **x** contains the values of the observations and the parameters sought.

Solving for **x**, the matrix equation (7) is an *under-determined* set of linear equations involving n equations in $3n+4$ unknowns: there are generally infinitely many solutions to such sets. The solution required is the one that minimizes the median of the squares of the residuals; equation (6). The solution to such an under-determined set of equations, equation (7), is the sum of the characteristic (or particular) solution, $\mathbf{x}_c$, and the solution of the homogeneous system $\mathbf{J}\cdot\mathbf{x}_h = 0$, giving $\mathbf{x}_h = \mathrm{ns}(\mathbf{J})\cdot\mathbf{p}$, where ns(**J**) is the nullspace of **J**, and **p** is the free vector, reflecting the under-determined nature of the set of equations (e.g. Strang 1988). Equation (7) becomes $\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}_0 = \mathbf{c}_c + \mathbf{x}_h$ giving:

$$\mathbf{x} = \mathbf{x}_0 + (\mathbf{x}_c + \mathrm{ns}(\mathbf{J})\cdot\mathbf{p}). \qquad (8)$$

As **J** is of dimension n by $3n+4$, ns(**J**) is of dimension

$2n+4$ by $3n+4$, and, therefore, **p** is of length $2n+4$, corresponding to the degree that equation (7) is underdetermined. For a particular data-set and starting guess, **x** depends only on the free vector, **p**, as $\mathbf{x}_0$, $\mathbf{x}_c$ and ns(**J**) are known. For a particular **p**, the median of the squares of the residuals on the observations can be calculated from the first $3n$ elements of $\mathbf{e} = \mathbf{x}_c + \mathbf{x}_h$, while **T** can be calculated from the last four elements of **x**. To find the *least* median of squares estimate of **T**, it is necessary to find the **p** which gives the smallest median of squares of the residuals. This is done analogously to the two-dimensional case (Fig. 2) in which pairs of data were used to define a line with respect to which the median of the residuals of the remaining data are calculated. For the more general problem, subsets of data are used to solve equation (8) for **p**, thus defining a trend with respect to which the median of the residuals of the remaining data are calculated.

To see how each subset is processed, note that equation (8) can be rearranged to give:

$$\mathrm{ns}(\mathbf{J})\cdot\mathbf{p} = \mathbf{x} - \mathbf{x}_0 - \mathbf{x}_c \qquad (9)$$

which, if **x** is known and **p** is to be solved for, is an *over*-determined system of equations. Since **p** is of length $2n+4$, only $2n+4$ rows of ns(**J**) and the corresponding elements of the right-hand side of (9) are required to solve for **p**. Denoting this subset by ′, if **x** is set equal to $\mathbf{x}_0$, then (9) becomes:

$$\mathrm{ns}(\mathbf{J})'\cdot\mathbf{p} = -\mathbf{x}'_c$$

which can be easily solved for **p**. ns(**J**)′ is the 'reduced' nullspace of **J**. Given **p**, back-substitution into equation (8) gives **x**, and the median of squares of the non-zero residuals can be calculated.

Many subsets need to be processed in order to be certain that a subset is chosen that contains no outliers. Ideally, all possible subsets, taking $2n+4$ from $3n$ measurements, should be examined, but, since there are

$$\binom{3n}{2n+4}$$

ways to choose $2n+4$ from $3n$ items, the number of possible combinations increases rapidly with increasing n. For instance, if $n = 20$ (i.e. 20 fault planes have been measured in the field) the number of possible combinations is already $\approx 1.5 \times 10^{14}$. As a consequence, the procedure described above will normally be executed only for a limited number of times, k, with the subsets chosen at random from among the $3n$ measurements. For instance, in the example from Crete (see Appendix), 100 subsets (i.e. $k = 100$) were examined for each of the eight starting guesses (see Appendix).

Once the vector **p** giving the LMS is found, the reduced stress tensor, **T**, can be calculated from the parameters x, a, b and c in the last four elements of **x**. From **T**, the orientations of the principal stresses can be determined using an eigenvalue decomposition of **T**. These LMS estimates, in contrast to LS ones, will be independent of the presence of outliers in the data. An application of the LMS approach to paleostress analysis using the *PSALMS* code is given in the Appendix.

