Appendix I: Algorithm

Least-Squares Formalism

I assume that all geologic and paleomagnetic data that constrain displacement, strain, and/or rotation in a particular timestep have been transformed to scalar rate estimates r_k^n. (The subscript $k = 1, \ldots, K$ identifies the datum, and the superscript n identifies the timestep.) Let the corresponding scalar rate predictions derived from the velocity field of the finite-element model in a particular timestep be called p_k^n. (In the remainder of this section I will discuss only a single computational time-step, so the superscript n will be suppressed.) I further assume that each scalar rate r_k has an uncertainty that can be approximated by a Gaussian probability distribution with standard deviation σ_k, and that the errors in these rates are independent. (A Gaussian probability distribution is reasonably appropriate for the numerator in each rate, which is an amount of strain or displacement. It is not appropriate for the denominator, the elapsed time, which typically comes with a hard upper limit based on cross-cutting relations but without any lower limit. Because of this, I will iterate the solution of the entire history in a way that allows additional degrees of freedom in each rate history; this method is described below. However, in each individual iteration, the elapsed time is held constant, giving a Gaussian distribution for the rate.)

Finally, I assume that there is a probability $0 \le q_k \le 1$ that each datum is relevant to the timestep. Normally q_k is unity because irrelevant data are simply omitted from the list $k = 1, \ldots, K$. However, in the cases of certain paleostress data fractional q_k are required. The natural logarithm of the density of the joint probability that the velocity model matches all the relevant rates is then formed from the individual probability densities (Φ) as:

$$S \equiv \ln\left\{\prod_{k=1}^{K}\left[\Phi(p_k = r_k)\right]^{q_k}\right\} = \sum_{k=1}^{K} q_k \ln\left[\Phi(p_k = r_k)\right] = -\sum_{k=1}^{K} q_k \left[\frac{(p_k - r_k)^2}{2\sigma_k^2} + \ln(\sigma_k) + \ln\sqrt{2\pi}\right]. \quad (1)$$

I refer to this quantity S as the "score" of the velocity solution, which is to be maximized. That is, the sum of squares of the relevant prediction errors (each divided by the variance of the corresponding rate) is to be minimized.

On the surface of a spherical planet with radius *R*, define a coordinate system of colatitude (θ) measured southward from the North Pole, and longitude (ϕ) measured eastward from the prime meridian. The unknowns in each velocity solution are the horizontal θ-components and ϕ-components of the velocity of the surface. The predicted rates p_k can be expressed as a linear combination of the velocity components *v* (Southward) and *w* (Eastward) at each of the *J* nodes of a finite element grid:

$$p_k = c_k + \sum_{j=1}^{J} \left(f_{kj} v_j + g_{kj} w_j \right) \tag{2}$$

With this linear relation, *S* is a quadratic form in the nodal-velocity-component values v_j and w_j, so it is maximized by finding the single stationary point in multi-dimensional velocity space where

$$\frac{\partial S}{\partial v_i} = 0 = \frac{\partial S}{\partial w_i}; \quad i = 1, \ldots, J . \tag{3}$$

Algebraically, this leads to a $2J \times 2J$ linear system, which can be thought of as being partitioned into 4 submatrices times two subvectors equaling two subvectors:

$$\left[\begin{array}{c|c} A_{ij} & B_{ij} \\ \hline C_{ij} & D_{ij} \end{array} \right] \left[\begin{array}{c} v_j \\ \hline w_j \end{array} \right] = \left[\begin{array}{c} E_i \\ \hline F_i \end{array} \right] \tag{4}$$

using the abbreviations

$$\begin{aligned} A_{ij} &= \sum_{k=1}^{K} \frac{f_{ki} f_{kj}}{\sigma_k^2} \quad & B_{ij} &= \sum_{k=1}^{K} \frac{f_{ki} g_{kj}}{\sigma_k^2} \\ C_{ij} &= \sum_{k=1}^{K} \frac{g_{ki} f_{kj}}{\sigma_k^2} \quad & D_{ij} &= \sum_{k=1}^{K} \frac{g_{ki} g_{kj}}{\sigma_k^2} \end{aligned} \tag{5}$$

$$\begin{aligned} E_i &= \sum_{k=1}^{K} \frac{f_{ki} \left(r_k - c_k \right)}{\sigma_k^2} \\ F_i &= \sum_{k=1}^{K} \frac{g_{ki} \left(r_k - c_k \right)}{\sigma_k^2} \end{aligned} \tag{6}$$

Boundary Conditions

The equations stated above are singular in the absence of boundary conditions. Some edge(s) of the model must be fixed (or moved in a predetermined way) to provide a velocity reference frame. I replace the row equations that state that *S* is stationary with respect to variations in these nodal velocity components with simpler equations stating the desired values of these components.

This method permits only velocity boundary conditions, not stress boundary conditions. Along each axis (θ or ϕ) one boundary should be constrained and one left free (because integrating strain information to find velocities is like solving first-order differential equations). When one margin of a continent is facing a subduction zone, it is best to leave that boundary free.

A-priori or Pseudo-Data

An essential context for all the geologic data showing locally intense straining is that they should be overlaid on a set of *a-priori* data (or "pseudo-data") stating that in other places the strain-rate is close to zero. An appropriate formalism is to assign a zero target strain-rate, with a statistical uncertainty. A larger standard deviation should be attached to this null hypothesis in complex regions where unknown faults and orogenic phases might have been buried or overlooked.

To implement these constraints, the score S of any velocity solution, which is to be optimized, is first augmented by a term:

$$\Delta S = -W \iint_U \frac{\dot{\varepsilon}_{\theta\theta}^2 + \dot{\varepsilon}_{\theta\theta}\dot{\varepsilon}_{\phi\phi} + \dot{\varepsilon}_{\phi\phi}^2 + \dot{\varepsilon}_{\theta\phi}^2}{\mu^2} dU \tag{7}$$

where W is the weight coefficient for all pseudo-data (in units of m^{-2}), μ is the standard deviation of the (nominally zero) strain-rate (in s^{-1}), and U is the area without active faults. In practice, any finite element that has no active fault crossing it is a part of U. Weighting by area is used to make this new term roughly independent of local variations in finite element size. However, an appropriate value for the weight coefficient W is approximately the inverse of the mean of the areas of all finite elements, as explained below.

This term has the effect of causing unfaulted areas to behave as Newtonian-viscous sheets of lithosphere. The algorithm will adjust the velocity in each time step to minimize the area integral of squared strain-rates for these elements; this is exactly the result one obtains by beginning from the momentum equation (in the absence of horizontal forces), adopting a linear rheology, and solving for velocity with inhomogeneous boundary conditions.

