CAP Moment Tensor Inversion

Vipul Silwal

Geophysical Institute

University of Alaksa

Fairbanks, Alaska

Last compiled: December 3, 2013

Contents

1   *Zhao and Helmberger* (1994)

1.1   time shift

Cross-correlation function The relative time shift of the data f(t) from the synthetics g(t) can be

estimated with the use of the cross-corrleation function:

$$C(t) = \frac{\int_{-\infty}^{\infty} f(\tau)g(t+\tau)d\tau}{(\int_{-\infty}^{\infty} f^2(\tau)d\tau \int_{-\infty}^{\infty} g^2(\tau)d\tau)^{1/2}} \tag{1}$$

This is the normalized form of $C(t)$.

1.2 Seismic Moment

The moement is the ratio of the peak of the amplitude of the data to that of the synthetics

$$M_0 = \frac{Max(|f(t)|)}{Max(|g(t)|)} \tag{2}$$

1.3 Misfit

The error estimation in this earlier version of code involved both L1 and L2 norm.

$$L1 : ||f||_1^1 = \int_{t1}^{t2} |f(t)|dt \tag{3}$$

$$L2 : ||f||_2^2 = \int_{t1}^{t2} f^2(t)dt \tag{4}$$

un which [t1, t2] is the time interval, in which the seismogram is used. **The L1 norm emphasizes the high frequencies of the data, whereas the L2 emphasizes the low frequency**. Using equation 2, the synthetics is defined as:

$$d(t) = M_0 g(t) \tag{5}$$

The error is defined as (using either L1 or L2 norm):

$$e_{Lk} = \frac{||f - d||_k}{(||f||_k \times ||d||_k)^{1/2}} \tag{6}$$

Finally the error are implemented in the following manner:

$$e_1 = \frac{(e_{L1} + e_{L2} + (2e_{L1}^2 + 2e_{L2}^2)^{1/2})}{4} \tag{7}$$

For one station e_2 is defined the same way as e_1, except that the M_0 used in the equation 5 is the average moment $(1/n \sum_1^n M_0)$ of all the components used from the station. e_2 is a measure of the consistency of all components of one station, i.e. a minimum of e_2 gives closest ratios of the different components, say $SH/P_n l$ of the synthetics to data. **In short, e_1 emphasizes the fit of the individual components, whereas e_2 emphasizes the consistence of all the components of one station**.

In the equation 7 equal weightage have been given to the low frequency and high frequency components of the data.

The best solution is obtained by conducting the grid search in strike, dip and rake space, and finding the minimum of:

$$SOL : \min(E) = \frac{1}{n} \sum_{i=1}^{n} (e_{1i} + e_{2i}) \tag{8}$$

2 *Zhu and Helmberger* **(1996)**

The equations used in this paper are implemented in CAP.

2.1 Misfit

We define an object function to measure the misfit error between u and s and search through the parameter space to find the global minimum of the object function. Misfit error is defined as the norm-k (L1 or L2) of the difference between u and s normalized by the norms of both u and s.

$$e = \frac{||u - s||_k}{||u||_k \times ||s||_k} \tag{9}$$

Because Pnl usually has smaller amplitude than surface waves, this normalization helps to weight Pnl and surface waves equally. It also prevents the inversion from being completely dominated by the strongest station, which is usually the nearest station, if several stations at different distance ranges are used. However, the amplitude information is lost during the normalization. Some of this information, such as amplitude ratios of Pnl-to-surface waves and SV-to-SH, provide important constraints on the source orientation and depth. A more severe problem with this normalization is that it introduces singularities in the source parameter space at those points where source orientation generates nodal synthetics (where the norm of synthetics vanishes). In the case when the data include nodal records, the grid search will miss the true minimum. The misfit error using true amplitudes without normalization:

$$e = ||u - s||_k \tag{10}$$

2.2 Distance scaling factor

Using true-amplitude waveforms for source inversion usually leads to the problem of the closest station dominating the inversion when stations are distributed over a large distance range. The misfits of surface waves have larger scatter than body waves, which are expected because surface waves are more easily affected by shallow heterogeneity. It has been shown that Pnl at a range of 300 to 1000 km is quite stable (Helmberger and Engen, 1980) and easily inverted for source mechanisms (Wallace and Helmberger, 1982). At closer range, the details of the Moho transition plays a more important role as well as the PL waves trapped in the shallow crust (Song and Helmberger, 1996). Since both of these features show strong local variation, we should expect the large scatter. The misfit errors show a rapid decay with distance. Since radiation patterns have been taken out, this decay is related to the amplitude decay due to geometrical spreading and attenuation. To compensate for this decay, we introduce a distance range scaling factor and define the misfit error for a record at a distance r as:

$$\boxed{e = \left(\frac{r}{r_0}\right)^p ||u - s||_k} \tag{11}$$

here p is a scaling factor to give the record at r the same weight as that at reference distance $r0 = 100$. Any kind of norm-k **L1 or L2** can be chosen for calculating the data misfit.