DISCUSSION

The LMS approach not only allows reliable estimation of the paleostress tensor consistent with the majority of fault planes in a data-set, but also the identification of any outliers in this set. Noting that the first $3n$ elements in **x** are the calculated, 'corrected' d, p and i values which correspond to the LMS reduced stress tensor, **T**, outliers can be identified as those values that lie far away from the observations. The decision whether a residual is large, and hence identifies an outlier, or whether a residual is small and characterizes a 'good' datum is not obvious. In order to make this decision, the residuals on the observations are compared with a scale estimate, σ_{fit}, calculated from the data, which, obviously, has to be robust itself, only depending on the 'good' data and not being affected by any outliers. A simple scale estimate is the minimal median itself; if it is multiplied by a sample correction factor depending on the size of the data-set, it approximates the classical scale estimate of LS for Gaussian distributed residuals Rousseeuw & Leroy 1987). The robust scale estimate used is:

$$\sigma_{\mathrm{fit}} = 1.4826\left(1 + \frac{5}{n-m}\right)\sqrt{\mathrm{median}\ (e_i)^2},$$

where n is the number of observations and m is the number of the parameters (Rousseeuw & Leroy 1987). In terms of the paleostress problem, $m = 4$ and median $(e_i)^2$ is equal to the minimal least median of squares obtained from a given data-set. The scale estimate is used to find the outliers in the data-set; data are classified as outliers if they fail the test:

$$|e_i| < 2.5\,\sigma_{\mathrm{fit}}.$$

As pointed out by Rousseeuw & Leroy (1987), the bound is arbitrary but reasonable, because, for Gaussian distributed residuals, only 0.5% of the residuals would be expected to be larger in magnitude than $2.5\,\sigma_{\mathrm{fit}}$.

Although it is an important advance to be able to undertake paleostress analysis of data-sets with outliers, it is also important to be able to obtain a measure of the range of stress tensors consistent with the data. Such measures are easy to obtain, in the form of a covariance matrix, from LS analysis, but more difficult from LMS analysis. The best way to proceed is (Rousseeuw & Leroy 1987):

(1) apply LMS to a data-set to find $\mathbf{T}_{\mathrm{LMS}}$;

(2) identify outliers via $\mathbf{T}_{\mathrm{LMS}}$ and $e_i < 2.5\,\sigma_{\mathrm{fit}}$, and remove these data;

(3) apply LS on the resulting outlier-trimmed data-set; such analysis is referred to as re-weighted least squares (RLS). This analysis will give $\mathbf{T}_{\mathrm{RLS}}$ and the covariance matrix of the elements of $\mathbf{T}_{\mathrm{RLS}}$; it can be undertaken using any of the existing LS computer programs.

An indication of the range of **T** consistent with a data-set can be obtained by means of a modified χ^2-test (Wonnacott & Wonnacott 1981) applied to σ_{fit} calculated from the median of squares of the residuals for each of the subsets processed. This allows the identification of paleostress tensors which are within a 95% confidence interval of the LMS paleostress tensor; these tensors are compatible with the data and yield statistically satisfactory solutions to the inverse problem. The disadvantage of this approach is that a large number of subsets need to be processed in order to properly define a 95% confidence interval.

It is interesting to note that the residuals, and therefore $\sqrt{\text{median}\,(e_i)^2}$, are in radians. Given that measurements of the angles, d, p and i, are unlikely to be much better than ±4°, or ±0.07 radians, a value of the least median of squares which is an approximate lower achievable limit would be $0.07^2 = 0.005$; at this limit, at least half of the measured angles are fitted to better than ±4°. The corresponding value for ±10° is 0.03. Given this natural scale for the value of the least median of squares, approximate ranges for acceptable values of the orientations of the paleostress tensor could be obtained once an acceptable maximum ± on the measured angles was decided upon.

Acknowledgements—T M Will acknowledges financial support from a Melbourne University Postgraduate Scholarship and the Studienstiftung des Deutschen Volkes Chris Wilson is thanked for comments on an earlier version of this manuscript and the "creation of the book of psalms" D J Sanderson and an unknown reviewer are thanked for their helpful comments

REFERENCES

Angelier, J 1975. Sur l'analyse de mesures recueillies dans des sites faillés: l'utilité d'une confrontation entre les méthodes dynamiques et cinématiques *C r hebd Séanc Acad. Sci , Paris* **D281**, 1805–1808.