The 2×2 strain-rate tensor $\dot{\tilde{\varepsilon}}$ on the spherical surface is calculated by summing spatial derivatives of the nodal functions. The nodal functions that I use were introduced by *Kong and Bird* [1995] and shown to satisfy the requirements of horizontality, continuity, and completeness:

$$\begin{bmatrix} v(\theta,\phi) \\ w(\theta,\phi) \end{bmatrix} = \sum_{j=1}^{J} \begin{bmatrix} G_{1,1}^{j}(\theta,\phi) & G_{2,1}^{j}(\theta,\phi) \\ G_{1,2}^{j}(\theta,\phi) & G_{2,2}^{j}(\theta,\phi) \end{bmatrix} \begin{bmatrix} v_j \\ w_j \end{bmatrix}. \tag{8}$$

In this notation, the superscript *j* on the vector nodal function $\vec{G}_x^j$ or nodal function component $G_{x,y}^j$ identifies the node that has unit velocity (all other nodes having zero velocity in this particular nodal function). Subscript $x=1$ indicates the nodal function associated with unit southward velocity *v*; subscript $x=2$ indicates the nodal function associated with unit eastward velocity *w*. Subscript $y=1$ indicates the southward or θ-component of the vector nodal function $\vec{G}_x^j$, and subscript $y=2$ indicates the eastward or ϕ-component.

The coefficients of the linear system are augmented by

$$\begin{aligned}
\Delta A_{ij} = \frac{W}{R^2}\sum_{\ell=1}^{L}\Bigg\{ \frac{1}{\mu_\ell^2}\iint_{a_\ell} \Bigg[& 2\frac{\partial G_{1,1}^i}{\partial\theta}\frac{\partial G_{1,1}^j}{\partial\theta} + \csc\theta\left(\frac{\partial G_{1,1}^i}{\partial\theta}\frac{\partial G_{1,2}^j}{\partial\phi} + \frac{\partial G_{1,2}^i}{\partial\phi}\frac{\partial G_{1,1}^j}{\partial\theta}\right) + \cot\theta\left(\frac{\partial G_{1,1}^i}{\partial\theta}G_{1,1}^j + G_{1,1}^i\frac{\partial G_{1,1}^j}{\partial\theta}\right) + \\
& 2\left(\csc\theta\frac{\partial G_{1,2}^i}{\partial\phi} + \frac{G_{1,1}^i}{\tan\theta}\right)\left(\csc\theta\frac{\partial G_{1,2}^j}{\partial\phi} + \frac{G_{1,1}^j}{\tan\theta}\right) + \frac{1}{2}\left(\csc\theta\frac{\partial G_{1,1}^i}{\partial\phi} + \frac{\partial G_{1,2}^i}{\partial\theta} - \frac{G_{1,2}^i}{\tan\theta}\right)\left(\csc\theta\frac{\partial G_{1,1}^j}{\partial\phi} + \frac{\partial G_{1,2}^j}{\partial\theta} - \frac{G_{1,2}^j}{\tan\theta}\right)\Bigg] da_\ell \Bigg\} \\
\Delta B_{ij} = \frac{W}{R^2}\sum_{\ell=1}^{L}\Bigg\{ \frac{1}{\mu_\ell^2}\iint_{a_\ell} \Bigg[& 2\frac{\partial G_{1,1}^i}{\partial\theta}\frac{\partial G_{2,1}^j}{\partial\theta} + \csc\theta\left(\frac{\partial G_{1,1}^i}{\partial\theta}\frac{\partial G_{2,2}^j}{\partial\phi} + \frac{\partial G_{1,2}^i}{\partial\phi}\frac{\partial G_{2,1}^j}{\partial\theta}\right) + \cot\theta\left(\frac{\partial G_{1,1}^i}{\partial\theta}G_{2,1}^j + G_{1,1}^i\frac{\partial G_{2,1}^j}{\partial\theta}\right) + \\
& 2\left(\csc\theta\frac{\partial G_{1,2}^i}{\partial\phi} + \frac{G_{1,1}^i}{\tan\theta}\right)\left(\csc\theta\frac{\partial G_{2,2}^j}{\partial\phi} + \frac{G_{2,1}^j}{\tan\theta}\right) + \frac{1}{2}\left(\csc\theta\frac{\partial G_{1,1}^i}{\partial\phi} + \frac{\partial G_{1,2}^i}{\partial\theta} - \frac{G_{1,2}^i}{\tan\theta}\right)\left(\csc\theta\frac{\partial G_{2,1}^j}{\partial\phi} + \frac{\partial G_{2,2}^j}{\partial\theta} - \frac{G_{2,2}^j}{\tan\theta}\right)\Bigg] da_\ell \Bigg\} \\
\Delta D_{ij} = \frac{W}{R^2}\sum_{\ell=1}^{L}\Bigg\{ \frac{1}{\mu_\ell^2}\iint_{a_\ell} \Bigg[& 2\frac{\partial G_{2,1}^i}{\partial\theta}\frac{\partial G_{2,1}^j}{\partial\theta} + \csc\theta\left(\frac{\partial G_{2,1}^i}{\partial\theta}\frac{\partial G_{2,2}^j}{\partial\phi} + \frac{\partial G_{2,2}^i}{\partial\phi}\frac{\partial G_{2,1}^j}{\partial\theta}\right) + \cot\theta\left(\frac{\partial G_{2,1}^i}{\partial\theta}G_{2,1}^j + G_{2,1}^i\frac{\partial G_{2,1}^j}{\partial\theta}\right) + \\
& 2\left(\csc\theta\frac{\partial G_{2,2}^i}{\partial\phi} + \frac{G_{2,1}^i}{\tan\theta}\right)\left(\csc\theta\frac{\partial G_{2,2}^j}{\partial\phi} + \frac{G_{2,1}^j}{\tan\theta}\right) + \frac{1}{2}\left(\csc\theta\frac{\partial G_{2,1}^i}{\partial\phi} + \frac{\partial G_{2,2}^i}{\partial\theta} - \frac{G_{2,2}^i}{\tan\theta}\right)\left(\csc\theta\frac{\partial G_{2,1}^j}{\partial\phi} + \frac{\partial G_{2,2}^j}{\partial\theta} - \frac{G_{2,2}^j}{\tan\theta}\right)\Bigg] da_\ell \Bigg\}
\end{aligned} \tag{9}$$

where $\ell = 1,\ldots,L$ identifies the "nonfaulting" elements comprising *U*, with individual areas a_ℓ. In practice, area integrals within each element are performed numerically, using 7 Gauss points with associated weights [*Zienkiewicz*, 1971].

