Journal of Structural Geology 30 (2008) 1593–1602

Contents lists available at ScienceDirect

Journal of Structural Geology

journal homepage: www.elsevier.com/locate/jsg

Non-linear least squares ellipse fitting using the genetic algorithm with applications to strain analysis

Anandaroop Ray [1], Deepak C. Srivastava*

Department of Earth Sciences, IIT Roorkee, Roorkee, UA 247667, India

ARTICLE INFO

Article history:
Received 28 February 2008
Received in revised form
10 September 2008
Accepted 11 September 2008
Available online 1 October 2008

Keywords:
Best-fit ellipse
Algebraic methods
Geometric methods
Genetic algorithm
Strain analysis

ABSTRACT

Several methods of strain estimation require the best-fit ellipse through a set of points either for defining elliptical shapes of distorted objects, and/or for tracing the finite strain ellipse. Fitting an ellipse to scattered points by solving a least squares problem can involve a linear as well as non-linear formulation. This article outlines both approaches and their relative merits and limitations and, proposes a simple yet powerful non-linear method of solution utilizing the genetic algorithm.

Algebraic methods solve the linear least squares problem, and are relatively straightforward and fast. However, depending upon the type of constraints used, different algebraic methods will yield somewhat different results. More importantly, algebraic methods have an inherent curvature bias – data corrupted by the same amount of noise will misfit unequally at different curvatures. The genetic algorithm method we propose uses geometric as opposed to algebraic fitting. This is computationally more intensive, but it provides scope for placing visually apparent constraints on ellipse parameter estimation and is free from curvature bias. Algebraic and geometric approaches are compared critically with the help of a few synthetic and natural examples for strain estimation in rocks. The genetic algorithm almost always produces results with lower misfit when dealing with noisy data and more importantly, yields closer estimates to the true values.

1. Introduction

Fitting an ellipse to an arbitrary set of co-planar points is a problem of fundamental importance in many fields of applied science ranging from observational astronomy and digital image processing to structural geology. This article addresses the problem of identifying the best-fit ellipse with special reference to the analysis of strain in natural rocks.

Strain estimation for rock deformation typically involves either elliptical strain markers (Ramsay, 1967, pp. 202–220; Ramsay and Huber, 1983, pp. 75–86; Lisle, 1985, pp. 2–27) or data from strain markers that are analyzed to yield a strain ellipse (Wellman, 1962; Mitra, 1978; Fry, 1979; Erslev and Ge, 1990; Lisle, 1992; Shah and Srivastava, 2006, 2007; Srivastava and Shah, 2008). An issue for either situation is accurate measurement of the strain ellipse, particularly when the elliptical markers are incomplete due to processes such as pressure solution, or data populations are either small or only represent a portion of the ellipse perimeter. An important factor for improving accuracy is the precision in determining the best-fit ellipse from the available data.

The best-fit problem can be viewed in the following manner – given a set of points in a plane, estimate the parameters of the ellipse which "best-fit" the data. Various "least-squares" fitting approaches have been formulated over the years (Zhang, 1997), but they all fall into two categories: (1) algebraic methods, which are extensively used due to their linear nature, simplicity and computational efficiency, and (2) geometric methods that solve a non-linear problem. In this article, we demonstrate that the genetic algorithm, using a geometric fitting approach, provides a more robust alternative than algebraic fitting – although it is computationally more intensive. After a brief mathematical outline of algebraic and geometric methods, this article develops the geometric fitting approach using the genetic algorithm (Holland, 1975). It shows that the genetic algorithm is not only a powerful optimization technique but it is also fairly simple to implement.

Genetic algorithms are global optimization techniques that use the mechanics of natural selection or biological evolution to converge to a solution (Goldberg, 1989, 2002). This algorithm has already received acclaim in a wide variety of optimization/parameter estimation/inversion problems, such as in the solution of complicated non-linear optimization problems in traffic system design, computational biology, acoustical oceanography (Gerstoft,

* Corresponding author. Tel.: +91 1332 285558; fax: +91 1332 273560.
E-mail address: dpkesfes@iitr.ernet.in (D.C. Srivastava).
[1] Present address: Shell Technology India, Bangalore 560 048, India.

0191-8141/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsg.2008.09.003

1994), electronic circuit design, speech processing, as well as seismic data migration (Jervis et al., 1993). To the best of our knowledge, the potential of the genetic algorithm has not yet been explored in structural geology.

2. Algebraic methods

The general equation of an ellipse is given as

$$ax^2 + bxy + cy^2 + dx + ey + f = 0 \tag{1}$$

with

$$b^2 - 4ac \leq 0 \tag{2}$$

If we have a set of m points which lie exactly on an ellipse, then Eq. (1) would be satisfied for all such points. If the points, however, do not lie exactly on the ellipse, then we would have m equations of the following form:

$$ax_i^2 + bx_iy_i + cy_i^2 + dx_i + ey_i + f = r_i \tag{3}$$

This forms a linear system of m equations in a, b, c, d, e, f with a residual r_i. We can write this in matrix form as

$$\begin{pmatrix} x_1^2 & x_1y_1 & y_1^2 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_i^2 & x_iy_i & y_i^2 & x_i & y_i & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_m^2 & x_my_m & y_m^2 & x_m & y_m & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \end{pmatrix} = \begin{pmatrix} r_1 \\ \vdots \\ r_i \\ \vdots \\ r_m \end{pmatrix} \tag{4}$$

or, more compactly as

$$X\mathbf{u} = \mathbf{r} \tag{5}$$

with $\mathbf{u}$ given as

$$\mathbf{u} = (a \quad b \quad c \quad d \quad e \quad f)^t \tag{6}$$

The problem then is to estimate the model vector $\mathbf{u}$ such that $\mathbf{r}$ is minimized. It requires finding:

$$\|X\mathbf{u}\| = \min \tag{7}$$

with $\mathbf{u}$ subject to some specific constraint such as the Singular Value Decomposition (SVD) constraint with $\|\mathbf{u}\| = 1$. Explicitly, the SVD is a factorization of rectangular matrices into a diagonal matrix pre-multiplied by an orthogonal matrix and post-multiplied by a transposed orthogonal matrix. As demonstrated by Varah (1996) various other constraints can also be prescribed.

In essence, algebraic least squares methods attempt to find a set of parameters $\mathbf{u}$ that makes the system of Eq. (4) as close as possible to that of the equation of a perfect ellipse. The methods proposed by Erslev and Ge (1990), Mulchrone et al. (2003), and Mulchrone and Roy Choudhury (2004) amply describe the use of algebraic least squares methods. Fitzgibbon et al. (1999) present an efficient ellipse fitting technique solved directly by a generalized eigensystem. They, however, incorporate a specific ellipticity constraint $b^2 - 4ac = -1$ to pull solutions away from singularities and be specific to ellipse fitting rather than generalized conic fitting. Hart and Rudman (1997) use linear least squares to fit the Cartesian equation of an ellipse centred at the origin. Their approach transforms the data not centred at the origin such that the centroid of the shape is shifted to the origin. The method for calculation of this centroid can be found in Mulchrone et al. (2003).

