The Trade-Off of Resolution and Variance Estimated from Ensembles of Solutions
Bill Menke, November 4, 2018

Recipe for a Dirichlet Trade-Off Curve

Step 1. Estimate the mean model $\mathbf{m}$ (say of length M) and its covariance $\mathbf{C}_m$ of the distribution $p\left(\mathbf{m}\middle|\mathbf{d}^{\text{obs}}\right)$, by taking the sample mean and sample covariance of the ensemble of solutions drawn from that distribution.

Step 2. Estimate the mean model $\mathbf{m}^{(\text{A})}$ (of length M) and its covariance $\mathbf{C}_m^{(\text{A})}$ of the distribution $p_A(\mathbf{m})$, by taking the sample mean and sample covariance of the ensemble of solutions drawn from that distribution.

Step 3. Calculate the eigenvalues $\lambda^{(i)}$ of $\mathbf{C}_m$ and sort them into increasing size.

Step 4. Let N be an integer in the range $1 \leq N \leq M$. For each value of N, calculate the Direchlet spread of resolution $J^{(N)}$ and the size of variance $K^{(N)}$:

$$J^{(N)} = M - N \quad \text{and} \quad K^{(N)} = \sum_{i=1}^{N} \lambda^{(i)}$$

Step 5. The trade-off curve is $K^{(N)}$ plotted against $J^{(N)}$.

Step 6. For comparison, calculate the size of the prior variance, $K^{(A)} = \text{tr}(\mathbf{C}_m^{\text{A}})$ and plot it, too.

Step 7. Identify the values of N for which $K^{(N)} > K^{(A)}$. They correspond to localized averages with variances controlled by the observed data $\mathbf{d}^{\text{obs}}$; the variances of the others are being controlled by the prior information.

Step 8: For any specific value of N, construct a diagonal matrix $\mathbf{\Lambda}^{(N)}$ containing the N smallest eigenvalues $\lambda^{(i)}$ of $\mathbf{C}_m$ and a matrix $\mathbf{V}^{(N)}$ containing the N corresponding eigenvectors. The resolution matrix is $\mathbf{R}^{(N)} = \mathbf{V}^{(N)}\mathbf{V}^{(N)T}$; the localized averages are $\mathbf{m}^{(N)} = \mathbf{R}^{(N)}\mathbf{m}$; and the covariance of the localized averages is $\mathbf{C}_m^{(N)} = \mathbf{V}^{(N)}\mathbf{\Lambda}^{(N)}\mathbf{V}^{(N)\text{T}}$.

Recipe for the Resolution of Deviation with Respect to the Prior Model

Step 1. Estimate, as before, the mean model $\mathbf{m}$ (say of length M) and covariance $\mathbf{C}_m$ of the distribution $p\left(\mathbf{m}\middle|\mathbf{d}^{\text{obs}}\right)$, by taking the sample mean and sample covariance of the ensemble of solutions drawn from that distribution.

Step 2. Estimate, as before, the mean model $\mathbf{m}^{(\text{A})}$ (of length M) and covariance $\mathbf{C}_m^{(\text{A})}$ of the distribution $p_A(\mathbf{m})$, by taking the sample mean and sample covariance of the ensemble of solutions drawn from that distribution.

Step 3: Calculate the resolution matrix $\mathbf{R}^{\text{G}}$ for deviations from the prior model as:

$$\mathbf{R}^{\text{G}} = \mathbf{I} - \mathbf{C}_m\left[\mathbf{C}_{\boldsymbol{m}}^{(A)}\right]^{-\mathbf{1}}$$