If we assume a spherical geometrical spreading for body waves and cylindrical geometrical spreading for surface waves, an appropriate choice of p would be p=1 for body waves and p=0.5 for surface waves.

3 Implementation: CAP

The same misfit function as equation 11 is used and the **L2 norm** is chosen.

3.1 Distance weighting factor

First the distance weighting factor is computed

$$\mu = \left(\frac{r}{r_0}\right)^p \tag{12}$$

where r is the diatance and r_0 is the reference distance $= 100$. Usually p=1 for body waves and p=0.5 for surface waves.

$$\mu_{pnl} = \left(\frac{r}{r_0}\right)^1 \tag{13}$$

$$\mu_{surf} = \left(\frac{r}{r_0}\right)^{0.5} \tag{14}$$

At this point body and surface wave weight are imposed to evaluate the overall weighting factor

$$w_{pnl} = (\mu_{pnl}) \times weight_{pnl} \tag{15}$$

$$w_{surf} = (\mu_{surf}) \times weight_{surf} \tag{16}$$

P.S. This weight is given in the input cap command and not the weight file.

3.2 L2 norm of data and green's function

Before computing the norm, both the data and the green's function are subjected to the weighting factor:

$$u(t)_i = w_i \times [f(t)_i]; \quad i = [Pnl, Surf] \tag{17}$$

$$s(t)_i = w_i \times [g(t)_i]; \quad i = [Pnl, Surf] \tag{18}$$

here f(t) is the recorded seismogram and g(t) is the green's function. Individual components of the data and green's function are picked and then multipled with the suitable weighting factor depending it is the Pnl of Surface wave section. Each component (R,T,Z) of seismograms is choped into Pnl and Surface wave setion.

Example for a particular station:

$$[u(t)_{pnl}]_R = w_{pnl} \times [f(t)_{pnl}]_Z \tag{19}$$

$$[u(t)_{pnl}]_Z = w_{pnl} \times [f(t)_{pnl}]_Z \tag{20}$$

$$[u(t)_{surf}]_R = w_{surf} \times [f(t)_{surf}]_R \tag{21}$$

$$[u(t)_{surf}]_Z = w_{surf} \times [f(t)_{surf}]_Z \tag{22}$$

$$[u(t)_{surf}]_T = w_{surf} \times [f(t)_{surf}]_T \tag{23}$$

3.2.1 L2 norm

$$L2 : ||u||^2 = \int_{t1}^{t2} u^2(t)dt \tag{24}$$

$$L2 : ||s||^2 = \int_{t1}^{t2} s^2(t)dt \tag{25}$$

For a vector signal, it is the inner product of the vector with itself.

$$L2 : ||\mathbf{u}||^2 = (\underline{\mathbf{u}}^T)(\underline{\mathbf{u}}) \tag{26}$$

$$L2 : ||\mathbf{s}||^2 = (\underline{\mathbf{s}}^T)(\underline{\mathbf{s}}) \tag{27}$$

3.3 Cross Correlation

The normalized cross correction function between u and s are computed using the same as equation 1.

$$C(t) = \frac{\int_{-\infty}^{\infty} u(\xi)s(\xi - t)d\xi}{(||u||^2 \, ||s||^2)^{1/2}} \tag{28}$$

$$= \frac{(u \star s)(t)}{(||u||^2 \, ||s||^2)^{1/2}} \tag{29}$$

This correlation function is the correlation between data and green's function. This is the normalized correlation function (or the **correlation coeffcient** whose value is between 0 and 1. For find the actual actual correlation between two function (unnormalized):

$$C(t) = (u \star s)(t) \tag{30}$$

$$= \int_{-\infty}^{\infty} u(\xi)s(\xi - t)d\xi \tag{31}$$

This is like convolution in some sense. To find the correlation of two functions, one function is slided over the another function. At some particular value of time τ_{max} this function $C(t)$ will have it maximum. This maximum correlation point is what we are interested inn matching the wave forms. And the shift τ_{max} is the relative shift between synthetics and data.