Angelier, J. 1979. Determination of the mean principal stresses for a given fault population. *Tectonophysics* **56**, T17–T26

Angelier, J 1984. Tectonic analysis of fault slip data sets. *J geophys Res.* **89**, 5835–5848.

Angelier, J 1989 From orientations to magnitudes in paleostress determinations using fault slip data. *J. Struct Geol* **11**, 37–50

Angelier, J. & Goguel, J 1979 Sur une méthode simple de détermination des axes principaux des constraintes pour une population de failles. *C r. hebd Séanc. Acad. Sci., Paris* **D288**, 307–310

Angelier, J. & Mechler, P 1977. Sur une méthode graphique de recherche des contraintes principales également utilisable en tectonique et en séismologie: la méthode des dièdres droits *Bull. Soc Géol Fr.* **19**, 1309–1318

Angelier, J , Tarantola, A., Valette, B & Manoussis, S 1982 Inversion of field data in fault tectonics to obtain the regional stress—1 Single phase fault populations: a new method of computing the stress tensor. *Geophys. J R. astr. Soc.* **69**, 607–621.

Armijo, R., Carey, E & Cisternas, A. 1982. The inverse problem in microtectonics and the separation of tectonic phases *Tectonophysics* **82**, 145–160

Belsley, D. A , Kuh, E & Welsch, R. E. 1980 *Regression Diagnostics*. Wiley, New York.

Bott, M. H. P. 1959. The mechanics of oblique slip faulting. *Geol Mag.* **96**, 109–117.

Carey, E. & Brunier, B. 1974. Analyse théoretique et numérique d'un modèle méchanique élémentaire appliqué à l'étude d'une population de failles *C. r hebd Séanc Acad. Sci., Paris* **D279**, 891–894.

Etchecopar, A , Vasseur, G. & Daignières, M 1981 An inverse problem in microtectonics for the determination of stress tensors from fault striation analysis *J Struct Geol* **3**, 51–65

Huber, P J 1964 Robust estimation of a local parameter *Ann Math Stat* **35**, 73–101

Lisle, R J 1987 Principal stress orientations from faults an additional constraint *Annls Tectonicae* **1**, 155–158

Lisle, R J 1988 Romsa A Basic program for paleostress analysis using fault-striation data *Comput & Geosci* **14**, 255–259

Powell, R. 1985. Regression diagnostics and robust regression in geothermometer/geobarometer calibration the garnet-clinopyroxene geothermometer revisited *J metamorph Geol* **3**, 231–243.

Rousseeuw, P J 1984 Least median of squares regression *J Am Stat Ass.* **59**, 871–880

Rousseeuw, P J. & Leroy, A M 1987 *Robust Regression and Outlier Detection*. Wiley, New York

Strang, G 1988. *Linear Algebra and its Application* (3rd edn) Harcourt Brace Jovanovich, San Diego

Wallace, R E. 1951 Geometry of shearing stress and relation to faulting. *J. Geol.* **59**, 118–130

Will, T M 1990 Theoretical studies in metamorphic petrology and structural geology Unpublished Ph D thesis, University of Melbourne, Australia

Wilson, C J L , Will, T M., Cayley, R. A & Chen, S. In press Geologic framework, tectonic evolution and displacement history in the Stawell Zone, Lachlan Fold Belt, Australia. *Tectonophysics*

Wolfram, S 1988 *Mathematica™. A System for Doing Mathematics by Computer* Addison-Wesley, Redwood City

Wonnacott, T H & Wonnacott, R J. 1981 *Regression A Second Course in Statistics* Wiley, New York

APPENDIX

THE PALEOSTRESS PROGRAM, *PSALMS* v1.0.

A computer program, *PSALMS* ("*P*aleo*S*tress *A*nalysis by the *L*east *M*edian of *S*quares"), has been written to perform the calculations described above, it calculates paleostress orientations from fault plane measurements in the presence of outliers. It consists of three main parts (Fig. A1):

(1) input and internal recalculation of the input variables,

(2) calculation of a starting guess for part 3 through a preliminary estimate of the paleostress orientations using Angelier & Mechler's (1977) graphical, right-dihedra method in conjunction with an additional constraint proposed by Lisle (1987);

(3) calculation of the paleostress directions using the least median of squares (LMS) and the constraint equation (2).