Derivation

1. Estimates of Covariance from Ensembles. Suppose one has a large number, say L, of solutions $\mathbf{m}^{(i)}$ (each of length M) that sample a posterior probability distribution $p(\mathbf{m}|\mathbf{d}^{obs}) = p(\mathbf{d}^{obs}|\mathbf{m})\, p_A(\mathbf{m})$ (constructed, say, using the Metropolis-Hastings algorithm). The model $\mathbf{m}$ and its covariance $\mathbf{C}_m$ can be estimated as the sample mean and covariance:

$$\mathbf{m} \approx \frac{1}{L}\sum_i \mathbf{m}^{(i)} \quad \text{and} \quad [\mathbf{C}^{est}]_{jk} \approx \frac{1}{L}\sum_i \left(m_j^{(i)} - \bar{m}_j\right)\left(m_k^{(i)} - \bar{m}_k\right) \tag{1}$$

Note that these quantities can be computed by accumulating various sums and products during the iterative Metropolis-Hastings process.

```
counts=0;
msum = zeros(M,1);
mprod = zeros(M,M);
for i=[1:L]
    % compute a member, m, of the ensemble here
    % where m is a vector of M model parameters
    counts = counts+1;
    msum = msum + m;
    mprod = mprod + m*m';
end
end

mmean = msum / counts;
Cm = zeros(M,M);
for ii = [1:M]
for jj = [1:M]
     Cm(ii,jj) = (mprod(ii,jj)-msum(ii)*mmean(jj)-
mmean(ii)*msum(jj)+counts*mmean(ii)*mmean(jj))/(counts-1);
end
end
```

Furthermore, if the Metropolis-Hastings process is repeat again to sample just the prior distribution $p_A(\mathbf{m})$, the prior model $\mathbf{m}^{(\mathrm{A})}$ and covariance $\mathbf{C}_m^{(\mathrm{A})}$ can be estimated in similar manner. However, if $p_A(\mathbf{m})$ is Normal and embodies *linear* prior information of the form $\mathbf{h} = \mathbf{H}\mathbf{m}$ (with covariance $\mathbf{C}_h$), then these quantities can be estimates more efficiently as $\mathbf{m}^{(A)} = [\mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathrm{h}}^{-1}\mathbf{H}]^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathrm{h}}^{-1}\mathbf{h}$ (with covariance $\mathbf{C}_m^{(\mathrm{A})} = [\mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathrm{h}}^{-1}\mathbf{H}]^{-1}$).

(2)

Linear Combinations of Model Parameters with Small Variance. Suppose that $\mathbf{C}_m$ has eigenvalue decomposition $\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\mathrm{T}}$. Here $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues $\lambda^{(i)}$ and $\mathbf{V}$ and is a matrix whose columns are orthonormal eigenvectors $\mathbf{v}^{(i)}$. The eigenvalues are all non-negative, because they represent variances. Now consider the "data" $\mathbf{d}$ derived from the model parameters via $\mathbf{d} = \mathbf{G}\mathbf{m}$ with $\mathbf{G} = \mathbf{V}^{\mathrm{T}}$. The covariance $\mathbf{C}_m$ of these data are uncorrelated with covariance $\boldsymbol{\Lambda}$:

$$\mathbf{C}_d = \mathbf{G}\mathbf{C}_m\mathbf{G}^{\mathrm{T}} = \mathbf{V}^{\mathrm{T}}\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\mathrm{T}}\mathbf{V} = \boldsymbol{\Lambda}$$

(3)

(since $\mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{I}$). Now suppose that we restrict ourselves to the $N \leq M$ eigenvalues that are equal to or below a threshold; that is $\lambda^{(i)} \leq \sigma^2$ and group them into an eigenvalue matrix $\mathbf{\Lambda}^{(N)}$**.** The corresponding eigenvectors $\mathbf{V}^{(N)}$ satisfy $\mathbf{V}^{(N)T}\mathbf{V}^{(N)} = \mathbf{I}$; however, in general, $\mathbf{V}^{(N)}\mathbf{V}^{(N)T} \neq \mathbf{I}$, except when $N = M$. The N data satisfy $\mathbf{d}^{(N)} = \mathbf{G}^{(N)}\mathbf{m}$, and all have variances $\sigma_{d_i}^2 \leq \sigma^2$. These data represent the linear combinations of the model parameters with the *acceptable* variances (in the sense of being at or below the threshold).