Use of Balanced Cross-sections

Many structural geologists publish restored cross-sections from which they estimate the amount of shortening or extension along the line of section. Dividing the amount of extension (compression is negative extension) by the time available gives the relative velocity component (along the line of section) for the end points, which is the rate estimate r_k. If the endpoints of the section are marked by position vectors $\vec{b}_k$ and $\vec{d}_k$, then $c_k = 0$ and

$$\begin{aligned} f_{kj} &= G^j_{1,1}(\vec{b}_k)\cos\gamma(\vec{b}_k) - G^j_{1,1}(\vec{d}_k)\cos\gamma(\vec{d}_k) - G^j_{1,2}(\vec{b}_k)\sin\gamma(\vec{b}_k) + G^j_{1,2}(\vec{d}_k)\sin\gamma(\vec{d}_k) \\ g_{kj} &= G^j_{2,1}(\vec{b}_k)\cos\gamma(\vec{b}_k) - G^j_{2,1}(\vec{d}_k)\cos\gamma(\vec{d}_k) - G^j_{2,2}(\vec{b}_k)\sin\gamma(\vec{b}_k) + G^j_{2,2}(\vec{d}_k)\sin\gamma(\vec{d}_k) \end{aligned}, \quad (10)$$

where $\gamma(\vec{b}_k)$ and $\gamma(\vec{d}_k)$ are the forward azimuths at each end of the directed great circle arc $\vec{b}_k \to \vec{d}_k$ (each measured clockwise from North).

Use of Fault-slip data

A large fraction of the available data concern offsets on faults. While offset is actually a vector, I use only the larger of the dip-slip or strike-slip components and treat this as a scalar datum. This is because the strike-slip component of dominantly dip-slip faults is rarely known, while any dip-slip on strike-slip faults is irrelevant to relative horizontal velocities. After division by the time available, this scalar offset becomes a scalar relative velocity component across the fault.

When a fault is long enough to cross several finite elements, I impose the same slip and slip-rate in each element. In the case of rigid-microplate tectonics, where each fault connects to other faults at triple-junctions, this method is reasonably accurate. The other end-member is the case where no faults connect, but all terminate within the domain. In that case, each fault might be expected (on the basis of crack theory for linear materials) to have an ellipsoidal profile of slip versus length. Such "elliptical" faults would have a mean slip which is only 79% (π/4) of their maximum slip. Thus, my method might overstate strain by 27% in some cases where faults do not connect and where the geologic offsets reported are the maximum offsets. However, if the geologic offsets are considered to be determined at random points of convenience, then once again there is no systematic error.

If every fault extended continuously across the model from boundary to boundary, one could simply use its slip-rate as a constraint on the relative velocity of the nodes on opposite sides of the fault. However, the number of faults in many applications is so great that such customized grids are prohibitively expensive to work with. Thus, I have developed a more general approach, which allows any number of faults to cross a given finite element.

For each finite element, there are four steps: (a) Form the target strain-rate tensor for that element as the sum of the strain-rate tensors implied by all the active fault segments cutting that element; (b) Form the matrix of covariances of the strain-rate components in that element as the sum of the covariances added by all the active fault segments, plus the small covariance of the strain-rate in the continuum blocks between them; (c) Diagonalize the covariance matrix to find three principal axes (in strain-rate space) along which the uncertainties are independent, and also rotate the target strain-rates into this new coordinate system; (d) Add these 3 independent targets as scalar data with known uncertainties in the global system of equations.

The strain-rate tensor in the horizontal plane, $\dot{\tilde{\varepsilon}}$, is a second-rank tensor of size 2×2. I simplify the notation by treating the three independent components of the strain-rate tensor ($\dot{\varepsilon}_{\theta\theta} = \dot{\varepsilon}_{NS}, \dot{\varepsilon}_{\theta\phi} = \dot{\varepsilon}_{SE}, \dot{\varepsilon}_{\phi\phi} = \dot{\varepsilon}_{EW}$) as a one-subscript vector ($\dot{\varepsilon}_m$; $m = 1,2,3$), permitting us to write the covariance of strain-rates as a 3×3 matrix. If all the active fault segments that cut (even partway) through one finite element are numbered $z = 1, \ldots, Z$, then I express the strain-rate vector in the element as a linear combination of their scalar slip-rates s_z:

$$\dot{\varepsilon}_m = \sum_{z=1}^{Z} H_{zm} s_z ; \; m = 1,2,3 . \tag{11}$$

The covariance matrix of the strain-rate components is composed of two parts: the continuum compliance common to all parts of the lithosphere (see "*A-priori* or pseudo-data" above), and the terms arising from the standard deviations δs_z of the scalar slip-rates s_z:

$$\tilde{V} = \mu^2 \begin{bmatrix} 4/3 & 0 & -2/3 \\ 0 & 1 & 0 \\ -2/3 & 0 & 4/3 \end{bmatrix} + \sum_{z=1}^{Z} (\delta s_z)^2 \left[\vec{H}_z^{\mathrm{T}} \vec{H}_z \right] \tag{12}$$

To find $\vec{H}_z$ (the partial derivative of element strain-rate with respect to slip-rate of one active fault), I make the simplifying restriction that no node lies exactly on a fault. Also, I slightly straighten the traces of any fault segment that crosses the same element boundary more than once. Then, each fault segment (with its projected extensions, if necessary) must separate one node of the element from the other two. Let u_z be the index number of the isolated node. If node u_z is on the right side of the fault segment (when looking along its azimuth γ_z, measured clockwise from North), then I define the variable η_z as +1; otherwise, it is -1. Let κ_z be the fraction of the width of the element that is cut by the fault segment: $0 < \kappa_z \leq 1$.

In the case of a strike-slip fault, the scalar-slip rate s_z is defined as the right-lateral offset divided by the time available. (Left-lateral offsets are negative right-lateral offsets.) Then

$$\vec{H}_z = \frac{\eta_z \kappa_z}{R} \begin{pmatrix} \dfrac{\partial G_{1,1}^{u_z}}{\partial \theta} \cos\gamma_z - \dfrac{\partial G_{2,1}^{u_z}}{\partial \theta} \sin\gamma_z, \\ \dfrac{1}{2}\left(\dfrac{\partial G_{1,1}^{u_z}}{\partial \phi} \dfrac{\cos\gamma_z}{\sin\theta} - \dfrac{\partial G_{2,1}^{u_z}}{\partial \phi} \dfrac{\sin\gamma_z}{\sin\theta} + \dfrac{\partial G_{1,2}^{u_z}}{\partial \theta} \cos\gamma_z - \dfrac{\partial G_{2,2}^{u_z}}{\partial \theta} \sin\gamma_z - \dfrac{G_{1,2}^{u_z} \cos\gamma_z - G_{2,2}^{u_z} \sin\gamma_z}{\tan\theta} \right), \\ \dfrac{\partial G_{1,2}^{u_z}}{\partial \phi} \dfrac{\cos\gamma_z}{\sin\theta} - \dfrac{\partial G_{2,2}^{u_z}}{\partial \phi} \dfrac{\sin\gamma_z}{\sin\theta} + \dfrac{G_{1,1}^{u_z} \cos\gamma_z - G_{2,1}^{u_z} \sin\gamma_z}{\tan\theta} \end{pmatrix}. \quad (13a)$$