Algebraic methods all have the indisputable advantage of solving a linear least squares problem. The methods for this are well known and fast. However, it is intuitively unclear what it is we are minimizing geometrically – the right hand side of Eq. (4) is often referred to as the "algebraic distance" to be minimized. A geometric interpretation given by Bookstein (1979) clearly demonstrates that algebraic methods neglect points far from the centre. Closely following Zhang (1997) we can understand this last statement as follows: an origin-centred conic (Fig. 1 after Zhang, 1997) can be written as

$$Q = Ax^2 + Cy^2 + F = 0 \tag{8}$$

The algebraic distance of a point (x_i,y_i) from the conic is given as (Bookstein, 1979)

$$Q(x_i, y_i) = Ax_i^2 + Cy_i^2 + F = -F\left(\frac{d_i^2}{c_i^2} - 1\right) \tag{9}$$

where d_i is the distance of (x_i,y_i) from the centre O of the conic and c_i is the length of the ray from O to the conic along the straight line joining O to the point (x_i,y_i). For different points with the same Euclidean misfit $|d_i - c_i|$, the ratio d_i/c_i will be different at different curvatures. Thus with algebraic methods, we have a curvature bias for the same Euclidean misfit, i.e. the Cartesian distance $|d_i - c_i|$ (Fig. 1). Simply speaking, if the data are corrupted by the same amount of noise at all curvatures, the algebraic misfit (Eq. (9)) will be different at different curvatures! This characteristic is undesirable. Finally, different minimizing constraints on the model vector $\mathbf{u}$, such as the Bookstein constraint (Bookstein, 1979), produce very different ellipses. For example, this problem is evident from Fig. 2 (after Gander et al., 1994), where the dashed and solid lines correspond to different constraints on $\mathbf{u}$ for the same data set rotated and translated.

3. Geometric methods

The parametric equation of an ellipse is

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} a\cos\varphi \\ b\sin\varphi \end{pmatrix} \tag{10}$$

where x_0, y_0 are the co-ordinates of the centre of the ellipse, α is the angle of inclination of the axis of length a from the x co-ordinate axis, b is the length of the axis that is perpendicular to a, and φ is a parameter that runs from 0 to 2π anticlockwise from the a axis (Fig. 3).

Thus, the ellipse is completely characterized by the model vector $\mathbf{u}$ such that:

$$\mathbf{u} = (x_0 \quad y_0 \quad a \quad b \quad \alpha)^t \tag{11}$$

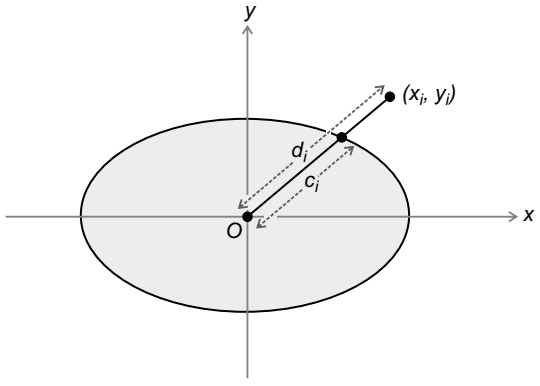

Fig. 1. Origin-centred conic (after Zhang, 1997). d_i is the distance of the point (x_i,y_i) from the centre O of the conic and c_i is the length of the ray from O to the conic along the straight line joining O to the point (x_i,y_i).

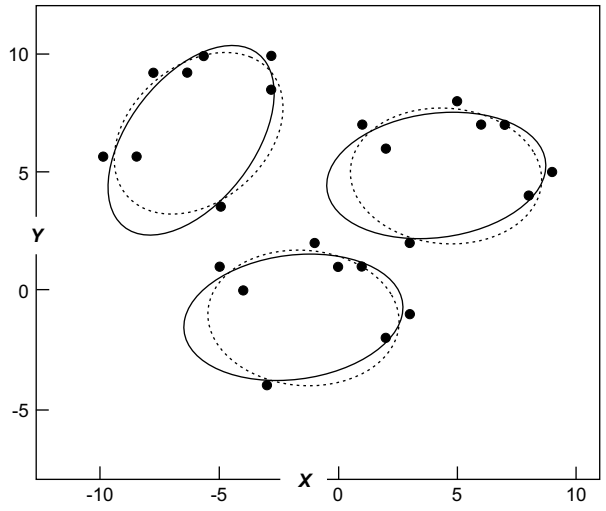

Fig. 2. Algebraically fit ellipses satisfying different Euclidean invariant constraints on the model vector $\mathbf{u}$ $(a, b, c, d, e, f)^t$ (after Gander et al., 1994).

Consider a case where the given point data do not lie exactly on the ellipse (Fig. 3). The distance of the ith data point from the centre of the proposed ellipse is given by $d_i(\mathbf{u})$, where

$$d_i(\mathbf{u}) = \left\| \begin{pmatrix} x_i - x_0 \\ y_i - y_0 \end{pmatrix} \right\| \tag{12}$$

The distance from the centre of this ellipse to the point on the ellipse along the same straight line for which Eq. (12) has been calculated is given by $c_i(\mathbf{u})$, where

$$\begin{aligned} c_i(\mathbf{u}) &= \left\| \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} a\cos\varphi_i \\ b\sin\varphi_i \end{pmatrix} \right\| \\ &= \left(a^2\cos^2\varphi_i + b^2\sin^2\varphi_i \right)^{1/2} \end{aligned} \tag{13}$$

Thus, we can define a function $F_i(\mathbf{u})$ as follows:

$$F_i(\mathbf{u}) = \|d_i(\mathbf{u}) - c_i(\mathbf{u})\| \tag{14}$$

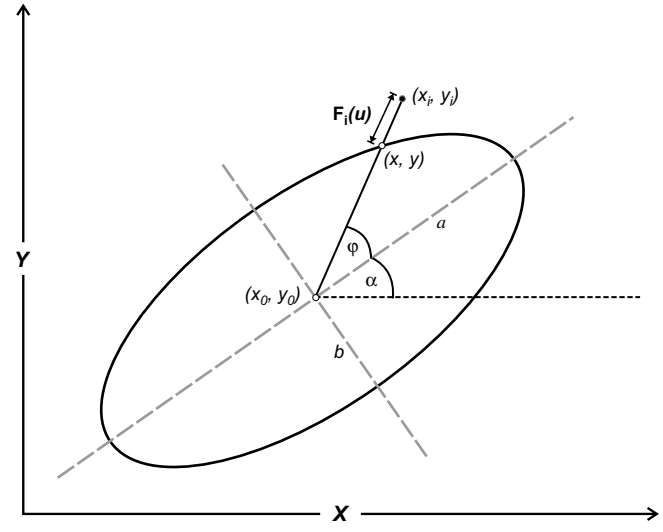

Fig. 3. Parameters characterizing an ellipse model vector. x_0, y_0 are the co-ordinates of the centre of the ellipse. α is the angle between the horizontal line and the axis of length a. b is the length of the axis perpendicular to a. φ is required in the parametric representation of the ellipse, running from 0 to 2π anticlockwise from the a axis. The distance between the ith data point (x_i, y_i) and a model ellipse characterized by a specific $\mathbf{u}$ is $F_i(\mathbf{u})$.

$F_i(\mathbf{u})$ is the distance of the ith data point from the ellipse (Fig. 3). $F_i(\mathbf{u})$ for all m data points no longer forms a linear set of equations in the model with parameters a, b, x_0, y_0, α. One, therefore, needs to solve a non-linear least squares problem to find the best-fit ellipse. Taking all the $F_i(\mathbf{u})$ into consideration, an objective function, i.e. a measure of misfit to be minimized can be formulated as the sum of $F_i^2(\mathbf{u})$. Thus the problem is then to find a model $\mathbf{u}$ such that

$$G(\mathbf{u}) = \sum_1^m F_i^2(\mathbf{u}) = \min \tag{15}$$

The calculation in Eq. (15) of the total misfit requires the evaluation of φ_i for each of the m data points. For a given model vector $\mathbf{u}$, the computation of $F_i(\mathbf{u})$ requires that φ_i be known. For the sake of readability, the computation of φ_i, though not complicated, is explained in Appendix A.