$$C_{max} = C(\tau_{max}) = \max\left[corr(u(t), s(t))\right] \tag{32}$$

$$= \max\left[(u \star s)(t)\right] \tag{33}$$

$$= \max\left[\int_{-\infty}^{\infty} u(\xi)s(\xi - t)d\xi\right] \tag{34}$$

$$= \int_{-\infty}^{\infty} u(\xi)s(\xi - \tau_{max})d\xi \tag{35}$$

$$= \max\left[(u \star s)(\tau_{max})\right] \tag{36}$$

Thus for the maximum correlation of two functions $u(t)$ and $s(t)$, we are multiplying the functions are shifting the synthetics by τ_{max} and then finding area under the curve $u(t)s(t + \tau_{max})$.

3.4 Moment magnitude

M_0, the is seismic moment $M_0 = \mu A D$, where μ is the shear modulus, A is the area of rupture and D is the average displacement. Relationship between moment magnitude M_w and seismic moment M_0 is given by *Kanamori* (1977) relation,:

$$M_w = \frac{2}{3}(\log_{10} M_0) + k' \tag{37}$$

Actual numbers are cited in Hanks and Kanamori 1979 GCMT catlaog, *Aki and Richards* (1980), and *Shearer* (1999) uses the following equation:

$$M_w = \frac{2}{3}(\log_{10} M_0) - \frac{2}{3}(16.1) \tag{38}$$

However in the actual code, CAP uses the following equation to estimate the scaled seismic moment A from moment magnitude M_w :

$$M_w = \frac{2}{3}(\log_{10}(A)) + \frac{2}{3}(20 - 16.1) \tag{39}$$

$$A = 10^{(1.5M_w + 16.1 - 20)} \tag{40}$$

To relate M_0 and A:

$$M_w = \frac{2}{3}(\log_{10}(A)) + \frac{2}{3}(20 - 16.1) \tag{41}$$

$$= \frac{2}{3}(\log_{10}(A) + 20) - \frac{2}{3}16.1 \tag{42}$$

Therefore $\tag{43}$

$$(\log_{10}(A) + 20) = \log_{10} M_0 \tag{44}$$

$$\log_{10}\frac{M_0}{A} = 20 \tag{45}$$

$$\frac{M_0}{A} = 10^{20} \tag{46}$$

$$A = \frac{M_0}{10^{20}} \tag{47}$$

P.S. Instead of M_0, A will be used to obtain the scaled moment tensor elements.

3.5   Fault plane solution to Moment tensor

$$M_{xx} = -(\sin\delta \, \cos\lambda \, \sin 2\Phi + \sin 2\delta \, \sin\lambda \, \sin^2\Phi) \tag{48}$$

$$M_{yy} = (\sin\delta \, \cos\lambda \, \sin 2\Phi - \sin 2\delta \, \sin\lambda \, \cos^2\Phi) \tag{49}$$

$$M_{zz} = (\sin 2\delta \, \cos\lambda) \tag{50}$$

$$M_{xy} = (\sin\delta \, \cos\lambda \, \sin 2\Phi + 0.5\sin 2\delta \, \sin\lambda \, \sin 2\Phi) \tag{51}$$

$$M_{xz} = -(\cos\delta \, \cos\lambda \, \cos\Phi + \cos 2\delta \, \sin\lambda \, \sin\Phi) \tag{52}$$

$$M_{yz} = -(\cos\delta \, \cos\lambda \, \sin\Phi - \cos 2\delta \, \sin\lambda \, \cos\Phi) \tag{53}$$

Φ, strike of the fault

δ, dip angle of the fault plane

λ, slip angle

3.6   Green's function to Synthetics

When conducting the grid search over strike, dip and rake, the moment tensor is evaluated. Using this moment tensor and the azimuth of the station, the radiation pattern (**horizontal radiation coefficient**) are found. See *Jost and Herrmann* (1989).

$$A1 = -\frac{1}{2}(M_{xx} - M_{yy})\cos(2az) + M_{xy}\sin(2az) \tag{54}$$

$$A2 = -M_{xz}\cos(az) - M_{yz}\sin(az) \tag{55}$$

$$A3 = \frac{2M_{zz} - M_{yy} - M_{xx}}{6} \tag{56}$$

$$A4 = -\frac{1}{2}(M_{xx} - M_{yy})\sin(2az) + M_{xy}\cos(2az) \tag{57}$$

$$A5 = -M_{yz}\sin(az) + M_{xz}\sin(az) \tag{58}$$

$$A6 = 0 \tag{59}$$

$$A7 = \frac{M_{zz} + M_{yy} + M_{xx}}{3} \tag{60}$$

$$A8 = 0 \tag{61}$$

6

To find the actual synthetics, these need to be multiplied by the scaled seismic moment A.