In the case of dip-slip faulting, it is most convenient to define s_z as the net horizontal extension perpendicular to the fault trace, divided by the time available. (Thrusting is considered to be negative extension.) In the case of detachment faulting, net horizontal extension is the distance from the breakaway fault in the foot-wall to the tip of the hanging-wall (reconstructed if necessary), regardless of whether the fault slipped at a low angle or, alternatively, slipped at a high angle and then rotated during further extension. In the more common case of dip-slip faulting without horizontal-axis rotation of foot-wall or hanging-wall, net horizontal extension is the relative vertical offset (throw) times the cotangent of the fault dip. My convention is that normal and detachment faulting have positive s_z and thrust faults have negative values. Then,

$$\vec{H}_z = \frac{\eta_z \kappa_z}{R} \begin{pmatrix} \dfrac{\partial G_{1,1}^{u_z}}{\partial \theta} \sin\gamma_z + \dfrac{\partial G_{2,1}^{u_z}}{\partial \theta} \cos\gamma_z , \\ \dfrac{1}{2}\left(\dfrac{\partial G_{1,1}^{u_z}}{\partial \phi} \dfrac{\sin\gamma_z}{\sin\theta} + \dfrac{\partial G_{2,1}^{u_z}}{\partial \phi} \dfrac{\cos\gamma_z}{\sin\theta} + \dfrac{\partial G_{1,2}^{u_z}}{\partial \theta} \sin\gamma_z + \dfrac{\partial G_{2,2}^{u_z}}{\partial \theta} \cos\gamma_z - \dfrac{G_{1,2}^{u_z} \sin\gamma_z + G_{2,2}^{u_z} \cos\gamma_z}{\tan\theta} \right), \\ \dfrac{\partial G_{1,2}^{u_z}}{\partial \phi} \dfrac{\sin\gamma_z}{\sin\theta} + \dfrac{\partial G_{2,2}^{u_z}}{\partial \phi} \dfrac{\cos\gamma_z}{\sin\theta} + \dfrac{G_{1,1}^{u_z} \sin\gamma_z + G_{2,1}^{u_z} \cos\gamma_z}{\tan\theta} \end{pmatrix} . \quad (13b)$$

The next step is to find the 3 positive eigenvalues (λ_h; $h = 1,2,3$) of $\widetilde{V}$ and their corresponding unit eigenvectors (Λ_{hm}). These eigenvectors indicate strain-rate patterns that are statistically uncorrelated; they have target amplitudes of $\sum_{m=1}^{3} \dot{\varepsilon}_m \Lambda_{hm}$ and standard deviations of $\sqrt{\lambda_h}$, respectively. Each of the three targets is now imposed as a scalar datum in the global system of equations. The corresponding coefficients of the nodal velocities are

$$\begin{aligned} f_{kj} &= \frac{1}{R} \left[\frac{\partial G_{1,1}^{j}}{\partial \theta}, \quad \frac{1}{2}\left(\csc\theta \frac{\partial G_{1,1}^{j}}{\partial \phi} + \frac{\partial G_{1,2}^{j}}{\partial \theta} - \frac{G_{1,2}^{j}}{\tan\theta} \right), \quad \csc\theta \frac{\partial G_{1,2}^{j}}{\partial \phi} + \frac{G_{1,1}^{j}}{\tan\theta} \right] \begin{bmatrix} \Lambda_{h1} \\ \Lambda_{h2} \\ \Lambda_{h3} \end{bmatrix} \\ g_{kj} &= \frac{1}{R} \left[\frac{\partial G_{2,1}^{j}}{\partial \theta}, \quad \frac{1}{2}\left(\csc\theta \frac{\partial G_{2,1}^{j}}{\partial \phi} + \frac{\partial G_{2,2}^{j}}{\partial \theta} - \frac{G_{2,2}^{j}}{\tan\theta} \right), \quad \csc\theta \frac{\partial G_{2,2}^{j}}{\partial \phi} + \frac{G_{2,1}^{j}}{\tan\theta} \right] \begin{bmatrix} \Lambda_{h1} \\ \Lambda_{h2} \\ \Lambda_{h3} \end{bmatrix} \end{aligned} . \quad (14)$$

(Note that k equals h plus a constant that indicates how many data have previously been incorporated into the system.)

Once the global velocity solution has been found for any timestep, it is necessary to do a local optimization calculation within each faulting element to find the *predicted* (model) rates p_z at which each fault ($z = 1, \ldots, Z$) is slipping, as well as the residual strain-rate $\dot{\varepsilon}_m^{\mathrm{c}}$ which is due to deformation of the continuum around the faults. The total strain-rate of the element must be the sum of the continuum and the fault contributions:

$$\dot{\varepsilon}_m^{\mathrm{c}} + \sum_{z=1}^{Z} H_{zm} p_z = \dot{\varepsilon}_m . \quad (15)$$

This problem is different from the global problem because the $\dot{\varepsilon}_m$ vector is known. Because of this constraint, it is an algebraic convenience to use the Lagrange multiplier method with three temporary weight variables (ς_1, ς_2, ς_3). Define the local score (in one element) that is to be optimized as:

$$S' \equiv -\sum_{z=1}^{Z} \frac{(p_z - s_z)^2}{(\delta s)^2} - \frac{\left(\dot{\varepsilon}_1^{\mathrm{c}\,2} + \dot{\varepsilon}_1^{\mathrm{c}} \dot{\varepsilon}_3^{\mathrm{c}} + \dot{\varepsilon}_3^{\mathrm{c}\,2} + \dot{\varepsilon}_2^{\mathrm{c}\,2}\right)}{\mu_\ell^2} - \sum_{m=1}^{3} \varsigma_m \left(\dot{\varepsilon}_m^{\mathrm{c}} + \sum_{z=1}^{Z} H_{zm} p_z - \dot{\varepsilon}_m \right) \qquad (16)$$

where δs_z is the standard deviation of each slip rate according to the input data. Then, to find a local solution that has all fault rates as close as possible to their goals, while the continuum strain-rate is close to zero, and the total strain-rate is correct, find the stationary point of S' with respect to variations in the p_z, the $\dot{\varepsilon}_m^{\mathrm{c}}$, and the ς_m in turn, leading to a linear system.
The slip-rate p_k^n that is finally recorded (for fault *k*, in timestep *n*) is the average of the rates p_z in all the elements the fault passes through, where the averaging weights are the segment lengths $\rho_{\ell z}$.

The method described here for incorporating faults effectively multiplies each known fault offset into roughly 1~2 scalar data per finite element crossed by the trace. In addition, 1~2 pseudo-data per element crossed are used to express the *a-priori* assumption of continuum stiffness. Therefore, for parity between faulting elements and non-faulting elements, the suggested value of *W* is the inverse of the mean area of finite elements.

Use of Paleomagnetic Data

I assume that the paleomagnetic dataset is restricted to sites that include some geologic or geochemical indication of the orientation of the paleo-horizontal plane at the time of magnetization, and that have been properly corrected for local structure. It is also best to exclude certain sedimentary rocks that are known to be especially prone to post-magnetization compaction, which can produce a non-tectonic inclination anomaly.