3.1. Resolvability of parameters and sensitivity analysis

Having formulated an inverse problem as in Eq. (15), a study of whether this formulation will enable us to estimate the parameters of interest, i.e. a, b, x_0, y_0, α, is in order. In other words, a study of how well these parameters can be resolved needs to be conducted to gauge the effectiveness of the chosen misfit function (Eq. (15)). The particular set of model parameters that produce a minimum in Eq. (15) will parametrize the ellipse that best-fits the observed m data points.

The misfit must appreciably differ for different input parameters. This can be demonstrated for points lying on an ellipse with parameters $(x_0, y_0, a, b, \alpha) = (2, 3, 0.3, 0.4, 35)$. Because a five-dimensional illustration is too complex, a two-dimensional model space is used for this purpose. If all best-fit ellipse parameters other than the axes lengths a and b were known, the problem would effectively reduce to a two-dimensional optimization problem (Fig. 4). Fig. 4 shows a misfit surface plotted against a range of ellipse axis lengths for a and b. The lowest point on the misfit surface is located at (0.3, 0.4, 0.0) on the a–b plane (horizontal). Thus the best-fit ellipse estimates the true axis lengths of 0.3 and 0.4, using the misfit defined by Eq. (15).

The above sensitivity to estimation analysis gives us glimpses of only a few slices into a five-dimensional model space, in which a model that produces minimum misfit is sought. One could perform a grid search and search every point in this five-dimensional model parameter space for the point of lowest misfit, but that would be computationally prohibitive. The search method utilized in this article is therefore the genetic algorithm, which avoids a grid search, quickly finds regions of minima in misfit and finally converges to the global minimum.

4. The genetic algorithm

In complicated multi-dimensional non-linear inverse problems, misfit functions are often oscillatory and strong parameter interdependencies make the estimation of a global misfit, such as that given by Eq. (15), highly difficult to solve. It is in such situations where genetic algorithms have been found to be particularly effective (Sen and Stoffa, 1995).

4.1. Method of operation

Genetic algorithms work with a population of models, and unlike in any other optimization method, model parameters are coded in binary. The basic steps are encoding, selection, crossover and mutation of models until the evolution of a satisfactory model, i.e. the convergence to a solution is reached. In brief, a description

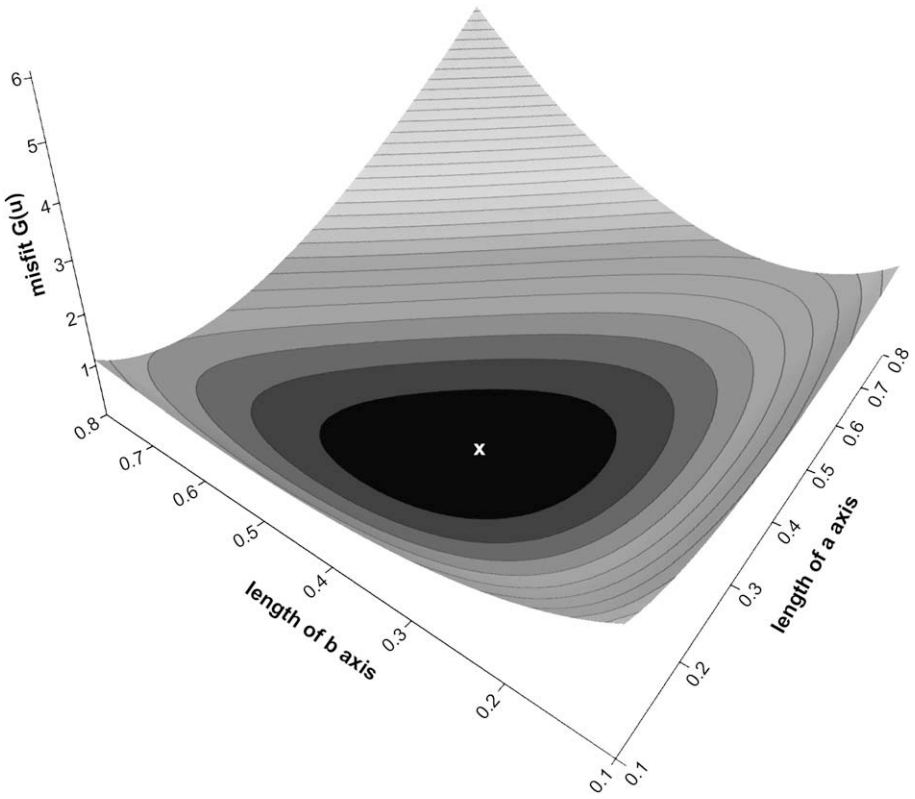

Fig. 4. A misfit surface which varies with the model ellipse axis lengths *a* and *b*. Lower values of misfit are darker, with a minimum (white cross) at (0.3, 0.4, 0.0). The true values of *a* and *b* are 0.3 and 0.4, respectively.

of the above is given here, specific to the case of finite strain estimation using elliptical distributions of points where the parameters of the best-fitting ellipse are sought. The basic principles discussed here are, however, equally applicable to the use of genetic algorithms in any complicated parameter estimation problem. Further, it must be mentioned at the outset that there are a few factors controlling the operation of genetic algorithms which at a first glance seem arbitrarily chosen. However, as is the case in the solution of any inverse problem, forward modelling is the key to understanding the solutions obtained. The use of synthetic data sets for the estimation of ellipse parameters using the genetic algorithm proved invaluable in deciding the controlling factors.

4.1.1. Encoding

Models in genetic algorithms are used in the same sense as chromosomes in biology. Further, model parameters are regarded as genes, which are the components of chromosomes. Thus, in Eq. (11), **u** is a model (chromosome) consisting of five model parameters (genes) a, b, x_0, y_0, α. Biologically speaking, genes are binary encoded. So in genetic algorithms, model parameters are encoded in binary. In this implementation, 8-bit encoding for each of the model parameters is used.

At the start of the genetic algorithm, 64 model vectors are randomly initialized. Each of the model parameters within a particular model are automatically randomly assigned values from within a range that is appropriate for the problem being considered. In the case of finite strain estimation, it is obvious that for elliptical distributions of points (Fig. 5), the centre (x_0,y_0) of the best-fit ellipse must lie within the region defined by the set of observed points. Further in this case, the maximum length of the long axis cannot exceed that of the diagonal of the rectangle defined by the most extreme x and y co-ordinates observed (Fig. 5). Finally, for the inclination α of axis a with respect to the datum, it is for sure that either of the short or long arms must exist in the first quadrant (see Fig. 3). Thus, we can restrict α to the range $(0, \pi/2)$. Specific situations could exist, where the search ranges for the genetic algorithm cannot be decided from the rectangle as described above. In such situations, any lenient estimate of the region constraining the centre and of the major axis would suffice to start the genetic algorithm. As expected, the smaller the

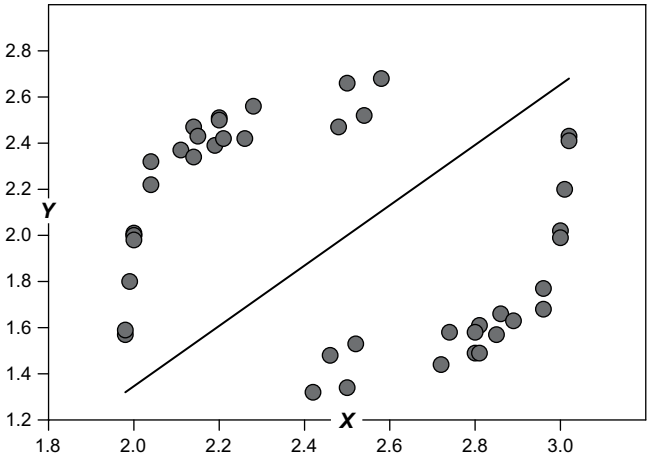

Fig. 5. A natural example that requires searching for the best-fit strain ellipse through points given by the Wellman (1962) method. The points are the corners of the parallelograms obtained by drawing parallels to the sides of the distorted centra in a vertebrate fossil (for details see Shah and Srivastava, 2007). The diagonal line belongs to a rectangle which bounds the data. This line is used for generating model parameter bounds.

constraining search ranges are, the quicker the algorithm converges to a solution.