$\mathbf{A} = A * [A1\ A2\ A3\ A4\ A5\ A6\ A7\ A8]$.

The synthetics can be computed using the 9 green's functions and these 6 radiation coffiients:

$$d_z = ZSS\ A1 + ZDS\ A2 + ZDD\ A3 + ZEP\ A7 \tag{62}$$

$$d_r = RSS\ A1 + RDS\ A2 + RDD\ A3 + REP\ A7 \tag{63}$$

$$d_t = TSS\ A4 + TDS\ A5 + TDD\ A6 + TEP\ A8 \tag{64}$$

where R,T,Z are radial, transverse and vertical components of SS(vertical strike-slip), DS(vertical dip-slip) and DD(45° dip-slip).

P.S. No Transverse components from vertical dip-slip fault and explosive source (TDD and TEP are 0)

3.6.1 time shift

However instead of using the green's function the correlation function $C(t)$ is used. Correlation between the data and corresponding green's function. (i.e. instead of green's func ZSS, the correlation of ZSS and vertical component of data is used.

$$v_z = corr(ZSS, u_z)\ A1 + corr(ZDS, u_z)\ A2 + corr(ZDD, u_z)\ A3 + corr(ZEP, u_z)\ A7 \tag{65}$$

$$v_r = corr(RSS, u_r)\ A1 + corr(RDS, u_r)\ A2 + corr(RDD, u_r)\ A3 + corr(REP, u_r)\ A7 \tag{66}$$

$$v_t = corr(TSS, u_t)\ A4 + corr(TDS, u_t)\ A5 + corr(TDD, u_t)\ A6 + corr(TEP, u_t)\ A8 \tag{67}$$

where v_z, v_r, v_t are correlation of data and synthetics for a particular component. The value of time for which this correlation value comes out to be maximum, is the corresponding time_shift(τ) for that component.

Pnl wave (time_shift $= \tau 1$)

$$P_z = corr(ZSS, u_z)\ A1 + corr(ZDS, u_z)\ A2 + corr(ZDD, u_z)\ A3 + corr(ZEP, u_z)\ A7 \tag{68}$$

$$P_r = corr(RSS, u_r)\ A1 + corr(RDS, u_r)\ A2 + corr(RDD, u_r)\ A3 + corr(REP, u_r)\ A7 \tag{69}$$

$$P_{max} = \max(w1 \cdot P_z + w2 \cdot P_r)_{\tau 1} \tag{70}$$

Rayleigh wave (time_shift $= \tau 2$)

$$S_z = corr(ZSS, u_z)\ A1 + corr(ZDS, u_z)\ A2 + corr(ZDD, u_z)\ A3 + corr(ZEP, u_z)\ A7 \tag{71}$$

$$S_r = corr(RSS, u_r)\ A1 + corr(RDS, u_r)\ A2 + corr(RDD, u_r)\ A3 + corr(REP, u_r)\ A7 \tag{72}$$

$$S_{max} = \max(w3 \cdot S_z + w4 \cdot S_r)_{\tau 2} \tag{73}$$

SH wave (time_shift $= \tau 3$)

$$L_t = corr(TSS, u_t)\ A4 + corr(TDS, u_t)\ A5 \tag{74}$$

$$L_{max} = \max(w5 \cdot L_t)_{\tau 3} \tag{75}$$

where w1 to w5 are the weights for particular component (as specified in the weight file).