Dirichlet Resolution Case. Now suppose that we invert these data using the minimum length method:

$$\mathbf{m}^{(N)} = \mathbf{G}^{-g}\mathbf{d}^{(N)} \quad \text{with} \quad \mathbf{G}^{-g} \equiv \mathbf{G}^{(N)T}\left[\mathbf{G}^{(N)}\mathbf{G}^{(N)T}\right]^{-1} = \mathbf{V}^{(N)}$$

(4)

The last equality follows from $\mathbf{G}^{(N)}\mathbf{G}^{(N)T} = \mathbf{V}^{(N)T}\mathbf{V}^{(N)} = \mathbf{I}$. We construct a one-parameter family of solutions $\mathbf{m}^{(N)}$, together with their covariance $\mathbf{C}_m^{(N)}$ and resolution $\mathbf{R}^{(N)}$, by varying N in the range $1 \leq N \leq M$:

$$\mathbf{m}^{(N)} = \mathbf{G}^{-g}\mathbf{d}^{(N)} = \mathbf{V}^{(N)}\mathbf{V}^{(N)T}\mathbf{m} = \mathbf{R}^{(N)}\mathbf{m} \quad \text{with} \quad \mathbf{R}^{(N)} \equiv \mathbf{V}^{(N)}\mathbf{V}^{(N)T}$$

$$\mathbf{C}_m^{(N)} = \mathbf{G}^{-g}\mathbf{C}_d\mathbf{G}^{-g\mathrm{T}} = \mathbf{V}^{(N)}\mathbf{\Lambda}^{(N)}\mathbf{V}^{(N)\mathrm{T}}$$

(5)

The elements of $\mathbf{m}^{(N)}$ (for fixed N) can be understood as weighted averages of $\mathbf{m}$, where the weights are given the resolution matrix $\mathbf{R}$. The covariance of these weighted averages is given by $\mathbf{C}_m^{(N)}$. The $N = M$ case just returns the original model parameter and original covariance:

$$\mathbf{m}^{(M)} = \mathbf{G}^{-g}\mathbf{d}^{(M)} = \mathbf{V}^{(M)}\mathbf{V}^{(M)T}\mathbf{m} = \mathbf{m}$$

$$\mathbf{C}_m^{(M)} = \mathbf{G}^{-g}\mathbf{C}_d\mathbf{G}^{-g\mathrm{T}} = \mathbf{V}^{(M)}\mathbf{\Lambda}^{(M)}\mathbf{V}^{(M)\mathrm{T}} = \mathbf{C}_m$$

$$\mathbf{R}^{(M)} = \mathbf{I}$$

(6)

The other cases represent weighted averages that have different (and as we will show, poorer) resolution.

The minimum length solution is known to minimize the *Direchlet* spread $J^{(N)}$ of resolution (Menke 2018), defined as:

$$J^{(N)} = \sum_i \sum_j D_{ij}^2 = \sum_j \left[\sum_i D_{ji}^{\mathrm{T}} D_{ij} \right] = \mathrm{tr}(\mathbf{D}^{\mathrm{T}}\mathbf{D}) \quad \text{with} \quad \mathbf{D} = \mathbf{R}^{(N)} - \mathbf{I}$$

(7)

The measure $J^{(N)}$ quantifies how close the resolution is to the perfect resolution case, for which $\mathbf{R}^{(N)} = \mathbf{I}$ and $J^{(N)} = 0$. Substituting $\mathbf{V}^{(N)}\mathbf{V}^{(N)T}$ for $\mathbf{R}^{(N)}$ yields:

$$\mathbf{D}^{\mathrm{T}}\mathbf{D} = \left(\mathbf{V}^{(N)}\mathbf{V}^{(N)T} - \mathbf{I}\right)^{\mathrm{T}}\left(\mathbf{V}^{(N)}\mathbf{V}^{(N)T} - \mathbf{I}\right) = \mathbf{V}^{(N)}\mathbf{V}^{(N)T}\mathbf{V}^{(N)}\mathbf{V}^{(N)T} - 2\mathbf{V}^{(N)}\mathbf{V}^{(N)T} + \mathbf{I} =$$

$$= \mathbf{V}^{(N)}\mathbf{V}^{(N)T} - 2\mathbf{V}^{(N)}\mathbf{V}^{(N)T} + \mathbf{I} = \mathbf{I} - \mathbf{V}^{(N)}\mathbf{V}^{(N)T}$$