The number of model vectors is somewhat arbitrary, and can be decided with experience and forward modelling. The fewer the number of model vectors, the faster the algorithm runs. However, the initial population of model solutions will have fewer approximations to the best-fitting model, so convergence to a good solution may be slow (Section 4.1.3).

4.1.2. Selection

In this process, models from the current model pool are selected for reproduction. Reproduction will lead to the formation of offspring models which will be better approximations to the best-fitting model. For purposes of selection, the misfit of each of these models is calculated using Eq. (15). There is a factor known as the reproduction factor f such that of the N models, 64 in this implementation, fN are selected for reproduction in a process known as crossover. The reproduction factor f has been taken as 0.5 in this implementation. The guiding principle for choosing f is that if too few parents are involved in crossover, new solutions will not be formed quickly enough, and if too many are crossed over then computational overhead increases without any significant improvement. Selection is based on a probability distribution (Sen and Stoffa, 1995), which ensures that a model with smaller misfit should have a greater chance of reproduction. A probability distribution that achieves this result is a normalized Boltzmann distribution. For the kth model vector, the selection probability is then given by

$$P(\mathbf{u}_k) = \frac{\exp[-G(\mathbf{u}_k)/T]}{\sum_1^{64} \exp[-G(\mathbf{u}_k)/T]} \tag{16}$$

where $G(\mathbf{u})$ is the objective function given by Eq. (15) and T is the minimum value of $G(\mathbf{u})$ in the model pool in the particular generation of models being considered for selection. Selection via Eq. (16) ensures that models with small values of objective function or misfit have a greater chance of being selected, whereas models with larger values of misfit have nearly similar chances of selection owing to the nature of the negative exponential curve (Fig. 6).

4.1.3. Crossover

Crossover is one of the most important steps in the genetic algorithm. Selected parents are randomly paired and allowed to reproduce. Two offspring models are produced as a consequence. Each parent model is allowed to reproduce only once (Fig. 7). The process of reproduction involves the crossing-over of parent model

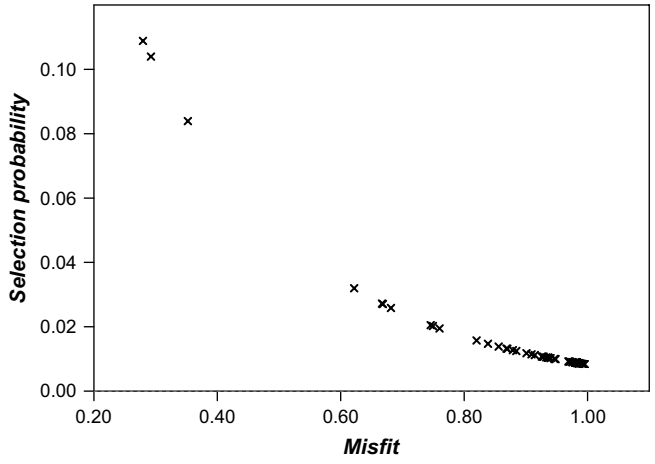

Fig. 6. The Boltzmann probability distribution.

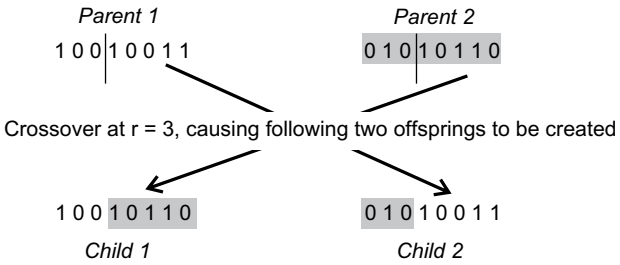

Fig. 7. Schematic of reproduction by crossover.

characteristics, i.e. genes, or model parameters to the offspring models. Crossover is at the heart of the combinatorial process that allows the genetic algorithm to converge towards a region of good solution model vectors without doing an exhaustive grid search of the five-dimensional model space.

For a particular parameter, for example the x co-ordinates of the centre of a candidate model ellipse, the process of crossover is as follows (Fig. 7).

A random bit from 0 to 7 (shown here from left to right) is selected. Say it is the rth bit.

1. The first child model is made to be a perfect, bit-by-bit copy of the first parent up to, but not including the rth bit. The remaining bits in the first child from r to 7 are copied perfectly, bit-by-bit from the other parent.
2. For the second child, the reverse procedure is followed. The second parent is perfectly copied first, and then from the rth bit onwards, the first parent is copied, bit-by-bit.

When only one model parameter is crossed over in a reproduction, it is called single point crossover. If all model parameters are crossed over, we have multiple point crossover, which is our approach.

4.1.4. Mutation

To prevent the genetic algorithm from converging to local minima in the objective function as well as to introduce a flavour of newness in the model pool, a mutation of parameters within the offspring models is triggered. This is made to occur with a low-order probability between 0.1 and 0.5. For a particular parameter within a given model, mutation is effected by randomly complementing a bit. Complementing is flipping a bit from 0 to 1 or vice versa. In this implementation, we applied mutations to all model parameters with the probability 0.4. Higher mutation rates will not allow the genetic algorithm to stabilize to a good solution whereas lower mutation rates will cause rapid convergence to solutions which may only be local minima, as opposed to the global minimum in misfit. The optimal mutation rate can be judiciously decided after experience in solving the problem at hand – namely, by forward modelling and solving for synthetic data sets.

4.1.5. Stopping criterion

After mutation, the fN newly created offspring models as well as the fittest $(1-f)$N of the original population are combined to form the new N member pool of models. These models form the next generation of models to be subjected to selection, crossover and mutation again. This process of evolution of fitter generations continues until the desired minimum in misfit is reached. In the case of noisy data, a certain number of maximum iterations are preferred as the stopping criterion. Since it is not known beforehand what minimum value will be reached, or what will be the maximum number of iterations required to reach it, the algorithm

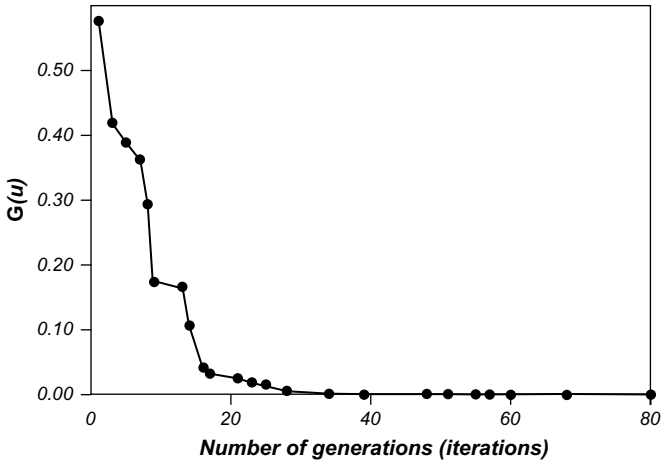

Fig. 8. Evolution of better fitting models with successive generations. A rapid convergence in the direction of the true solution is followed by a less rapid phase of solution refinement. Objective function $G(\mathbf{u})$ is the misfit.

must be run a few times and monitored to ensure that after a certain number of iterations, misfit values do not change substantially (Fig. 8), and that the values arrived at after this many iterations are nearly the same in all the trials.