The interpretation of paleomagnetic inclination and declination data in terms of North-South displacement and rotation requires the definition of a reference polar wander path. It is necessary to use the same velocity reference frame for polar wander and for velocity boundary conditions.

The inclination of a sample yields its magnetic paleo-latitude according to a simple dipole model for the paleo-field. This is converted to a paleo-latitude anomaly by comparison with the polar wander path. Multiplying this paleolatitude anomaly by the radius of the planet and dividing by the length of time available results in the mean velocity component in the paleo-direction of magnetic South (at the time of magnetization). This becomes the rate estimate r_k. For comparison, the model prediction is formed using

$$\begin{aligned} f_{kj} &= G^j_{1,1}\cos\gamma - G^j_{1,2}\sin\gamma \\ g_{kj} &= G^j_{2,1}\cos\gamma - G^j_{2,2}\sin\gamma \end{aligned} \tag{17}$$

and $c_k = 0$, where the nodal functions $G^j_{x,y}$ are evaluated at the datum location, and γ is the azimuth (measured clockwise) of North-pointing paleomagnetic declinations (at the time of magnetization) with respect to present geographic North (in the reference frame of the undeformed part of the continent) at the datum location.

While the magnetic declination of a sample is clearly related to its history of rotation about a local vertical axis, the relationship is non-unique and model-dependent. In general, the declination anomaly can only be converted to a vertical-axis rotation if the displacement of the sample is negligible or approximately known. Here I assume that these questions have been resolved by the assumption or approximation most appropriate to the particular problem, and that estimated net rotations about the local vertical axis ($\Delta\gamma$) are available as input data. I also assume that the sense of large rotations has been decided in advance based on regional tectonics.

These values determine the average rotation rates $r_k = \omega_k^{(\text{datum})} = \Delta\gamma/t$, where $\Delta\gamma$ is the vertical-axis rotation in going from past to present (counterclockwise positive), and t is the age of the magnetization. The interpretation of these rotations requires some assumptions about the shape and the stiffness of the bodies carrying the remnant magnetization; I assume that the outcrops selected for paleomagnetic sampling were rigid inclusions of equidimensional shape embedded in a deforming continuum. Therefore, the rotation rate that the magnetization would be predicted to record is

$$p_k = \omega_k^{(\text{model})} \cong \frac{1}{2R}\left(\frac{w}{\tan\theta} + \frac{\partial w}{\partial\theta} - \csc\theta\,\frac{\partial v}{\partial\phi}\right). \tag{18}$$

Consequently, the model predictions are formed using $c_k = 0$ and

$$
\begin{aligned}
f_{kj} &= \frac{1}{2R}\left(\frac{G_{1,2}^{j}}{\tan\theta} + \frac{\partial G_{1,2}^{j}}{\partial\theta} - \csc\theta \frac{\partial G_{1,1}^{j}}{\partial\phi} \right) \\
g_{kj} &= \frac{1}{2R}\left(\frac{G_{2,2}^{j}}{\tan\theta} + \frac{\partial G_{2,2}^{j}}{\partial\theta} - \csc\theta \frac{\partial G_{2,1}^{j}}{\partial\phi} \right)
\end{aligned}
. \qquad (19)
$$

The methods described in previous sections effectively multiply the continuum-stiffness assumption into one scalar datum per finite element, and multiply each fault offset into 1~2 scalar data per finite element traversed. If the finite-element grid has many elements, the paleomagnetic data will require a similar weighting in order to avoid being overwhelmed in the global solution. Accordingly, both f and g functions of (17) and (19) should be multiplied by a dimensionless weight factor P which is common to all paleomagnetic data. A suggested value for P is the square root of twice the number of finite elements, which is the mean number of elements traversed by a fault crossing the whole grid. Then, a paleomagnetic datum indicating an exotic terrane will receive the same procedural weight as a fault-offset datum on the terrane-bounding fault, and the outcome will depend on the relative values and their uncertainties. Different values of P can be used to investigate solutions in which paleomagnetic data are either less or more prominent.

Use of Stress Directions

One principal stress direction must always be perpendicular to the free surface of the Earth, or approximately vertical. Thus, the orientation of the stress tensor is described by the azimuth (γ; measured clockwise from North) of the most-compressive horizontal principal stress ($\hat{\sigma}_{1\mathrm{h}}$). This direction is geologically recorded as the strike of igneous dikes or other vertical veins that break through any laterally-homogeneous, isotropic rock. In some cases a population of faults with slickensides can be statistically analyzed to determine the stress direction [*e.g., Gephart*, 1990].

To relate this information about stress to my kinematic model, I approximate the lithosphere as horizontally isotropic, so that the principal directions of stress are the same as the principal directions of strain rate. There may be an error of up to 35° associated with this assumption;

even so, I believe that the solutions will typically be more accurate and reasonable than those that ignore stress data.

Once we know the azimuth of $\hat{\sigma}_{1\mathrm{h}}$, we can use this as the direction of a new local horizontal axis $\hat{\alpha}$, and also define a perpendicular horizontal axis $\hat{\beta}$ (right-handed: $\hat{\alpha}\times\hat{\beta}=\hat{r}$). In these coordinates, the requirement that $\hat{\alpha}$ is the most-compressive horizontal principal strain-rate direction can be stated in two parts: $\dot{\varepsilon}_{\alpha\beta}=0$ and $\dot{\varepsilon}_{\alpha\alpha}<\dot{\varepsilon}_{\beta\beta}$. In terms of the global coordinate system, the first becomes

$$\dot{\varepsilon}_{\theta\phi}\cos(2\gamma)+\frac{\dot{\varepsilon}_{\theta\theta}-\dot{\varepsilon}_{\phi\phi}}{2}\sin(2\gamma)=0\,. \tag{20}$$

In terms of derivatives of velocity, this is

$$\frac{1}{2R}\left\{\left(\csc\theta\frac{\partial v}{\partial\phi}+\frac{\partial w}{\partial\theta}-\frac{w}{\tan\theta}\right)\cos(2\gamma)+\left(\frac{\partial v}{\partial\theta}-\csc\theta\frac{\partial w}{\partial\phi}-\frac{v}{\tan\theta}\right)\sin(2\gamma)\right\}=0\,, \tag{21}$$

so the coefficients of the linear system can be computed from the factors

$$\begin{aligned}
f_{kj}&=\frac{1}{2R}\left\{\left(\csc\theta\frac{\partial G_{1,1}^{j}}{\partial\phi}+\frac{\partial G_{1,2}^{j}}{\partial\theta}-\frac{G_{1,2}^{j}}{\tan\theta}\right)\cos(2\gamma)+\left(\frac{\partial G_{1,1}^{j}}{\partial\theta}-\csc\theta\frac{\partial G_{1,2}^{j}}{\partial\phi}-\frac{G_{1,1}^{j}}{\tan\theta}\right)\sin(2\gamma)\right\}\\
g_{kj}&=\frac{1}{2R}\left\{\left(\csc\theta\frac{\partial G_{2,1}^{j}}{\partial\phi}+\frac{\partial G_{2,2}^{j}}{\partial\theta}-\frac{G_{2,2}^{j}}{\tan\theta}\right)\cos(2\gamma)+\left(\frac{\partial G_{2,1}^{j}}{\partial\theta}-\csc\theta\frac{\partial G_{2,2}^{j}}{\partial\phi}-\frac{G_{2,1}^{j}}{\tan\theta}\right)\sin(2\gamma)\right\}
\end{aligned} \tag{22}$$