(8)

so that

$$J^{(N)} = \mathrm{tr}(\mathbf{D}^{\mathrm{T}}\mathbf{D}) = \mathrm{tr}(\mathbf{I}) - \mathrm{tr}\left(\mathbf{V}^{(N)}\mathbf{V}^{(N)T}\right) = M - N$$

(9)

The last equality follows from:

$$\mathrm{tr}(\mathbf{I}) = N \quad \text{and} \quad \mathrm{tr}\left(\mathbf{V}^{(N)}\mathbf{V}^{(N)T}\right) = \sum_i \sum_j v_j^{(i)} v_j^{(i)} = \sum_{i=1}^{N} \left|\mathbf{v}^{(i)}\right|^2 = N$$

(10)

Note that the spread function $J^{(N)}$ monotonically decreases with increasing N, and is zero when $N = M$.

While the $N = M$ case has perfect resolution, it does not necessarily have acceptably small variance. The size $K^{(N)}$ of the variance can be quantified as:

$$K^{(N)} = \mathrm{tr}\left(\mathbf{C}_m^{(N)}\right) = \mathrm{tr}\left(\mathbf{V}^{(N)}\mathbf{\Lambda}^{(N)}\mathbf{V}^{(N)\mathrm{T}}\right) = \mathrm{tr}\left(\mathbf{\Lambda}^{(N)}\right) = \sum_{i=1}^{N} \left[\mathbf{\Lambda}^{(N)}\right]_{ii}$$

(11)

Here we have used the fact that the trace is invariant under orthogonal transformations. Note that $K^{(N)} < N\sigma^2$, since each $\lambda^{(i)} < \sigma^2$. The size $K^{(N)}$ monotonically increases with increasing N since the eigenvalues, being variances, are all positive.

The functions $J^{(N)}$ and $K^{(N)}$ define the *trade-off curve* $K(J)$. Adding data increases N and moves the solution along this curve. The spread $J^{(N)}$ of resolution monotonically decreases and the size $K^{(N)}$ of variance monotonically increases. Finding a solution with both arbitrarily small spread of resolution *and* arbitrarily small size of variance is impossible; at best, one can find a point along the curve where the two have acceptable values.

The size of the prior covariance, $K^{(A)} = \mathrm{tr}(\mathbf{C}_m^{\mathrm{A}})$, can be used as a reference value against which to judge a particular choice of N. Only the cases for which $K^{(N)} < K^{(A)}$ have a variance controlled by the data (as contrasted to the prior information).

Backus-Gilbert Resolution Case. A criticism of the Dirichlet measure is that it does not penalize distance from the main diagonal. The Backus-Gilbert measure weights by this distance and produces a better estimate of the degree to which resolution is localized. It is defined as:

$$J^{(N)} = \sum_{i=1}^{N}\sum_{j=1}^{M} w(l,k) R_{ij}^{(N)} R_{ij}^{(N)} \quad \text{with the constraint} \quad \sum_{j=1}^{\boldsymbol{M}} R_{ij}^{(N)} = 1 \ \text{ for all } i \tag{12}$$