Both the computer program and the theoretical background are covered in great detail by Will (1990). After a preliminary estimate of the paleostress directions (part 2), the core of the program (part 3) calculates the paleostress orientations subject to the constraint equation using LMS. The algorithm involves linearization of the constraint equations, followed by solution of the resulting under-determined set of linear equations from a calculation of the nullspace of the Jacobian. From the solution, the paleostress orientations are calculated and the outliers identified.

The program has been written on the Apple Macintosh using the software package Mathematica™ v1.2 (Wolfram 1988). You need to have Mathematica™ in order to execute the paleostress program, *PSALMS* Mathematica™, and therefore the current version of *PSALMS*, needs at least 4 MB of RAM and can be run on a Macintosh SE but is much faster on a Mac SE 30 or Mac II. A copy of the mathematica package, *PSALMS* v1.0, is available, as shareware, for $ A25 from the senior author (just send an ***uninitialized*** 3.5" disk). There is *no* IBM version available

Estimation of a starting guess for the paleostress directions

In the following description the data-set of Angelier *et al.* (1982) is used as an example. All figures were produced by *PSALMS*

A starting guess for the unknown elements in the reduced stress tensor is required in order to solve the constraint equation (2); this is independent of the regression estimator employed Even though different starting guesses could be chosen from geological intuition or arbitrarily, we calculate them by the right-dihedra method of Angelier & Mechler (1977) in conjunction with Lisle's (1987) additional constraint. However, in contrast to Lisle (1988), where the user has to provide one likely σ_1 direction, we use an internal reference grid consisting of 60 (this default value can be changed interactively by the user) equally spaced reference directions, x_i, from which the most likely principal stress orientations are calculated. This provides a quick and a much more efficient way to find possible σ_1 directions if the user has not already a good *a priori* knowledge of the paleostress directions. Each individual reference direction is treated as a likely candidate for being σ_1. The frequency of how often a reference direction is contained in the σ_1 dihedra is calculated and displayed graphically (Fig. A2). Since the principal stresses are mutually orthogonal the σ_3 vector has to lie on a plane 90° away from any direction that is likely to be σ_1. On that plane an array of new reference directions, z_i, is set up. The program checks whether these reference directions are contained within the σ_3 dihedra as defined by the faults and calculates the likelihood of those directions being σ_3. The combined probabilities for an x_i–z_i pair being σ_1 and σ_3 are calculated subject to Lisle's (1987) additional constraint. This yields a preliminary estimate of the paleostress directions (Fig. A3a). This estimate is used in the calculation of a starting guess for the unknowns of the reduced stress tensor, **T**.

Algebraically, the orientations of the principal stresses σ_1, σ_2 and σ_3 correspond to the three eigenvectors of **T**, with the direction cosines of the principal stresses identical to the normalized eigenvectors. In turn, each eigenvector is associated with an eigenvalue which is proportional to the point density in the direction of the principal stress. From the eigenvectors and their eigenvalues, the reduced stress tensor is obtained from:

$$\mathbf{T} = \mathbf{V}\Lambda\mathbf{V}^{-1}, \qquad \text{(A1)}$$

where **V** is an orthogonal matrix that contains the eigenvectors in its

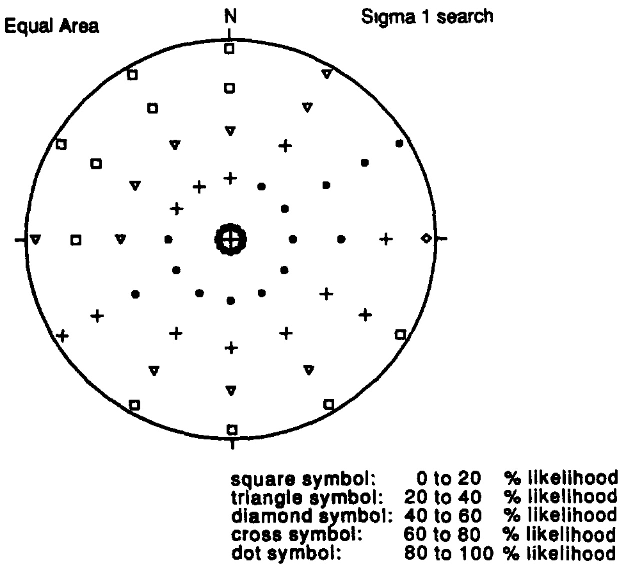

Fig. A2. Likelihood of a reference direction being contained in the σ_1-dihedra. Original *PSALMS* output. Equal-area, lower-hemisphere projection.