if we use $c_k=0$ and a rate estimate $r_k=0$. The difficulty is in deciding what standard deviation σ_k to associate with this constraint $\dot{\varepsilon}_{\alpha\beta}=0$, since we have transformed the constraint from one concerning an angle to one concerning a strain-rate component. When the calculation is first starting and there are no strain-rates known as yet, a purely arbitrary small strain rate uncertainty (ξ) must be assigned as σ_k. However, when strain rate estimates are available from a previous time step (or from a previous iteration of the current time step), it is better to use $\sigma_k=2(\delta\gamma)\sqrt{\dot{\varepsilon}_{\theta\phi}^2+\frac{1}{4}\left(\dot{\varepsilon}_{\theta\theta}-\dot{\varepsilon}_{\phi\phi}\right)^2}$, where the symbol $\delta\gamma$ indicates the standard deviation (in radians) of the azimuth γ of the direction $\hat{\sigma}_{1\mathrm{h}}$. This suggests that the velocity solution should be iterated within each timestep; in this project I have used 4 iterations per timestep.

The second requirement was the inequality $\dot{\varepsilon}_{\alpha\alpha}<\dot{\varepsilon}_{\beta\beta}$. During each iteration of the solution, I evaluate the strain rates $\dot{\varepsilon}_{\alpha\alpha}$ and $\dot{\varepsilon}_{\beta\beta}$ to see if this is true. If not, then in future iterations I impose an additional constraint $\dot{\varepsilon}_{\beta\beta} = \dot{\varepsilon}_{\alpha\alpha} + \xi$, where ξ is a small (positive) strain rate difference which must be arbitrarily chosen. In terms of the global coordinates,

$$\left(\dot{\varepsilon}_{\phi\phi} - \dot{\varepsilon}_{\theta\theta}\right)\cos(2\gamma) + 2\dot{\varepsilon}_{\theta\phi}\sin(2\gamma) = \xi . \tag{23}$$

This can be expressed in terms of velocity components as

$$\frac{1}{R}\left\{\left(\csc\theta\frac{\partial w}{\partial\phi} + \frac{v}{\tan\theta} - \frac{\partial v}{\partial\theta}\right)\cos(2\gamma) + \left(\csc\theta\frac{\partial v}{\partial\phi} + \frac{\partial w}{\partial\theta} - \frac{w}{\tan\theta}\right)\sin(2\gamma)\right\} = \xi , \tag{24}$$

so the coefficients of the linear system can be computed from the factors

$$\begin{aligned} f_{kj} &= \frac{1}{R}\left\{\left(\csc\theta\frac{\partial G^j_{1,2}}{\partial\phi} + \frac{G^j_{1,1}}{\tan\theta} - \frac{\partial G^j_{1,1}}{\partial\theta}\right)\cos(2\gamma) + \left(\csc\theta\frac{\partial G^j_{1,1}}{\partial\phi} + \frac{\partial G^j_{1,2}}{\partial\theta} - \frac{G^j_{1,2}}{\tan\theta}\right)\sin(2\gamma)\right\} \\ g_{kj} &= \frac{1}{R}\left\{\left(\csc\theta\frac{\partial G^j_{2,2}}{\partial\phi} + \frac{G^j_{2,1}}{\tan\theta} - \frac{\partial G^j_{2,1}}{\partial\theta}\right)\cos(2\gamma) + \left(\csc\theta\frac{\partial G^j_{2,1}}{\partial\phi} + \frac{\partial G^j_{2,2}}{\partial\theta} - \frac{G^j_{2,2}}{\tan\theta}\right)\sin(2\gamma)\right\} \end{aligned} \tag{25}$$

and $c_k = 0$ if we create a new rate estimate $r_k = \xi$. The same value of ξ is also used to set the standard deviation for this constraint as $\sigma_k = (0.83)\xi$, so that the Gaussian distribution (which my method forces me to use) will best approximate the desired Heaviside distribution near the origin.

Paleostress data are different from structural and paleomagnetic data because they are not integral constraints over time, but momentary samples. A thin igneous dike may form in a day. Therefore, it is necessary to distinguish between two types of published references on paleostress. The more desirable type summarizes many paleostress indicators of different ages to show that $\hat{\sigma}_{1\mathrm{h}}$ has remained constant over some time interval from age t_2 to age t_1. These data should be applied in each timestep of that interval ($t_1/\Delta t < n \le t_2/\Delta t$) with full relevance: $q_k^n = 1$. The less desirable type of reference concerns paleostress indicators whose age can only be constrained to be less than t_2 but more than t_1. These are relevant to one of the timesteps in the interval, but it is not clear to which. These should be applied in all timesteps of the interval, but with reduced weight due to their reduced (mean) relevance: $q_k^n = \inf\{1, \Delta t/(t_2 - t_1)\}$.

Ideally, very complete datasets on paleostress direction would impose a "smoothness" constraint on the velocity solutions, mimicking the smoothness that forward models have as a result of solving the momentum equation. In practice, there are rarely more than a dozen relevant paleostress indicators in any given timestep. A related problem is that if each paleostress indicator were only compared to the strain-rate tensor in a single finite element, then its influence on the solution would decrease as the finite element grid was refined. I attempt to solve both problems by interpolating paleostress directions (with associated uncertainties) for every finite element in the grid, based on the relevant paleostress data. The interpolation is by nonparametric statistics based on the spatial autocorrelation of the *present* stress field as represented by the World Stress Map [*Zoback*, 1992]. The interpolation method is given in *Bird and Li* [1996]; I use the simpler of their two methods in which the data are considered independent, because this permits the weighting of the original data by their relevance (q_k^n) values during the interpolation.

In cases where stress-direction data are sparse, it may be desirable or necessary to use active fault segments as additional indicators, assuming $\hat{\sigma}_{1\mathrm{h}}$ perpendicular to thrusts, *etc*. Only the first phase of movement on a fault should be used to indicate stress, because in later times the fault is an inherited plane of weakness.

Integration Over Time

I use the "predictor/corrector" method of time integration, as in earlier forward dynamic models of the history of North America [*Bird*, 1988; *Bird*, 1992]. Each time-step begins with an explicit “prediction” of new node locations. Using these, all nodal function derivatives, coefficients, and velocities are recomputed for the same timestep. Then, the “predicted” velocity for that time-step is “corrected” by adding one-half of the (vector) change between the solutions. The node locations are corrected accordingly. *Bird* [1989] presented studies of the accuracy of this method; for practical purposes, it is sufficient to use timesteps of 1~5 m.y..