3.6.2 New synthetics

After finding the point of maximum correlation, and shifting the synthetics, we obtain new synthetics of the form:

$$s'(t) = s(t - \tau) \tag{76}$$

The negaative sign is used because poistive time shift means, synthetics is earlier and needs to be shifted in positive t direction. In vector form :

$$\text{old synthetics}: \quad s(t) \quad = \mathbf{s} \tag{77}$$

$$\text{new synthetics}: \quad s'(t) \quad = \mathbf{s}' \tag{78}$$

$$\text{relationship}: \quad s'(t) \quad = s(t - \tau) \tag{79}$$

3.7 L2 norm synthetics

Compute synthetics as in equation 64 and find the L2 norm using the same formula as equation 25.

$$s_z = ZSS\ A1 + ZDS\ A2 + ZDD\ A3 + ZEP\ A7 \tag{80}$$

$$s_r = RSS\ A1 + RDS\ A2 + RDD\ A3 + REP\ A7 \tag{81}$$

$$s_t = TSS\ A4 + TDS\ A5 \tag{82}$$

L2 norm

$$||s||^2 = \int_{t1}^{t2} s^2(t)dt \tag{83}$$

$$= \int_{t1}^{t2} s'^2(t)dt \tag{84}$$

$$= \int_{t1+\tau}^{t2+\tau} s^2(t - \tau)dt \tag{85}$$

$$= (\mathbf{s}^T)(\mathbf{s}) \tag{86}$$

$$= (\mathbf{s}'^T)(\mathbf{s}') \tag{87}$$

3.8 Correlation between data and synthetics

The correlation between data $u(t)$ and synthetics $s'(t) = s(t - \tau)$ (new syntheics shifted by τ which gives maximum correlation), can be found by using equation 36

$$corr(u(t), s(t))_{max} = C_{max} = C(\tau_{max}) = \int_{t1}^{t2} u(t)s(t - \tau_{max})dt \tag{88}$$

$$= \int_{t1}^{t2} u(t)s'(t)dt \tag{89}$$

$$= (u \star s)(\tau_{max}) \tag{90}$$

3.9 Misfit

Finally the misfit e is given by:

$$\text{Integral form: } e = \int_{t1}^{t2} u^2(t)dt + \int_{t1}^{t2} s^2(t)dt - 2\int_{t1}^{t2} u(t)s(t - \tau_{max})dt \tag{91}$$

replacing $s(t)$ by shifted synthetics $s'(t)$

$$e = \int_{t1}^{t2} u^2(t)dt + \int_{t1}^{t2} s'^2(t)dt - 2\int_{t1}^{t2} u(t)s'(t)dt \tag{92}$$

$$= \int_{t1}^{t2} (u^2(t) + s'^2(t) - 2u(t)s'(t))dt \tag{93}$$

$$= \int_{t1}^{t2} (u(t) - s'(t))^2 dt \tag{94}$$

$$\text{Vector form: } e = (\underline{u}^T)(\underline{u}) + (\underline{s}'^T)(\underline{s}') - 2(\underline{u}^T)(\underline{s}') \tag{95}$$

$$= (\underline{u} - \underline{s}')^T(\underline{u} - \underline{s}') \tag{96}$$

This is the misfit error for a particular component of a particular station.

$u(t) \equiv \mathbf{u}$, is the recorded data (Pnl or Surface wave window)

$s(t) \equiv \mathbf{s}$, is the synthetic seismograms (Pnl or Surface wave window)

$s'(t-\tau) \equiv \mathbf{s}'$, is the synthetic seismograms shifted by τ to maximum correlation point (Pnl or Surface wave window)

For all N stations and all components:

$$E = \sum_{j=1}^{N}\sum_{i=1}^{5} e_{ij} = \sum_{j=1}^{N}\sum_{i=1}^{5} w_{ij}(\underline{u}_{ij} - \underline{s}'_{ij})^T(\underline{u}_{ij} - \underline{s}'_{ij}) \tag{97}$$

$$= \sum_{j=1}^{N}\sum_{i=1}^{5}(\underline{u}_{ij} - \underline{s}'_{ij})^T\mathbf{W}_{ij}(\underline{u}_{ij} - \underline{s}'_{ij}) \tag{98}$$

where w_{ij} is the weight for i^{th} component at the jth station. Since our data $\mathbf{u}_{ij}$ and synthetics $\mathbf{s}'_{ij}$ are vectors, $\mathbf{W}_{ij} = \mathbf{I}\,w_{ij}$. Dimensions of $\mathbf{I}$ are number of sample points in the corresponding time window.
The misfit is **not** normalized over number of stations or components.
This objective function E is minimized. The another way of looking at this is:

$$SOL : \min(E) = \min(||u||^2 + ||s||^2 - 2\,corr(u(t), s(t))) \tag{99}$$

$$= ||u||^2 + ||s||^2 - \max(2\,corr(u(t), s(t))) \tag{100}$$

$$= ||u||^2 + ||s||^2 - 2\max[(u \star s)(t)] \tag{101}$$

$$= ||u||^2 + ||s||^2 - 2[(u \star s)(\tau_{max})] \tag{102}$$

3.10   Correlation percentage

Cross-correlation coefficent are computed using the normalized form. For a particular component, the correlation percentage is computed as follows:

$$cp_i = 100 * \left(\frac{\max[corr(u(t), s(t))]}{\sqrt{||u||^2\,||s||^2}} \right) \tag{103}$$

$$= 100 * \left(\frac{(u \star s)(\tau_{max})}{\sqrt{||u||^2\,||s||^2}} \right) \tag{104}$$

could be simplified to:

$$cfg = 100 * \left(\frac{||u - s||^2}{\sqrt{||u||^2\,||s||^2}} \right) \tag{105}$$

3.11 Variance reduction

Variance reduction is defined as improvement in the solution from some standard reference. In this case, the standard reference is when synthetics $s(t) = 0$ (no valid reason for assuming this). At $s(t) = 0$, the Variance Reduction $VR = 0$. And if our synthetics $s(t)$ perfectly matches the data $u(t)$, i.e. $s(t) = u(t)$, the $VR = 100$.

$$VR = 100 * \left(1 - \frac{E}{||u||^2}\right) \tag{106}$$

$$= 100 * \left(1 - \frac{||u - s||^2}{||u||^2}\right) \tag{107}$$

Dreger uses different Variance reduction formula,

$$VR = 100 * \left(1 - \frac{\sqrt{||u - s||^2}}{\sqrt{||u||^2}}\right) \tag{108}$$

For an acceptable solution, variance reduction is usually greater than 70%.

3.12 Depth Test

To obtain the best depth solution, inversion is performed at differnet depths and the minimum is found. To further get a better estimate, the minimum of the best fittng parabola is used.

However, for error-depth plot, a different measure is taken instead of misfit error, and then plotted against depth.

$$E\% = f \times \left(\frac{E_d}{E_{min}} - 1\right) \tag{109}$$

$E\%$, quantifying the error of depth d w.r.t. to error at best depth

f, degree of freedom where $f = n(N_{samp})$, where n is number of freedom per sample and N_{samp} is the total number of sampling points used for inversion (All stations body and surface waves).

E_d, error at depth d

E_{min}, error at best depth $E_{min} = \min(E_d)$

The default value of degree of freedom per sample $n = 0.01$. Changing this by order of 10, we can also control the steepness of the depth curve.

4 Changes

4.1 Changes in grid search

For homogeneous parameterization over dip, δ, and isotropic parameter, ζ, (*Tape and Tape*, 2012a), we sample the parameters homogeneously in their cosine space, and then use the arccosine of those equally spaced value. Subsequently, it appears that spacing is larger near equator ($\delta=0$ and densely sampled near the pole ($\delta=90$).

δ_0, be the first element of search range in dip

δ_d, be the search increment

N, number of samples to be generated

$\delta_1 = \delta_0 + (N-1)\delta_d$, would be the final element of search range in dip

Then

$$\delta_i = \cos^{-1}\left(\cos(\delta_0) - i\left[\frac{\cos(\delta_0 - \cos(\delta_1))}{N}\right]\right) \tag{110}$$

Similarly for Isotropic components:

$$\zeta_i = \sin^{-1}\left(\sin(\zeta_0) + i\left[\frac{\sin(\zeta_1 - \sin(\zeta_0))}{N}\right]\right) \tag{111}$$

4.2   Changes in misfit function

1. Normalized the misfit error by number of components used. Error may now actually decrease by addition of 'good' stations. **normalization method**

$$||E_1(\mathbf{m})||_2^2 = \frac{1}{(P_{comp} + S_{comp})}\left[\sum_{i=1}^{P_{comp}} \frac{\int W_i ||(\mathbf{u}_p)_i - (\mathbf{s}_p)_i||_2^2 dt}{(N_p)_i} + \sum_{i=1}^{S_{comp}} \frac{\int W_i ||(\mathbf{u}_s)_i - (\mathbf{s}_s)_i||_2^2 dt}{(N_s)_i}\right]$$
$$\tag{112}$$

$\mathbf{m}$, Model parameters (Strike Φ,Dip δ,Rake λ)

u_p, observed P waveform

P_{comp}, number of P components used

$(N_p)_i$, number of sampling points for P (may vary from station to station)

u_s, observed S waveform

S_{comp}, number of S components used

$(N_s)_i$, number of sampling points for S (may vary from station to station)