4.2. Summary of steps involved in the genetic algorithm

1. A population of model vectors is randomly initialized.
2. The misfit $G(\mathbf{u})$ of each model vector is found.
3. fN models are selected for reproduction with probability of selection given by Eq. (16).
4. Parent models are randomly paired and crossed over to produce two new offspring models.
5. These offspring are subject to mutation with a low-order probability, set between 0.1 and 0.5.
6. The fittest $(1-f)$N of the parent models, together with the fN offspring models form the next N member generation of models.
7. This new model population is then subjected to steps 2–6 until the stopping criterion is met.

5. Application: synthetic data sets and demonstrations of curvature bias

5.1. Noiseless data

Estimation of ellipse parameters for both noiseless and noisy data is made by using our program written in C++. For noiseless data inversion, the data set corresponding to the sensitivity plots was used (Section 3.1). Fig. 8 shows the manner of convergence of the genetic algorithm to the solution (Table 1). Fig. 9 shows both the synthetic data set as well as the ellipse fit by the genetic algorithms.

Table 1
Results of noiseless parameter estimation. x_0, y_0 are the co-ordinates of the centre of the ellipse. α is the angle between the horizontal line and the axis of length a. b is the length of the axis perpendicular to a.

Parameter	True value	Estimate bounds (automatically calculated)	Inverted value
x_0	2.0	1.66–2.33	1.99783
y_0	3.0	2.63–3.37	2.99956
a	0.3	0–0.99	0.30115
b	0.4	0–0.99	0.39893
α	35.0	0–90	34.9412

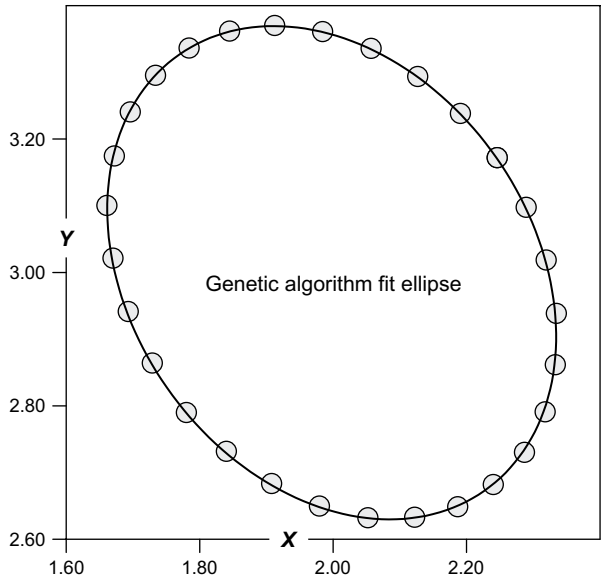

Fig. 9. The best-fit ellipse through the noiseless data in a synthetic example. Table 1 gives results in detail.

The evolution of fitter models (Fig. 8) till estimation of the solution took less than 80 generations. Refinements to the solution after the 40th generation were minor. The genetic algorithm inversions must be run a number of times to gain confidence in the inverted values, because of the random manner in which models are initialized. Genetic algorithms always converge to good solutions, and in 2–3 trials it is obvious what the correct parameter values are.

5.2. Noisy data sets and curvature bias

It is well known that most of the data sets obtained in geological situations are noisy. For this reason, we now examine three noisy but origin-centred data sets for which the true parameters, i.e. the axial ratio and the inclination of the true ellipse are known. The points in the first two examples (Fig. 10a,b) are located at opposite ends of lines passing through the centre, whereas the points in the third example do not satisfy this condition (Fig. 10c). These examples are particularly instructive for examining an important undesirable aspect of algebraic methods, namely, curvature bias.

The first data set is a striking example of curvature bias in algebraic methods (Fig. 10a). By design, there are about five times more data points in areas of low curvature than in the high curvature areas. The algebraically best-fit ellipse using the method of Fitzgibbon et al. (1999) estimates an ellipse of axial ratio 3.62 whereas the genetic algorithm estimates the ratio as 4.03 for a true ratio of 4.0 (Table 2). This difference in results is primarily because the algebraic method gives unequal weightage to the misfit of points at the curved sections of the best-fit ellipse. As mentioned earlier in Section 2 and pointed out by Bookstein (1979) and Zhang (1997) – this is an inevitable feature of solving this particular least squares problem algebraically. However, if one were to use Euclidean distances as the geometric methods do, then this problem is absent. Euclidean or geometric misfit (Eq. (15)) is curvature independent.

In the next two examples, the distribution of points around the ellipse is progressively less skewed (Fig. 10b,c). The results from both data sets lead to a similar inference, i.e. the algebraic method does not weigh points misfitting at high curvature in the same way

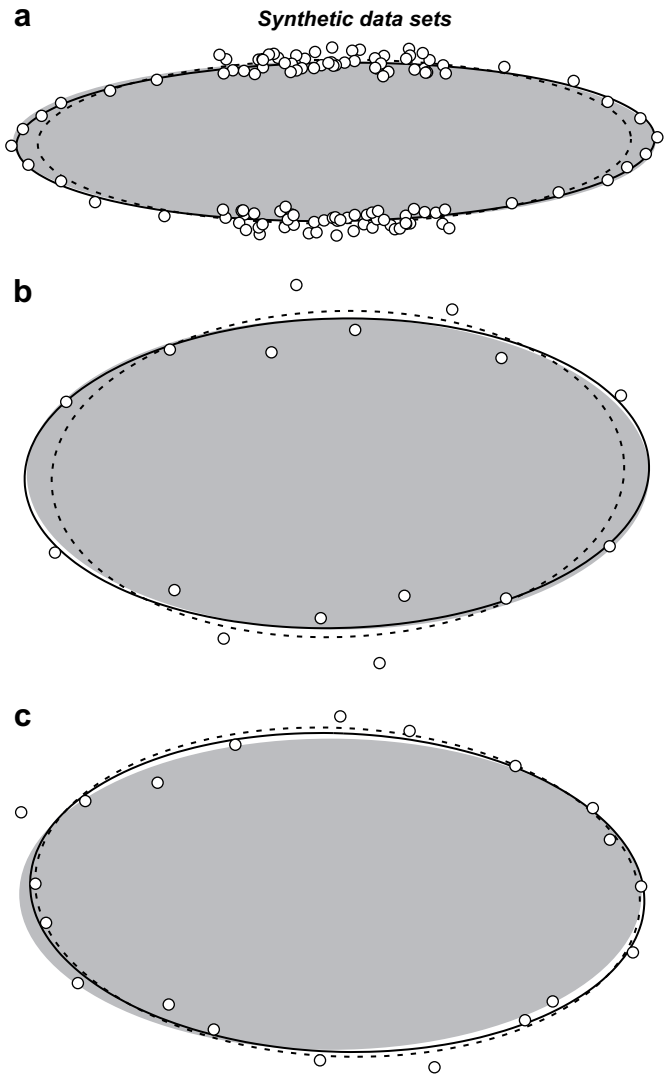

Fig. 10. (a)–(c) Three synthetic examples of noisy data sets with differences in the nature of data distribution with respect to the ellipse centre. The best-fit ellipses obtained by the genetic algorithm and the algebraic method of Fitzgibbon et al. (1999) are shown by solid and dashed lines, respectively. The true ellipse is the grey ellipse. Details of results are given in Table 2.

as it does points at flatter sections of the proposed best-fit ellipse, whereas the geometric genetic algorithm method does (Table 2).