Iterative Revision of Rate Histories

Because the events in geologic history that can be dated are not always those we would choose, many data about strain or displacement come with loose time windows, bracketing but not specifying the true duration of deformation. However, adjacent data with better constraints should cause strain-rates to rise in the correct period. If the model predicted rate for any datum is larger than the tentative goal rate, this is probably an indication that the goal rates for that datum should be revised to permit more rapid deformation in a portion of the time window.

My method is to assign new target rates $\left(r_k^n\right)^*$ (for datum k, in each timestep n), based on the predicted rates p_k^n from the previous model, but adjusted by a constant factor to achieve the correct total strain, displacement, or rotation. Assuming a sign convention such that all goal rates are positive,

$$\left(r_k^n\right)^* = \left[\frac{\sum_n r_k^n}{\sum_n \sup\left(p_k^n, 0\right)}\right] \sup\left(p_k^n, 0\right) \qquad (26)$$

where the asterisk on the left indicates new goals for the next iteration, and all other quantities are old values from the previous iteration. (The truncation of the actual rates at zero is used in order to forestall a possible instability in the computation in which the denominator might become very small because of a predicted history that includes an unanticipated sense reversal.)

There are data for which this formula is inadequate: those whose time window extends back before the earliest time considered in the palinspastic reconstruction. My method for such cases is based on the assumption that strain and displacement in the earlier (unmodeled) period had the same sign as the net strain or displacement in the time period of the model. Therefore, if the model predicts too much total strain in the modeled interval, the targets must be reduced using (26), in order to prevent implied earlier rates from switching sign. However, if the model predicts too little strain, then the targets are not adjusted, because strain of the same sense in the unmodeled period can make up the difference.

Another possible problem with method (26) is that in some cases it can cause numerical instability (*i.e.*, self-amplification of small rate variations). The method is stable when applied to data with a constructive interaction (e.g., two fault segments that together form a microplate boundary), neutrally stable for isolated data that interact only with the *a-priori* stiffness, and unstable for groups of data with a destructive interaction (e.g., two parallel thrust faults). In the unstable case, the repeated application of (26) leads to a history in which only one of the data has a non-zero rate in any particular timestep. This is undesirable because the details of their rate histories arise from the solution process and not from the data. To prevent this, I only apply (26) to those data (and in those iterations) where the predicted rate in at least one timestep exceeds the corresponding goal rate; this is a sign of a constructive interaction.

The entire computation is now repeated, beginning at the present. In many cases, data that have tight time constraints are able to pinpoint the time of an orogeny locally, but their influence is diluted by other data with broad time windows. Thus, the peak in strain-rate (as a function of time) in the computed history is initially modest. However, with iteration the tentative rates for the latter data will be adjusted to reflect a shorter, more intense event.

Idealized Test Cases

The program Restore (Appendix II) which realizes this algorithm has been successfully tested in the following cases:

1. With no data except the *a-priori* constraint, and with a few boundary nodes fixed, the grid remains static.

2. With no data except the *a-priori* constraint, and with velocity boundary conditions at two nodes that imply plate rotation about a local Euler pole, the grid rotates as a rigid plate with no deformation. During rotation at angular rate 10^{-15}/s, internal strain rates are less than 2×10^{-19}/s.

3. Rotation of this "rigid" plate through finite timesteps does not add significant error. For example, when the whole grid is rotated 60° in steps of 6° around a local Euler pole with the predictor/corrector method all fictitious strains are ≤0.8%, much less than the fictitious strains of 5.6% = $\left[\cos(6^\circ)\right]^{-10} - 1$ that would result from explicit time integration.

4. Regions with no data behave as isotropic sheets of incompressible viscous material when forced to deform. For example, when a domain of 27° longitude range and 10° latitude range lying along the equator is stretched E-W, the strain in the center can be described by

$\varepsilon_{\theta\theta} = \varepsilon_{rr} = -\varepsilon_{\phi\phi}/2$. When the uncertainty of the *a-priori* zero strain-rate (μ) is laterally heterogeneous, computed strain-rates ($\dot{\varepsilon}$) are higher in regions of higher μ. However, the scaling in this case is $\dot{\varepsilon} \sim \mu^2$, not $\dot{\varepsilon} \sim \mu$ as one might guess.

5. When a single strike-slip fault that is an arc of a small circle cuts across the domain from one side to the other, the resulting solution has relative rotation of two rigid plates about the pole of the small circle. When a single dip-slip fault that is an arc of a great circle cuts across the domain from one side to the other, the resulting solution has relative rotation of two rigid plates about a point 90° away on the great circle.

6. When an active fault terminates in the model interior there is a conflict between the datum showing fault activity and the *a-priori* assumption of no intraplate deformation. If the plate stiffness (W/μ^2) is small and the fault's slip-rate uncertainty (δs) is also small, then the reduction in slip rate is negligible. If both parameters are large, the fault may nearly be prevented from slipping.

A quantitative measure of this effect can be derived from the analytic solution of a simpler problem. Assume that an inverse model makes only one scalar prediction *(p)* which is to be compared with all data. These data include N_r entries showing a positive rate of *r* with uncertainty σ_r, and N_0 entries showing a rate of zero, with uncertainty σ_0. If we define a dimensionless parameter

$$\Gamma \equiv \frac{N_r \sigma_0^2}{N_0 \sigma_r^2} \tag{27}$$

then the result of a maximum-likelihood solution is

$$\frac{p}{r} = \frac{\Gamma}{\Gamma + 1}. \tag{28}$$

This case is more complicated because the model predicts a range of slip-rates along the fault trace and a range of continuum strain-rates in its neighborhood. However, the scaling is: $N_r \sim \sum_{\ell=1}^{L} \kappa_{\ell z}$ (the number of finite elements crossed by the fault); $N_0 \sim WI^2$ (where *I* is the length of the fault trace); $\sigma_r \sim \delta s$; and $\sigma_0 \cong \mu I$. Together these suggest a new dimensionless parameter

$$\Gamma_{\mathrm{f}} \equiv \tau \frac{\mu^2 \sum_{\ell=1}^{L} \kappa_{\ell z}}{(\delta s)^2 W} \tag{29}$$

in which τ is a dimensionless coefficient which depends on the type of faulting. Numerical tests with one isolated fault in the middle of a plate show that these equations give good predictions of mean slip-rate (within 5%) if the following coefficients are used: strike-slip $\tau_{\mathrm{s}} \cong 1.0$; thrust $\tau_{\mathrm{t}} \cong 0.2$; normal $\tau_{\mathrm{n}} \cong 5$; detachment $\tau_{\mathrm{d}} \cong 1.0$.