Here $w(l,k)$ is a penalty function that quantifies the physical distance between model parameters m_l and m_k and that satisfied $w(l,l) = 0$ and $w(l,k) > 0$ for $l \neq k$. For a one-dimensional organization of model parameters along the x-axis, the function $w(l,k) = (k-l)^2$ usually suffices. Backus and Gilbert (1967) showed that the solution to this problem is:

$$G_{kl}^{-g} = \frac{\sum_i u_i \left[\left\{\mathbf{S}^{(k)}\right\}^{-1}\right]_{il}}{\sum_i \sum_i u_i \left[\{\mathbf{S}^{(k)}\}^{-1}\right]_{ij} u_j} \quad \text{with} \quad u_j = \sum_i G_{jk}^{(N)}$$

$$\text{and} \quad \left[\mathbf{S}^{(k)}\right]_{ij} = \sum_l w(l,k)\, G_{il}^{(N)} G_{jl}^{(N)} \tag{13}$$

This form of the generalized inverse can be used in place of the minimum length generalized inverse $\mathbf{G}^{-g} = \mathbf{G}^{(N)T}\left[\mathbf{G}^{(N)}\mathbf{G}^{(N)T}\right]^{-1}$. However, the spread or resolution $J^{(N)}$ must be evaluated using Equation (12) and the size of variance by $K^{(N)} = \mathrm{tr}\left(\mathbf{G}^{-g}\boldsymbol{\Lambda}^{(N)}\mathbf{G}^{-g\mathrm{T}}\right)$, since the simplifications developed for the Dirichlet case do not apply.

Resolution of Deviations from the prior model. Consider a linear least squares problem with data equation $\mathbf{d} = \mathbf{Gm}$ (with covariance $\mathbf{C}_d$) and prior information equation $\mathbf{h} = \mathbf{Hm}$ (with covariance $\mathbf{C}_h$). Generalized Least Squares gives the estimated model parameters $\mathbf{m}^{\mathrm{est}}$ and posterior covariance $\mathbf{C}_m$ as:

$$\mathbf{m}^{\text{est}} = \mathbf{A}^{-1}[\mathbf{G}^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{d}^{obs} + \mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathrm{h}}^{-1}\mathbf{h}]$$

$$\mathbf{C}_m = \mathbf{A}^{-1} \quad \text{with} \quad \mathbf{A} \equiv [\mathbf{G}^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{G} + \mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathrm{h}}^{-1}\mathbf{H}]$$

(14)

As Menke (2018) points out, the problem is perfectly resolved as long as $\mathbf{A}^{-1}$exists; that is $\mathbf{R} = \mathbf{I}$ whenever the data and the prior information, taken together, are sufficient to specify a solution. This is the $N = M$ case we have been discussing above. Nevertheless, a non-trivial resolution matrix can be constructed for *deviations* $\Delta\mathbf{m} = \mathbf{m} - \mathbf{m}^{(A)}$ of the model parameters away from the prior solution $\mathbf{m}^{(A)}$. By prior solution, we mean the solution implied by the prior information, acting alone; that is: $\mathbf{m}^{(A)} = [\mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathrm{h}}^{-1}\mathbf{H}]^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathrm{h}}^{-1}\mathbf{h}$, which has covariance $\mathbf{C}_m^{(\mathrm{A})} = [\mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathrm{h}}^{-1}\mathbf{H}]^{-1}$. (Menke (2018) recommends adding very weak smallness prior information to the problem in cases where the prior information is not complete, so that $[\mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathrm{h}}^{-1}\mathbf{H}]^{-\mathbf{1}}$ always exists). The resolution matrix $\mathbf{R}^{\mathrm{G}}$ for $\Delta\mathbf{m}$ is:

$$\mathbf{R}^{\mathrm{G}} \equiv \mathbf{A}^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{G} = \mathbf{A}^{-1}\left(\mathbf{A} - \mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathrm{h}}^{-1}\mathbf{H}\right) = \mathbf{I} - \mathbf{A}^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathrm{h}}^{-1}\mathbf{H}$$

$$\mathbf{R}^{\mathrm{G}} = \mathbf{I} - \mathbf{C}_m\mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathrm{h}}^{-1}\mathbf{H} = \mathbf{I} - \mathbf{C}_m\left[\mathbf{C}_{\boldsymbol{m}}^{(A)}\right]^{-\mathbf{1}}$$

(15)

Note that the resolution is exactly zero when $\mathbf{C}_m = \mathbf{C}_m^{(A)}$. This is the case where the data contributes no information, so that the posterior covariance of the model parameters is just their prior covariance.

