Singular values of covariance matrices under localization

Travis Harty

1 Optimal linear transform

We update our state $x \sim N(\mu, \mathbf{P})$ using observation $y = \mathbf{H}\mu + \epsilon$, where $\epsilon \sim N(0, \mathbf{R})$, $\dim x = N_x$, and $\dim y = N_y$. We will perform this update by first transforming our variables using linear transformations $\mathbf{T}_x$ and $\mathbf{T}_y$ for x and y respectively such that $\mathbf{T}_x x = \tilde{x} \sim N(\tilde{\mu}, \mathbf{I})$ and $\mathbf{T}_y y = \tilde{y} = \mathbf{\Sigma}\tilde{\mu} + \tilde{\epsilon}$, where $\mathbf{\Sigma}$ is diagonal, and $\tilde{\epsilon} \sim N(0, \mathbf{I})$.

Take,

$$x' = \mathbf{P}^{-1/2} x \text{ and } y' = \mathbf{R}^{-1/2} y.$$

We then have,

$$\begin{aligned} y &= \mathbf{H}\mu + \epsilon \\ \mathbf{R}^{1/2} y' &= \mathbf{H}\mathbf{P}^{1/2}\mu' + \epsilon \\ y' &= \mathbf{R}^{-1/2}\mathbf{H}\mathbf{P}^{1/2}\mu' + \epsilon', \end{aligned}$$

where $\epsilon \sim N(0, \mathbf{I})$. We can then take the singular value decomposition of $\mathbf{R}^{-1/2}\mathbf{H}\mathbf{P}^{1/2}$ yielding,

$$\begin{aligned} y' &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mu' + \epsilon' \\ \mathbf{U}^T y' &= \mathbf{\Sigma}(\mathbf{V}^T\mu') + \tilde{\epsilon} \\ \tilde{y} &= \mathbf{\Sigma}\tilde{\mu} + \tilde{\epsilon} \end{aligned}$$

where $\mathbf{\Sigma}$ is diagonal and $\tilde{\epsilon} \sim N(0, \mathbf{I})$. We have $\mathbf{T}_x = \mathbf{V}^T\mathbf{P}^{-1/2}$ and $\mathbf{T}_y = \mathbf{U}^T\mathbf{R}^{-1/2}$.

2 Dimension reduction optimal linear transform

We again update our state $x \sim N(\mu, \mathbf{P})$ with observations $y = \mathbf{H}x + \epsilon$, where $\epsilon \sim N(0, \mathbf{R})$. We will now calculate

$$\mathbf{P} = \mathbf{Q}_x \mathbf{\Lambda}_x \mathbf{Q}_x^T \text{ and } \mathbf{R} = \mathbf{Q}_y \mathbf{\Lambda}_y \mathbf{Q}_y^T$$

and choose N_{λ_x} and N_{λ_y} that are the number of eigenvalues to keep for P and R respectively such that,

$$\dim(\mathbf{Q}_x) = (N_x, N_{\lambda_x});\ \dim(\mathbf{\Lambda}_x) = (N_{\lambda_x}, N_{\lambda_x})$$

and

$$\dim(\mathbf{Q}_y) = (N_y, N_{\lambda_y});\ \dim(\mathbf{\Lambda}_y) = (N_{\lambda_y}, N_{\lambda_y})$$

We can then repeat the above calculations while reducing the transformed variables' dimensions. Take,

$$x' = \mathbf{\Lambda}_x^{-1/2}\mathbf{Q}_x^T x \text{ and } y' = \mathbf{\Lambda}_y^{-1/2}\mathbf{Q}_y^T y$$