W_i, weight given to that particular waveform

This equation can be simplified to following form:

$$||E_1(\mathbf{m})||_2^2 = \frac{\sum_{i=1}^{N_{comp}}(\mathbf{u}_i - \mathbf{s}_i)^t(\frac{W_i}{N_i})(\mathbf{u}_i - \mathbf{s}_i)}{N_{comp}} \tag{113}$$

where,

$\mathbf{u}_i$, observed waveform (body or surface wave)

$\mathbf{s}_i$, synthetic waveform (body or surface wave)

N_{comp}, number of components used (all body and surface waveforms)

N_i, number of sampling points for the waveform (may vary from station to station)

W_i, weight given to that particular waveform

Normalized misfit could be written as:

$$||E_r(\mathbf{m})||_2^2 = \frac{\sum_{i=1}^{N_{comp}}(\mathbf{u}_i - \mathbf{s}_i)^t(\frac{W_i}{N_i})(\mathbf{u}_i - \mathbf{s}_i)}{\sum_{i=1}^{N_{comp}}(\mathbf{u})^t(\frac{W_i}{N_i})(\mathbf{u})} \tag{114}$$

This is useful when we are comparing across different events. Other common names in the literature for this normalized or reduced misfit is 'reduced chi-square'. χ^2 could be converted to variance reduction (which have many different formulas in literature).

$$
\begin{align}
V.R. &= (1 - ||E_r(\mathbf{m})||_2^2) \times 100 \tag{115}\\
&= (1 - ||E_r(\mathbf{m})||_2) \times 100 \tag{116}\\
&= \ln\left(\frac{||E_1(\mathbf{m})||_2^2}{||d||_2^2}\right) \times 100 \tag{117}\\
&= \ln\left(\frac{||E_1(\mathbf{m})||_2}{||d||_2}\right) \times 100 \tag{118}
\end{align}
$$

Since variation reduction is only good for comparing

2. If higher weight is given to a good station whose waveform is matcing well (higher correlation value), then overall misfit value reduces. vice versa, if higher weight is given to a bad station whose waveform is not matcing well (low correlation value), then overall misfit value increases.

5   Creating posterior samples

Our misfit function needs to be modified because of two following reasons:

1. Low misfit value
   This basically is because of small amplitudes in displacement and velocity field of seismograms, hence also in L2 norm of data, synthetics and misfit. One common way of overcoming this is by normalizing it with data norm, but this leads to giving high weight in case of nodal stations (*Zhu and Helmberger*, 1996). Another way to get around this is dividing the misfit function by its minimum value and measuring its logarithmic variation.

$$
\ln\left(\frac{E}{E_{min}}\right) \tag{119}
$$

2. Another issue is with misfit function being too smooth. Possible reasons for this could be :

   (a) Not using the proper parameterization for moment tensor inversion (*Tape and Tape*, 2012a).
   (b) Smooth misfit function also arises because of using exact magnitude M_w and not using full 6-parameter space for generating posterior samples. Our misfit function has only 3 model parameter (strike,dip,rake).

   To overcome this we scale the misfit by:

$$
\frac{T}{\ln\left(\frac{E_{max}}{E_{min}}\right)} \tag{120}
$$

Our final misfit function which is used for creating posterior pdf is:

$$
E_2(\mathbf{m}) = E(\phi, \delta, \lambda) = \frac{\ln\left(\frac{E}{E_{min}}\right)}{\ln\left(\frac{E_{max}}{E_{min}}\right)} \times T \tag{121}
$$

where T is any multiplication factor. This T could be thought of as temparature in the energy function for *simulated annealing*.

5.1 Misfit to Posterior pdf

The scaled logarithmic variation Eq.121 of misfit could be used to get the posterior probability of solution in model space.

$$\sigma(\mathbf{m}) = \frac{\exp(-E_2(\mathbf{m}))}{k} \tag{122}$$

where $k = \int \exp(-E_2(\mathbf{m}) \, d\mathbf{m}$ is the normalization factor and $\mathbf{m}$ are our model parameters (ϕ, δ, λ). $\boldsymbol{\sigma}(\mathbf{m})$ is the posterior probability density function. However, this is does not include any knowledge of the *shape* of the space in itself, i.e., the information about the homoegenous probability of model space parameters. (*Tarantola*, 2006). To get 'the' solution, homogenous state of information $(\boldsymbol{\mu}(\mathbf{m}))$ must be introduced in order to achieve the actual likelihood of solution. This could be done by:

$$L(\mathbf{m}) = \frac{\boldsymbol{\sigma}(\mathbf{m})}{\boldsymbol{\mu}(\mathbf{m})} \tag{123}$$