Example. We examine a weakly nonlinear test scenario, with $M = 11$ model parameters uniformly spaced along the x-axis, with true values $m_i^{true} = 1$. The use the data equation:

$$d_i = \sum_j G_{ij}^0 m_j + q_0 \sum_j Q_{ij} m_j^2$$

(16)

The matrices $G_{ij}^{(0)} = c^{(i)} \exp\{-c^{(i)}(j-1)\}$ and $Q_{ij} = d_j^{(i)} \exp\left\{-d_j^{(i)}(j-1)\right\}$ are chosen to exponentially decay with column number j, with decay constants that increase with row number i according to $c^{(i)} = 0.03 \times i$ and $d^{(i)} = 0.03 \times (i - ½)$. The problem is made weakly nonlinear by setting $q_0 = 0.1$, leading to a solution that differs by about 10% from the $q_0 = 0$ linear solution. The p.d.f. $p(\mathbf{d}^{obs}|\mathbf{m})$ is taken to be Normal, uncorrelated and with uniform variance $\sigma_d^2 = 10^{-4}$, and centered on $\mathbf{d}^{true} = \mathbf{d}(\mathbf{m}^{true})$ The prior information equation is of the form $\mathbf{Hm} = \mathbf{h}$ with the first $M - 1$ rows specifying the first differences of adjacent model

parameters and the last row specifying the value of m_M. The variance of the prior information is taken to be $\sigma_h^2 = 1$, which is much larger than the variance of the data. The prior information $\mathbf{h}^{true} = \mathbf{H}\mathbf{m}^{true}$ implies a solution $\mathbf{m}^{(A)} = [\mathbf{H}^{\mathrm{T}}\mathbf{C}_h^{-1}\mathbf{H}]^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{h}^{true}$ with covariance $\mathbf{C}_m^{\mathrm{A}} = [\mathbf{H}^{\mathrm{T}}\mathbf{C}_h^{-1}\mathbf{H}]^{-1}$. The prior p.d.f. $p_A(\mathbf{m})$ is taken to be Normal, with covariance $\mathbf{C}_m^{\mathrm{A}}$, and centered on $\mathbf{m}^{(A)}$.

The solution and its variance and resolution is inferred from 10^6 realizations drawn from the posterior distribution $p(\mathbf{m}|\mathbf{d}^{obs}) = p(\mathbf{d}^{obs}|\mathbf{m})\ p_A(\mathbf{m})$. They compare well with the solution and covariance estimated using Linearized Generalized Least Squares (Menke 2018), utilizing the gradient:

$$G_{ij} \approx \frac{\partial d_i}{\partial m_j} = G_{ij}^0 + 2q_0 Q_{ij} m_j \tag{17}$$

The Direchlet trade-off curve (Figure 1A) demonstrates that the size of variance $K^{(N)}$ are below the reference value $K^{(A)}$ for all values of N. Decreasing N from 11 to 7 decreases the size of variance by a factor of about three, while causing only a modest degradation of the resolution (Figure 1B).

The Backus-Gilbert trade-off curve (Figure 2) is broadly similar to the Dirichlet curve, except that, as expected, the elements of the resolution matrices are wider than the Dirichlet case, but have edges that decline very rapidly with distance from the main diagonal

The resolution matrix $\mathbf{R}^{\mathrm{G}}$ for deviations of the model about the prior model (Figure 3) is not fully-resolved, but is moderately-well localized.