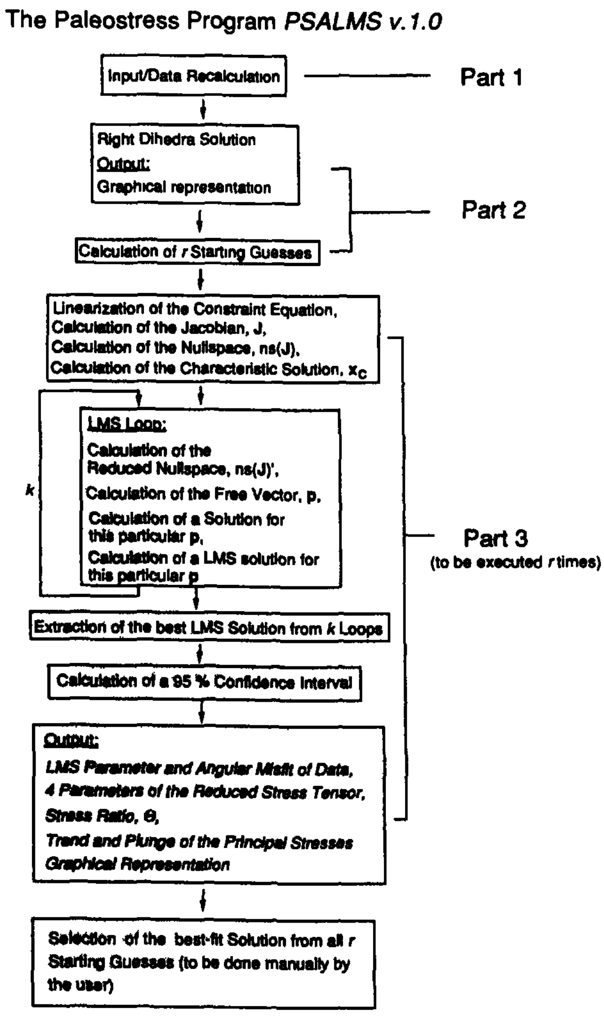

Fig. A1. Flowchart of the computer program *PSALMS*. The LMS loop should be executed for *at least* 100 times (i.e. $k = 100$; see text). It is recommended that part 3 is executed for r times with r being the number of different starting guesses. By default, $r = 8$, this value, however, can be changed by the user.

columns and Λ is the diagonal matrix of the eigenvalues. Since **T** is symmetric we obtain six constraint equations. However, there are seven unknowns, the three eigenvalues and the four parameters of **T**, x, a, b and c; therefore, the problem is unsolvable except for the trivial (or zero) solution. Therefore, in order to obtain a starting guess for **T**, the parameter x is fixed arbitrarily at a series of values in $\pi/4$ steps in the range 0 to 2π. The remaining parameters a, b and c are calculated from equations (3) and (A1) for the eight different default values of x. This produces eight different starting guesses for **T**. Iterating from each of the eight starting guesses should arrive at the same LMS estimate for **T**, although it is possible that iteration might find a local minimum in the function, equation (7). However, our experience is that just one iteration is generally required from the arbitrarily fixed x value closest to the LMS solution.

The LMS solution

The LMS solution to the inverse problem, using the different starting guesses for **T**, is found by minimizing the median of the squares of the residuals subject to the constraint equation (2). As noted in the text, this is a combinatorial problem, analogous to the two-dimensional case shown in Fig. 2. *PSALMS* calculates the Jacobian of the constraint equation (2), its nullspace, and the vector of the free variables for many subsets of the data, to find the stress tensor which minimizes the median of the residuals. Additionally, a modified χ^2-test is applied and all the paleostress orientations that are compatible with the data at a 95% confidence level are identified. Use of *PSALMS* is now illustrated with the help of an example.