In case of double couple, homogeneous state of information $\boldsymbol{\mu}(\mathbf{m}) = \boldsymbol{\mu}\phi, \delta, \lambda$ is giben by:

$$\mu(\phi) = c_1 \tag{124}$$
$$\mu(\delta) = \cos^{-1} c_2 \tag{125}$$
$$\mu(\lambda) = c_3 \tag{126}$$

where c_1, c_2 and c_3 are constants. The importance of likelihood function is that it is the ratio of probabilites (not the probability in itself) which makes it invariant of transformation of variables. A property similar to volumetric proability $V(\mathbf{m})$.

For a double couple source, the shape of beachball always remains the same. What changes is the orientation of it. Keeping this in mind, it would be more sensible to estimate the uncertainity in terms of rotation of beachball from its minimum misfit (maximum likelihood) orientation rather that separately evaluating the uncertainities for different parameters (whether strike-dip-rake or moment-tensors elements). Also we can't do this in strike-dip-rake space because of the complications of clearly defining the domain. Parameterization would definately have a great impact on uncertainty estimation. Here even though we are only using one roatation angle (minimum), in reality we should using the whole rotation matrix in order to clearly understand and quantify the misfit variation in parameter space.

We transform our parameter space from (ϕ, δ, λ) to $(\omega$ or $\xi)$. This

Let the orientation of minimum misfit solution be $\mathbf{m_0} = \phi_0, \delta_0, \lambda_0$. Distance between all the samples, $\mathbf{m}$, and minimum misfit solution $\mathbf{m_0}$ can be written as:

$$[\mathbf{m}, \mathbf{m_0}] = \boldsymbol{\xi} \tag{127}$$

This is an example of surjective mapping (not injective). There are four different angles in quaternion space that make this rotation possible (*Kagan* (1991), *Tape and Tape* (2012b)). If we only take the minimum angle of rotation required for this transformation, the mapping becomes injective.

Using this transformation, we can find $\sigma'(\boldsymbol{\xi})$ from $\sigma(\mathbf{m})$. However, they are not the same (due the transformation of variables). What still remains same is the likelihood function. Mathematically, we can write

$$\sigma(\mathbf{m}) \neq \sigma'(\boldsymbol{\xi}) \qquad \left[\because \sigma'(\boldsymbol{\xi}) = \sigma(\mathbf{m}) \left| \frac{d\mathbf{m}}{d\boldsymbol{\xi}} \right| \right] \tag{128}$$

$$\frac{\sigma(\mathbf{m})}{\mu(\mathbf{m})} = \frac{\sigma'(\boldsymbol{\xi})}{\mu'(\boldsymbol{\xi})} \tag{129}$$

$$L(\mathbf{m}) = L'(\boldsymbol{\xi}) \tag{130}$$

References

Aki, K., and P. G. Richards, *Quantitative Seismology, Theory and Methods*, W. H. Freeman, San Francisco, Calif., USA, 1980.

Jost, M., and R. Herrmann, A students guide to and review of moment tensors, *Seismological Research Letters*, *60*, 3757, 1989.

Kagan, Y. Y., 3-D rotation of double-couple earthquake sources, *Geophysical Journal International*, *106*, 709–716, 1991.

Kanamori, H., The energy release in great earthquakes, *Journal of Geophysical Research*, *82*, 2981–2987, 1977.

Shearer, P. M., *Introduction to Seismology*, Cambridge U. Press, Cambridge, UK, 1999.

Tape, W., and C. Tape, A geometric setting for moment tensors, *Geophysical Journal International*, *190*, 476–498, 2012a.

Tape, W., and C. Tape, Angle between principal axes triples, *Geophysical Journal International*, *191*, 813–831, 2012b.

Tarantola, A., Popper, bayes and the inverse problem, *Nature Physics*, *2*, 492–494, 2006.

Zhao, L. S., and D. V. Helmberger, Source estimation from broadband regional seismograms, *Bulletin of the Seismological Society of America*, *84*, 91–104, 1994.

Zhu, L., and D. V. Helmberger, Advancement in source estimation techniques using broadband regional seismograms, *Bulletin of the Seismological Society of America*, *86*, 1634–1641, 1996.