6. Applications in strain analysis: natural examples

Pebbles (Brun et al., 1981), oolites (Dunnet, 1969; Boulter, 1976), amygdales (Riberio et al., 1983), varioles (Barr and Coward, 1974), xenoliths (Coward, 1976; Ramsay, 1983), and reduction spots (Siddans, 1979) are some common examples of elliptically shaped strain markers. The shapes of these markers may always not be perfectly elliptical. In such cases, their elliptical shapes are obtained by best-fitting the ellipse through the available points on the circumference of the objects. The best-fit approach, besides approximating the elliptical shapes of the markers, is also useful in tracing the strain ellipse through a set of points (Wellman, 1962; Mitra, 1978; Erslev and Ge, 1990; Lisle, 1992; Shah and Srivastava, 2006, 2007; Srivastava and Shah, 2008).

For finite strain estimation using elliptical markers, the genetic algorithm automatically determines the search ranges for model ellipse parameters as described in Section 4.1.1. Thus, all that is required for the program to run is a list of files containing the scattered data to be fit to different ellipses, and a limit on the maximum number of iterations.

Table 2
Comparison of results obtained by application of the algebraic method (Fitzgibbon et al., 1999), and the genetic algorithm method (this study) on synthetically produced noisy data sets. RMS – root mean square.

Example	Parameter	True ellipse	Best-fit ellipse method	
			Algebraic	Genetic algorithm
Fig. 10a (120 points)	Axial ratio	4.0	3.621	4.033
	Inclination	0.0	0.484	0.706
	Total square misfit	0.0	1.995	1.439
	RMS misfit	0.0	0.129	0.110
Fig. 10b (16 points)	Axial ratio	2.0	1.757	2.017
	Inclination	0.0	1.952	1.412
	Total square misfit	0.0	0.338	0.275
	RMS misfit	0.0	0.145	0.131
Fig. 10c (20 points)	Axial ratio	2.0	1.843	1.930
	Inclination	0.0	−2.793	2.118
	Total square misfit	0.0	0.171	0.164
	RMS misfit	0.0	0.092	0.091

6.1. Examples

A situation where determination of the strain ellipse from distorted vertebrae requires finding the best-fit ellipse through points given by the Wellman method (Wellman, 1962) was introduced in Fig. 5. We now consider more examples for the application of the genetic algorithm method – one for tracing elliptical shapes of distorted markers and the other three for tracing strain ellipses.

The first example searches for the best-fit ellipse through the corners of distorted corals (*Favosites*) occurring in the Silurian slates of Pembrokeshire, South Wales (after Fig. 8.14 in Ramsay and Huber, 1983, p. 135). One approach for strain estimation in these *Favosites* is to assume that the shape of each distorted *Favosites* was a perfect regular hexagon and use the Mohr circle method as described in Ramsay and Huber (1983, pp. 145–146). The other approach is to approximate shapes of the *Favosites* as ellipses by fitting the ellipses through the corners of each distorted *Favosite* using the genetic algorithm and use the R_f/ϕ method (Lisle, 1985). For illustration, we have fitted the ellipses through the corners of 19 distorted *Favosite* forms (Fig. 11a, Table 3). The application of the R_f/ϕ method reveals that the two-dimensional strain ratio is 1.61, and the major axis of the strain ellipse makes an angle of 156° from an arbitrary reference line. Having determined the axial ratio and the orientation of the strain ellipse, it is possible to retrodeform the image of the distorted *Favosites* with the help of computer aided graphics software (e.g., Srivastava and Shah, 2006a,b). Such a retrodeformation reveals that all the ellipses do not assume circular shapes in the undistorted image. It follows, therefore, that the undistorted *Favosites* were not perfectly regular hexagons (Fig. 11b).

Strain estimation in the second example requires searching for the best-fit ellipse through the polar plots of the stretches in rutile needles (Fig. 8 in Mitra, 1978). In the third example, distorted oolites are strain markers (Fig. 7.7 in Ramsay and Huber, 1983, p. 112) and the best-fit ellipse passes through the rim of high density points given by the enhanced normalized Fry method (Erslev and Ge, 1990). The fourth and final example estimates strain in a flattened parallel fold by tracing the best-fit ellipse through the polar plots of inverse thickness at different angles of the limb dip (Fig. 2 in Lisle, 1992).

The best-fit strain ellipses were determined by the genetic algorithm as well as the algebraic method of Fitzgibbon et al. (1999)

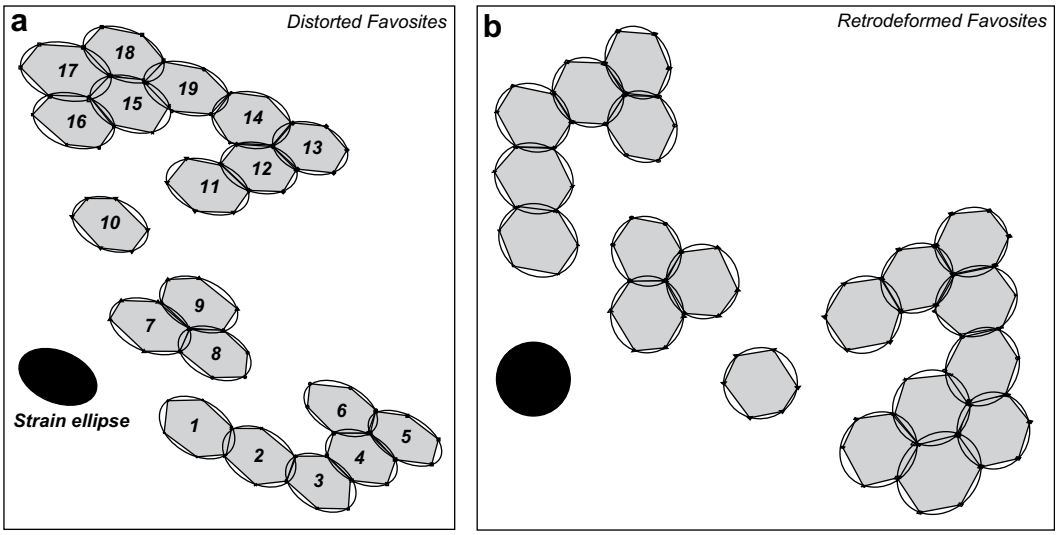

Fig. 11. An example demonstrating use of the genetic algorithm in tracing the elliptical shapes of distorted markers. (a) The best-fit ellipse through the corners of 19 hexagonal forms of distorted *Favosites* (1–19). Axial ratios and orientations of the best-fit ellipses through the corners of these hexagonal forms are given in Table 3. The axial ratio of strain ellipse is 1.61 and the major axis of strain ellipse makes an angle of 156° with respect to horizontal datum. (b) Retrodeformed image of the distorted fossils.

in the second, third and fourth examples (Fig. 12a–c and Table 4). The distribution of points in Fig. 12a and, particularly in Fig. 12b is such that the results obtained by the genetic algorithm and the algebraic method are not significantly different (Table 4). This result is, however, not true for the flattened fold example in Fig. 12c, where the distribution of points is such that the curvature bias becomes an important issue – note the lack of points along the highly curved portions of the ellipse in Fig. 12c and a similar situation in Fig. 10b. As a consequence of the nature of the distribution pattern of the data points, the algebraic method and the genetic algorithm yield significantly different axial ratios, 1.98 and 2.24 respectively, in this case (Fig. 12c and Table 4).

7. Discussion and conclusions

Algebraic methods of ellipse fitting are linear and computationally efficient. They have four limitations: (i) different algebraic methods may produce significantly different results due to different minimizing constraints (Fig. 2), (ii) algebraic methods may yield ambiguous results for data sets where only a few points are available and geological evidence such as cleavage traces or stretching lineations are lacking to constrain the direction of maximum stretching – Fig. 4 in Gander et al. (1994) demonstrates this aspect, (iii) algebraic methods generally have an inherent curvature bias (Fig. 10b, also Fig. 12c), and (iv) algebraic methods do not use Euclidean misfit as fitting criterion at all – so it is intuitively unclear what is being minimized.