7. In order to test the iteration of the tectonic history with (26), I divided the domain into two plates with one great-circle strike-slip fault. Then, I artificially segmented the fault, assigning the same offset to each segment, but assigning different (overlapping) time windows for slip on each segment. With iteration, the algorithm slowly converged on the correct solution, in which the slip history is the same for all segments, with all slip taking place within the shortest of the time windows. If the assigned uncertainties in fault slip-rate are small and/or the continuum stiffness (W/μ^2) is low, then convergence is slow, because actual slip rates are only slightly different from goal rates, and hence the goals are only adjusted in small steps. Scaling suggests that the number of iterations required for convergence should be proportional to ($\Gamma_{\mathrm{f}} + 1$), and numerical trials confirm this. ***Thus, there is a trade-off between fitting the data closely and converging quickly on the best deformation history.*** A good computational strategy is to set the parameters for small values of Γ (≅1~3) in the early iterations, then to raise Γ in later iterations; such a method is similar to simulated annealing.

8. Balanced cross-section data yield the correct relative displacement of their endpoints when there is no conflict with other data. For example, when two cross-sections each span a long thrust fault with very uncertain offset (*e.g.*, σ = 1000 km), they determine its slip distribution and also determine the relative rotation of the rigid plates on each side.

9. If a single cross-section with nonzero extension is located in a plate interior away from any faults or other data, there is a conflict between the cross-section datum and the *a-priori* assumption of no intraplate deformation. I have found a dimensionless parameter for cross-sections,

$$\Gamma_{\mathrm{c}} \equiv \tau_{\mathrm{c}} \frac{\mu^2}{\sigma_{\mathrm{c}}^2 W} \tag{30}$$

in which σ_{c} is the uncertainty of the extension rate of the cross-section. Used with equation (28), this parameter gives a good prediction of model rates of cross-section extension when the non-dimensional factor $\tau_{\mathrm{c}} = 0.8$.

10. When several cross-sections showing equal shortening (or extension) are placed en-echelon in a band (like stitches), they overcome the *a-priori* constraint of plate rigidity and define a band of orogeny (or taphrogeny). In order for the far-field velocity difference to be similar to the shortening (extension) velocity along each cross-section, the spacing between cross-sections must be comparable to their individual lengths, or less.

11. When a microplate is completely free to move (*e.g.*, isolated from the boundary conditions by a small-circle strike-slip fault of unknown offset) and it contains a paleolatitude anomaly datum, the microplate moves toward or away from the paleopole in accordance with the datum.

12. When a paleomagnetic site with a paleolatitude anomaly occurs in a plate interior isolated from active faults or other data, there is a conflict with the *a-priori* assumption of no deformation, and the solution is a compromise. There is a dimensionless factor for paleolatitude anomalies

$$\Gamma_{\mathrm{p}} \equiv \tau_{\mathrm{p}} \frac{\mu^2 P}{\sigma_{\mathrm{p}}^2 W} \tag{31}$$

(where σ_{p} is the uncertainty in the mean N-S velocity in m/s, and $\tau_{\mathrm{p}} \cong 1$) which can be used with (28) to predict the result. For example, with typical values $P = 100$, $W = 1 \times 10^{-10}$ /m^2, and $\sigma_{\mathrm{p}} = 6 \times 10^{-10}$ m/s (5°/30 Ma), the N-S motion of the site will only exceed half its nominal value for $\mu > 5\times10^{-16}$ /s; this is a large value that might be assigned in British Columbia, but would not be appropriate in Wyoming.

13. When a plate containing a single paleomagnetic site indicating a vertical-axis rotation is constrained at only one node, it rotates at the correct rate without negligible internal deformation.

14. When an paleomagnetic site showing vertical-axis rotation occurs in a plate interior isolated from active faults or other data, there is a conflict with the *a-priori* assumption of no deformation, and the solution is a compromise. There is a dimensionless factor for rotation data

$$\Gamma_{\mathrm{r}} \equiv \tau_{\mathrm{r}} \frac{\mu^2 P}{\sigma_{\mathrm{r}}^2 W a} \tag{32}$$

(where σ_{r} is the uncertainty in the rotation rate in radians per second, a is the area of a typical finite element, and $\tau_{\mathrm{r}} \cong 0.2$) which can be used with (28) to predict the result.

15. If uniform stress-direction data are given for each element, in a problem where other data do not completely dictate the stress direction, then the velocity solution changes to honor the stress constraints. One problem tested had a rectangular plate fixed on its eastern side, and forced to deform by several paleomagnetic sites with equal latitude anomalies along its western edge. If the stress-direction was not constrained, the solution was a combination of dextral simple shear and clockwise rotation. If N-S $\sigma_{1\mathrm{h}}$ was specified, the solution changed to a combination of dextral simple shear and E-W extension, with little rotation. Three refinements of the velocity solution were enough for convergence.

Assignment of Uncertainties to Rate Estimates

Most structural and paleomagnetic data indicate that some strain or displacement (symbolized here by the generic ε) occurred after time t_2 and before time t_1. They do not imply that the rate was constant during that period. The computation begins by using the average rate as the rate estimate in each timestep (except, less at each end):

$$r_k^n = \frac{\varepsilon_k}{(t_2 - t_1)_k} \quad \frac{\inf\begin{cases} \Delta t, \\ n\,\Delta t - (t_1)_k, \\ (t_2)_k - (n-1)\Delta t \end{cases}}{\Delta t}. \tag{33}$$

However, it would be inadequate to estimate the rate uncertainties σ_k^n by dividing the standard deviation of the deformation ($\delta\varepsilon_k$) by the time interval $(t_2 - t_1)_k$. That would confuse the uncertainty in the rate during one timestep with the uncertainty in the average rate. Also, it would have the perverse effect of making data less important as their geologic age constraints became tighter!

A related practical problem concerns the convergence of the iteration process by which we attempt to revise the rate histories (26). Fault offsets are often known with a very small uncertainties. If the offset uncertainty is the only contribution to σ_k^n, then the rate uncertainty is so small that the model prediction of the rate is virtually identical to the rate goal. Therefore, the adjustment of the rate history is unacceptably slow.

My solution is to begin the iteration process with a standard rate uncertainty $\sigma *$ assigned to each datum in a homogeneous class. This value is chosen so as to give Γ values of order unity for all data. This permits the data to interact, and permits adjustment of some rate histories in a reasonable number of iterations. Then, as the iteration continues, the rate uncertainties are gradually changed to values based only the uncertainties in the numerators, according to

$$\left(\sigma_k^n\right)_i = \sigma * \left[\frac{1}{\sigma *}\frac{\delta\varepsilon_k}{(t_2 - t_1)_k}\right]^{\frac{(i-1)}{(M-1)}} \tag{34}$$

(for all *n*, and for all *k* in the class) where *i* is the iteration number, up to a maximum of *M*. One way to think of this is that (most) rate uncertainties are initially increased arbitrarily to allow for the uncertainty in the time-history, or denominator. Once the time-history has been adjusted, they are smoothly returned to values based only on their numerators. Another way to say this is that the algorithm begins with a least-squares solution, passes through a phase of weighted-least-squares solutions, and finishes with a maximum-likelihood solution.