PSALMS is applied to a set of 33 normal faults from Crete given by Angelier *et al.* (1982). Results from previous studies and from our algorithm using Lisle's (1987) method are shown in Fig. A3(a). For each starting guess, a minimal least median of squares is found from 100 loops (i.e. $k = 100$) executed by the program. For each loop *PSALMS* calculates the outliers, the best LMS value, the four parameters of the reduced stress tensor, the stress ratio ($\Theta = (\sigma_2 - \sigma_3)/(\sigma_1 - \sigma_3)$; Angelier 1975) and the trend and plunge of σ_1, σ_2 and σ_3. For this data-set, the results produced are summarized in Table A1 and graphically displayed in Fig. A3(b). The LMS result that best fits the majority of the data corresponds to the minimum of the curve defined by the points in Fig. A3(b) and was found in run 3. Clearly, when one measurement of an observation is an outlier, the entire field observation is treated as an outlier. Even though it is possible to take the results from the best run and use the estimates for the parameters x, a, b and c as a new starting guess for further iterations, it has been found that this does not improve the solution in this particular case. This is because the linearization around the values of a, b and c calculated

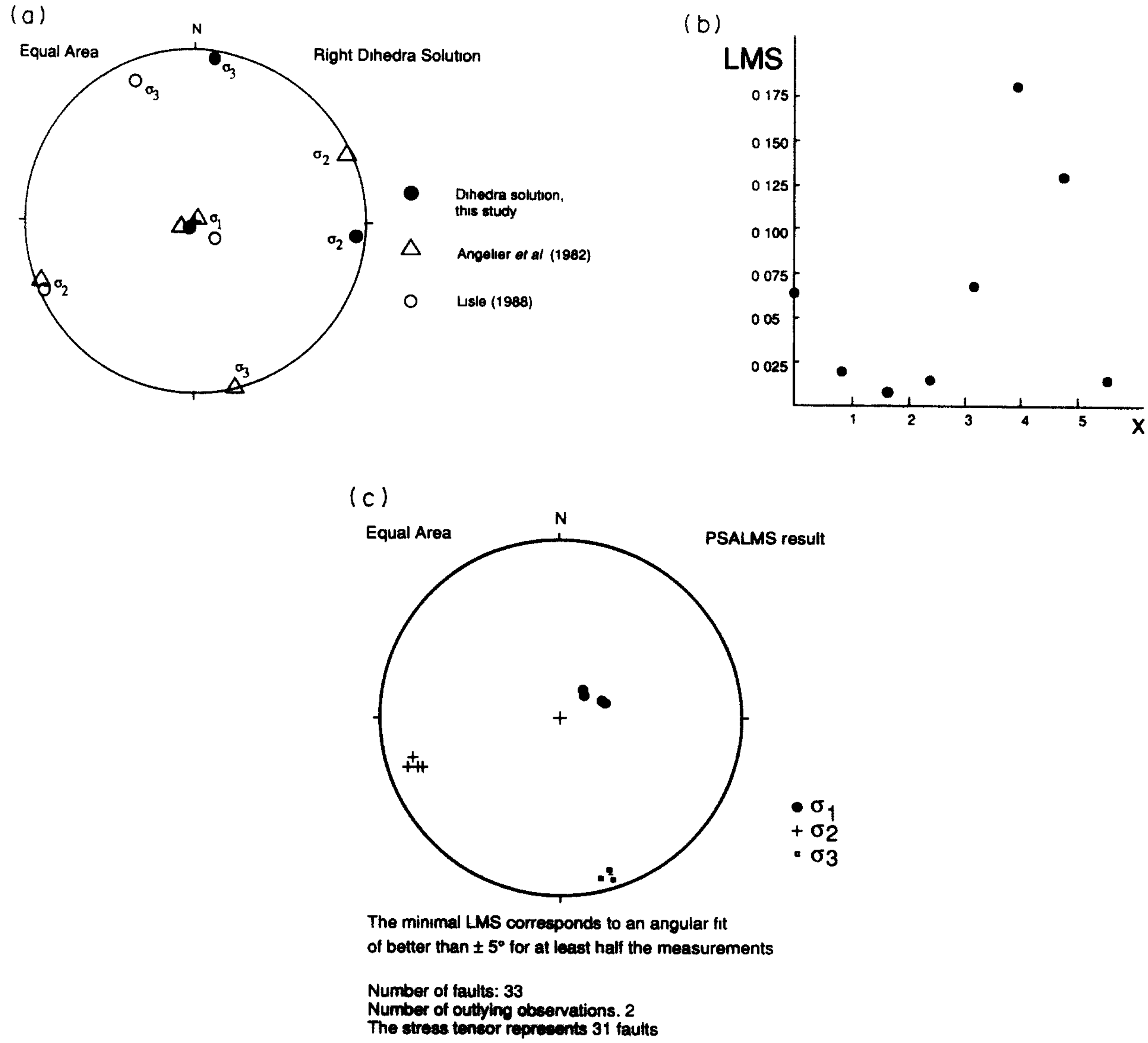

Fig. A3. (a) Results produced by part 2 of the program described here (bold circles) as compared with the results of Angelier *et al.* (1982) and Lisle (1988) (open triangles and open circles, respectively). The data-set analysed consists of 33 normal faults from Crete and is given by Angelier *et al.* (1982) The combined likelihood for the calculated x_i–z_i pair to be σ_1 and σ_3 is 71%. (b) LMS vs starting guess diagram. Note that the horizontal axis is in radians. Starting guess 3 produced the smallest LMS value and, thus gives the best solution for these data. (c) The LMS solution; original *PSALMS* output showing the orientations of all principal stresses whose LMS values pass the χ^2-test. For starting guess 3, the minimal median is 0.009 (i.e at least half of the angles are fitted better than ±5°) leading to a σ_{fit} of 0.16 and to acceptable answers at a 95% confidence level that have LMS values smaller than 0.013 (i.e. an angular misfit of less than ±6.5°). Equal-area, lower-hemisphere projection.