Although geometric fitting eliminates these limitations, it was previously avoided because most geometric algorithms require an initial starting estimate usually provided by an algebraic method, and is computationally many orders of magnitude more intensive. The latter limitation has been largely overcome by the computational capacity available in common computers nowadays. The genetic algorithm method is a robust alternative to algebraic fitting and requires only very broad search ranges to start (Section 4.1.1). It is particularly preferable over algebraic methods in situations where only a few data points are available and the data are noisy. A further advantage of the geometric method with the genetic algorithm is that unlike the Gauss–Newton or other such geometric methods, no calculation of derivatives for Jacobian or Hessian matrices or successive Givens rotations are required.

In summary, algebraic methods succeed when the distribution of data points around the true ellipse leads to minimal curvature bias. For example, both the algebraic and genetic algorithm methods yield similar strain ratios (Fig. 12a,b) in the examples of stretched rutile needles (Mitra, 1978) and distorted oolites (Ramsay and Huber, 1983, p. 112). However, where such distributions of points are lacking, e.g. the flattened parallel fold in Fig. 12c (Lisle, 1992), curvature bias for the algebraic method is an issue and the strain ratios obtained from the two approaches could be significantly different (Table 4). Often the distribution of data points cannot be controlled, and the fact that algebraic methods weigh points at different curvatures unequally is an important consideration. Weighting schemes can be used with algebraic methods to address this issue, but they do not minimize geometric distances no matter how good the weighting. Furthermore, weighting schemes have problems with regard to stability – small errors in observed data can lead to very incorrect estimates.

Table 3
The axial ratios and orientations of the best-fit ellipses given by genetic algorithm method, through the corners of 19 hexagonal forms of the distorted *Favosites* in Fig. 11a.

Hexagon number (Fig. 11a)	Axial ratio of the best-fit ellipse	Orientation of the best-fit ellipse	Root mean square misfit
1	1.51282	148.235	0.00706574
2	1.45455	150.00	0.0496767
3	1.51282	147.529	0.0386598
4	1.55405	151.412	0.0118186
5	1.81819	149.647	0.0418701
6	1.65279	151.765	0.0478829
7	1.61643	154.235	0.0305336
8	1.66666	158.118	0.00488621
9	1.62857	156.706	0.0172958
10	1.53522	158.471	0.0160585
11	1.63236	160.941	0.0366858
12	1.56338	165.177	0.0203292
13	1.52779	163.412	0.0302161
14	1.40789	158.353	0.0481277
15	1.57143	158.823	0.0255242
16	1.57144	153.059	0.0559155
17	1.60869	156.941	0.00709206
18	1.68182	157.412	0.0198016
19	1.72307	163.412	0.0286962

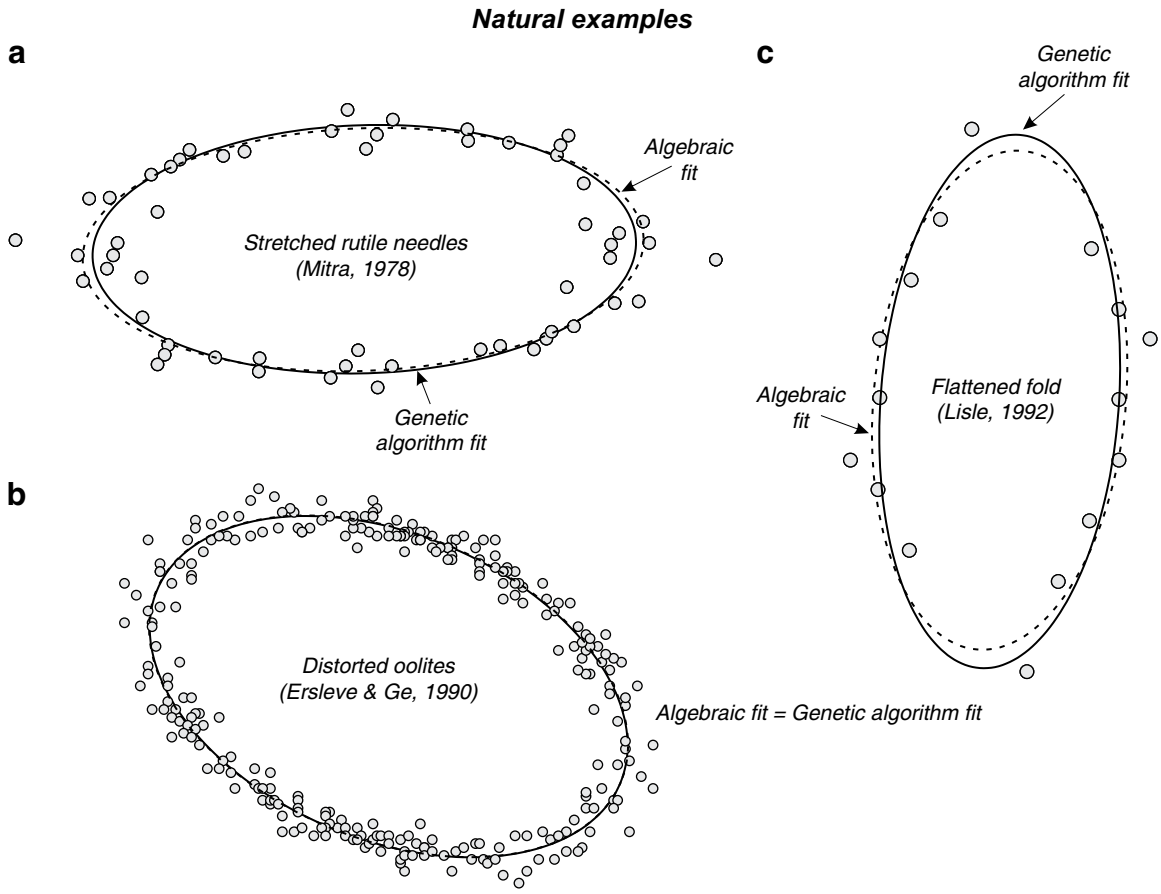

Fig. 12. Natural examples of best-fit strain ellipses. The solid and dashed best-fit ellipses are given by the genetic algorithm method and the algebraic method (Fitzgibbon et al., 1999), respectively. Data sets in (a), (b) and (c) are obtained from analyses of stretched rutile needles (Mitra, 1978), distorted oolites (Ramsay and Huber, 1983, p. 112) and flattened parallel fold (Lisle, 1992), respectively. Details of results are given in Table 4.

Other misfit functions than the one described by Eq. (15) can be used in conjunction with the genetic algorithm. Two prospective alternatives are the use of the L_1 norm, i.e. misfit for the ith point will be defined as $|x - x_i| + |y - y_i|$ in Fig. 3, and use of the shortest distance from a data point to the model ellipse. The alternate objective functions, however, present difficulties such as finding roots of a quadratic equation for the shortest distance method. This need will add to the computational overhead already involved in the process of selection for crossover. No conceptual barriers, however, prevent the use of alternate misfit functions owing to the flexibility of the genetic algorithm. Finally, it is hoped that with this study, the usefulness of genetic algorithms in finite strain estimation is demonstrated and that they will be applied in future work in this field.

Table 4
Comparison of results obtained by application of the algebraic method (Fitzgibbon et al., 1999), and the genetic algorithm method (this study) on three natural data sets (Fig. 12a–c). RMS – root mean square.