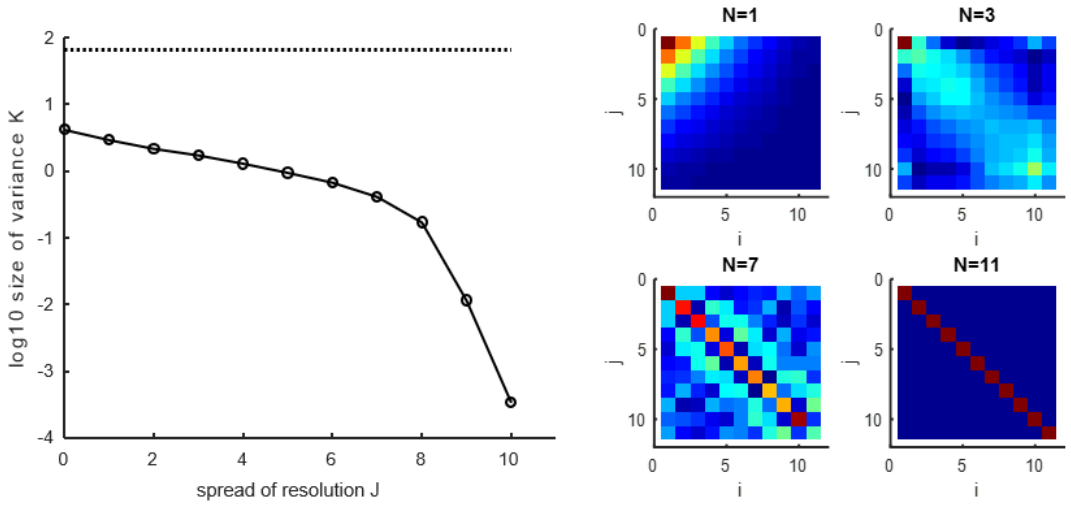

Figure 1. Dirichlet resolution case. (A) Trade off curve (curve with circles) for size of variance $K^{(N)}$ versus spread of resolution $J^{(N)}$ for the sample problem. The reference value $K^{(A)}$ (dotted line) is everywhere above the curve. (B) The resolution matrix $\mathbf{R}^{(N)}$ (colors) of the sample problem for selected values of N. The reference value $K^{(A)}$ (dotted line) is everywhere above the curve. The resolution becomes more localized (concentrated along the main diagonal) as the value of N is increased.

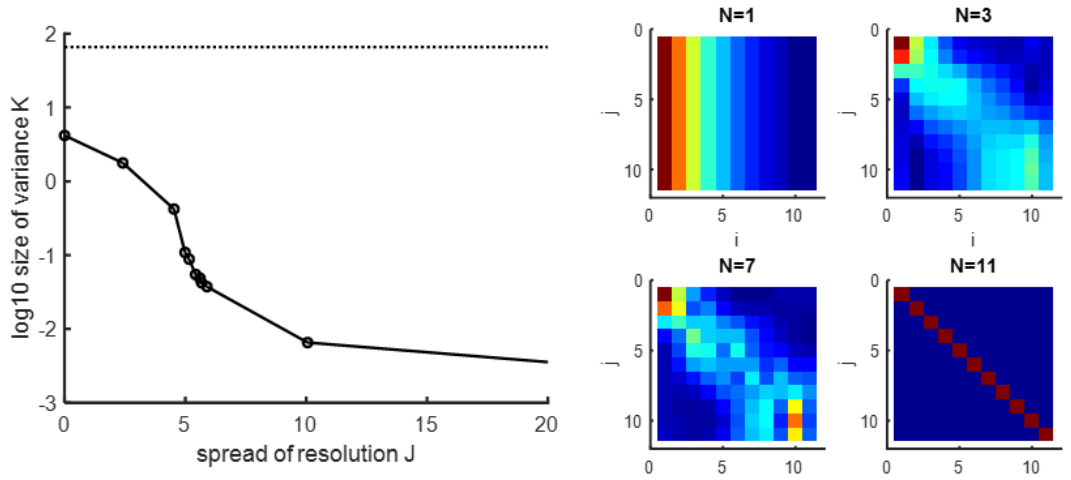

Figure 2. Backus-Gilbert resolution case. (A) Trade off curve (curve with circles) for size of variance $K^{(N)}$ versus spread of resolution $J^{(N)}$ for the sample problem. The reference value $K^{(A)}$ (dotted line) is everywhere above the curve. (B) The resolution matrix $\mathbf{R}^{(N)}$ (colors) of the sample problem for selected values of N. The reference value $K^{(A)}$ (dotted line) is everywhere above the curve. The resolution becomes more localized (concentrated along the main diagonal) as the value of N is increased.