from the dihedra method, with the appropriate x starting guess, will normally apply at the solution. However, if the dihedra method gives a poorly constrained starting guess, further iteration is advisable. All the paleostress orientations plotted on Fig. A3(c) represent a region of parameter space that contains principal stress directions that are compatible with the given data-set and are appropriate solutions to the inverse problem. Applications of *PSALMS* to multimodal cases are given in Will (1990) and Wilson *et al.* (in press)

Table A1. Results for the Angelier *et al.* (1982) data-set. One iteration of 100 loops was executed for each starting guess for x, with the values of a, b and c calculated from the results of the dihedra method

No. of run	Starting guess for x	Best least median of squares for each run	Parameters of the reduced stress tensor	Paleostress orientations
1	0	0.064	$x = 0.2855$ $a = 1.2627$ $b = -0.8531$ $c = -1.4764$	σ_1 31 → 061 σ_2 56 → 271 σ_3 14 → 160
2	$\pi/4$	0.0196	$x = 1.3646$ $a = 0.1199$ $b = -0.0572$ $c = -0.2976$	σ_1 66 → 082 σ_2 24 → 265 σ_3 01 → 174
3	$\pi/2$	0.0090	$x = 1.6744$ $a = 0.0955$ $b = -0.1093$ $c = -0.2672$	σ_1 74 → 075 σ_2 14 → 264 σ_3 03 → 173
4	$3\pi/4$	0.0123	$x = 2.1866$ $a = 0.0151$ $b = -0.1020$ $c = -0.1873$	σ_1 82 → 059 σ_2 04 → 181 σ_3 07 → 271
5	π	0.0684	$x = 2.6419$ $a = -0.2078$ $b = -0.3043$ $c = -0.2625$	σ_1 77 → 020 σ_2 14 → 163 σ_3 10 → 255
6	$5\pi/4$	0.1824	$x = 3.7233$ $a = -0.4654$ $b = -0.3483$ $c = -1.4316$	σ_1 49 → 107 σ_2 16 → 358 σ_3 37 → 256
7	$3\pi/2$	0.1297	$x = 6.5283$ $a = 2.9400$ $b = 0.8533$ $c = 1.2419$	σ_1 20 → 233 σ_2 69 → 048 σ_3 02 → 142
8	$7\pi/4$	0.0147	$x = 5.7996$ $a = 1.3874$ $b = 0.3268$ $c = 0.4209$	σ_1 11 → 234 σ_2 75 → 100 σ_3 11 → 326