Example	Parameter	Best-fit ellipse method	
		Algebraic	Genetic algorithm
Rutile needles Fig. 12a (56 points)	Axial ratio	2.3138	2.19048
	Inclination	2.2982	1.7647
	Total square misfit	4.1563	4.0208
	RMS misfit	0.2724	0.2680
Oolites Fig. 12b (276 points)	Axial ratio	1.6413	1.6441
	Inclination	−22.4199	−22.5882
	Total square misfit	0.3183	0.3180
	RMS misfit	0.0340	0.0339
Flattened fold Fig. 12c (16 points)	Axial ratio	1.9853	2.2449
	Inclination	84.6830	85.7647
	Total square misfit	1764.910	1552.6438
	RMS misfit	10.5030	9.8509

Acknowledgements

We thank Bill Dunne, Declan De Paor, K. Mulchrone, Richard Lisle and Michal Nemcok for constructive reviews, and Jyoti Shah and Aditi Pal for testing various examples. Richard Lisle provided consistent encouragement and the Department of Science and Technology, India funded this study.

Appendix A. Calculation of φ_i

Referring to Fig. 3, for the ith data point (x_i,y_i), the angle φ_i from the axis of length a is calculated by considering the point (x,y) on the ellipse at the intersection of the ellipse itself with the line joining (x_i,y_i) to the centre of the ellipse. We write

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} a\cos\varphi_i \\ b\sin\varphi_i \end{pmatrix} \tag{A.1}$$

The slope of the line joining the centre, (x,y) and (x_i,y_i), is given by p, where

$$p = \frac{y - y_0}{x - x_0} = \frac{y_0 - y_i}{x_0 - x_i} \tag{A.2}$$

For a given data set and a proposed model ellipse, (x_i,y_i) and (x_0,y_0) are known. Using these values in the RHS of Eq (A.2), we calculate p. Substituting p in Eq (A.1):

$$\frac{1}{p} = \frac{a\cos\alpha\cos\varphi_i - b\sin\alpha\sin\varphi_i}{a\sin\alpha\cos\varphi_i + b\cos\alpha\sin\varphi_i}$$
$$\Rightarrow b\sin\varphi_i(p\sin\alpha + \cos\alpha) = a\cos\varphi_i(p\cos\alpha - \sin\alpha)$$
$$\Rightarrow \tan\varphi_i = \frac{a}{b}\left(\frac{p\cos\alpha - \sin\alpha}{p\sin\alpha + \cos\alpha}\right) \quad \text{(A.3)}$$

For the case where the line is vertical, p goes to infinity and the expression (A.3) reduces to

$$\tan\varphi_i = \frac{a}{b\tan\alpha} \quad \text{(A.4)}$$

While initializing the model ellipse pool, zero lengths for the a and b axes are automatically rejected by the genetic algorithm and reset to non-zero values. This ensures that φ_i values are always computed without a 0/0 situation arising and more importantly, taking care of meaningless axial ratios.

References

Barr, M., Coward, M.P., 1974. A method for the measurement of volume change. Geological Magazine 111, 293–296.

Bookstein, F.L., 1979. Fitting conic sections to scattered data. Computer Graphics and Image Processing 9, 56–71.

Boulter, C.A., 1976. Sedimentary fabrics and their relation to strain-analysis methods. Geology 4, 141–146.

Brun, J.P., Gapais, D., Le Theoff, B., 1981. The mantled gneiss domes of Kuopio (Finland). Tectonophysics 74, 283–304.

Coward, M.P., 1976. Archean deformation patterns in southern Africa. Philosophical Transaction of the Royal Society of London A283, 313–331.

Dunnet, D., 1969. A technique of finite strain analysis using elliptical particles. Tectonophysics 7, 117–136.

Erslev, E.A., Ge, H., 1990. Least squares centre-to-centre and mean object ellipse fabric analysis. Journal of Structural Geology 10, 201–209.

Fitzgibbon, A., Pilu, M., Fisher, R.B., 1999. Direct least squares fitting of ellipses. IEEE Transactions on Pattern Analysis and Machine Intelligence 21, 476–480.

Fry, N., 1979. Random point distributions and strain measurement in rocks. Tectonophysics 60, 89–105.

Gander, W., Golub, G.H., Rolf, S., 1994. Fitting of circles and ellipses least squares solution. BIT Numerical Mathematics 34, 556–577.

Gerstoft, P., 1994. Inversion of seismoacoustic data using genetic algorithms and *a posteriori* distributions. Journal of the Acoustical Society of America 95, 770–782.

Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization and Machine Learning. Kluwer Academic Publishers, Boston, MA.

Goldberg, D.E., 2002. The Design of Innovation: Lessons from and for Competent Genetic Algorithms. Addison-Wesley, Reading, MA.

Hart, D., Rudman, A.J., 1997. Least-squares fit of an ellipse to anisotropic polar data: application to azimuthal resistivity surveys in karst regions. Computers and Geosciences 23, 189–194.

Holland, J.H., 1975. Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor.

Jervis, M., Stoffa, P.L., Sen, M.K., 1993. 2-D migration velocity estimation using a genetic algorithm. Geophysical Research Letters 14, 1495–1498.

Lisle, R.J., 1985. Geological Strain Analysis. A Manual for the R_f/ϕ Technique. Pergamon press, Oxford, 99 pp.

Lisle, R.J., 1992. Strain estimation from flattened parallel folds. Journal of Structural Geology 14, 369–371.

Mitra, S., 1978. Microscopic deformation mechanisms and flow laws in quartzites within the South Mountain anticline. Journal of Geology 86, 129–132.

Mulchrone, K.F., O'Sullivan, F., Meere, P., 2003. Finite strain estimation using the mean radial length of elliptical objects with confidence intervals. Journal of Structural Geology 25, 529–539.

Mulchrone, K.F., Roy Choudhury, K.R., 2004. Fitting an ellipse to an arbitrary shape: implications for strain analysis. Journal of Structural Geology 26, 143–153.

Ramsay, J.G., 1967. Folding and Fracturing of Rocks. McGraw-Hill Book Company, New York, 568 pp.

Ramsay, J.G., 1983. Emplacement kinematics of a granitic diaper: the Chindamora batholith, Zimbabwe. Journal of Structural Geology 11, 191–209.

Ramsay, J.G., Huber, M.I., 1983. Strain analysis. The Techniques of Modern Structural Geology, vol. 1. Academic Press Inc., London, 307 pp.

Riberio, A., Kullberg, M.C., Possolo, A., 1983. Finite strain estimation using anti-clustered distribution of points. Journal of Structural Geology 5, 233–244.

Sen, M., Stoffa, P.L., 1995. Global Optimization Methods in Geophysical Inversion. Elsevier Science, B.V.

Shah, J., Srivastava, D.C., 2006. Strain estimation from flattened parallel folds: application of the Wellman method and Mohr circle. Geological Magazine 143, 243–247.

Shah, J., Srivastava, D.C., 2007. Strain estimation from distorted vertebrate fossils: application of the Wellman method. Geological Magazine 144, 211–216.

Siddans, A.W.B., 1979. Deformation, metamorphism and texture development in Permian mudstone of the Glarus Alps. Eclogae Geologicae Helvetiae 72, 601–621.

Srivastava, D.C., Shah, J., 2006a. A rapid method for strain estimation from flattened parallel folds. Journal of Structural Geology 28, 1–8.

Srivastava, D.C., Shah, J., 2006b. Digital method for strain estimation and retro-deformation of bilaterally symmetric fossils. Geology 34, 593–596.

Srivastava, D.C., Shah, J., 2008. The 'isogon rosette' method for rapid estimation of strain in flattened folds. Journal of Structural Geology 30, 444–450.

Varah, J.M., 1996. Least squares data fitting with implicit functions. BIT Numerical Mathematics 36, 842–854.

Wellman, H.G., 1962. A graphic method for analyzing fossil distortion caused by tectonic deformation. Geological Magazine 99, 348–352.

Zhang, Z., 1997. Parameter estimation techniques: a tutorial with application to conic fitting. Image and Vision Computing 15, 59–76.