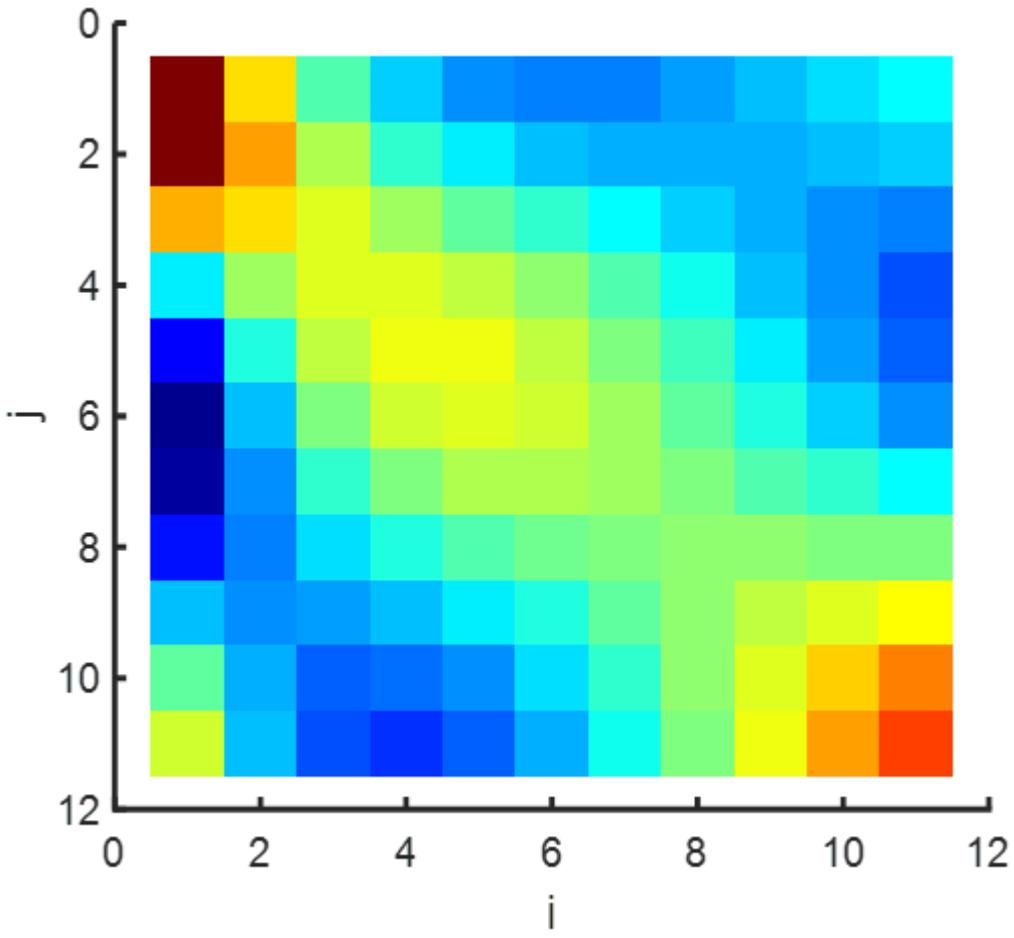

Figure 3. The resolution matrix $\mathbf{R}^G$ for deviations of the solution about the prior information.

Backus, G. and F. Gilbert (1967). Numerical application of a formalism for geophysical inverse problems, Geophys. J. R. Astr. Soc. 13, 247-276.

Menke, W., 2018. Geophysical Data Analysis: Discrete Inverse Theory, 4th Edition, Elsevier, 322pp.