Note that we have,

$$x' \sim N(\mathbf{\Lambda}_x^{-1/2}\mathbf{Q}_x^T\mu, \mathbf{\Lambda}_x^{-1/2}\mathbf{Q}_x^T\mathbf{P}\mathbf{Q}\mathbf{\Lambda}_x^{-1/2}) = N(\mu', \mathbf{\Lambda}_x^{-1/2}\mathbf{\Lambda}_x\mathbf{\Lambda}_x^{-1/2}) = N(\mu', \mathbf{I}_{N_{\lambda_x}})$$

$$y' \sim N(\mathbf{H}\mu', \mathbf{I}_{N_{\lambda_x}})$$

and

$$x \approx \mathbf{Q}_x\mathbf{\Lambda}_x^{1/2}x' + (\mathbf{I} - \mathbf{Q}_x\mathbf{Q}_x^T)\mu$$

where there is equality when N_{λ_x} is the true rank of $\mathbf{P}$.

Assuming that N_{λ_x} and N_{λ_y} are the ranks of P and R respectively, we have,

$$y = \mathbf{H}\mu + \epsilon$$

$$\mathbf{Q}_y\mathbf{\Lambda}_y^{1/2}y' = \mathbf{H}\mathbf{Q}_x\mathbf{\Lambda}_x^{1/2}\mu' + \epsilon$$

$$y' = \mathbf{\Lambda}_y^{-1/2}\mathbf{Q}_y^T\mathbf{H}\mathbf{Q}_x\mathbf{\Lambda}_x^{1/2}\mu' + \epsilon',$$

where $\epsilon' \sim N(0, \mathbf{I}_{N_{\lambda_y}})$. We can then take the singular value decomposition of $\mathbf{\Lambda}_y^{-1/2}\mathbf{Q}_y^T\mathbf{H}\mathbf{Q}_x\mathbf{\Lambda}_x^{1/2}$ yielding,

$$y' = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mu' + \epsilon'$$

$$\mathbf{U}^T y' = \mathbf{\Sigma}(\mathbf{V}^T\mu') + \tilde{\epsilon}$$

$$\tilde{y} = \mathbf{\Sigma}\tilde{\mu} + \tilde{\epsilon}$$

where $\mathbf{\Sigma}$ is diagonal and $\tilde{\epsilon} \sim N(0, \mathbf{I}_{N_{\lambda_y}})$. We have $\mathbf{T}_x = \mathbf{V}^T\mathbf{\Lambda}_x^{-1/2}\mathbf{Q}_x^T$ and $\mathbf{T}_y = \mathbf{U}^T\mathbf{\Lambda}_y^{-1/2}\mathbf{Q}_y^T$. Note that we have,

$$\mathbf{T}_x\mathbf{T}_x^{-R} = \mathbf{V}^T\mathbf{\Lambda}_x^{-1/2}\mathbf{Q}_x^T\mathbf{Q}_x\mathbf{\Lambda}_x^{1/2}\mathbf{V} = \mathbf{I}_{N_{\lambda_x}},$$

but there is no left inverse. However, the right inverse is something like a left inverse in that

$$\mathbf{T}_x^{-R}\mathbf{T}_x = \mathbf{Q}_x\mathbf{\Lambda}_x^{1/2}\mathbf{V}\mathbf{V}^T\mathbf{\Lambda}_x^{-1/2}\mathbf{Q}_x^T$$

$$\mathbf{T}_x^{-R}\mathbf{T}_x = \mathbf{Q}_x\mathbf{Q}_x^T$$

which is equal to the identity matrix only if P is full rank, and all eigenvectors are kept. However, if μ is in the span of the columns of $\mathbf{Q}_x$, then we have,

$$\mathbf{T}_x^{-R}\mathbf{T}_x\mu = \mathbf{Q}_x\mathbf{Q}_x^T\mu = \mu$$

even if $\mathbf{T}_x$ does not have a left inverse. Furthermore, if μ is not in the span of the columns of $\mathbf{Q}_x$, this is possible even if the range of $\mathbf{P}$ and $\mathbf{Q}_x$ are equal, then we have,

$$(\mathbf{T}_x^{-R}\mathbf{T}_x + \mathbf{I}_{N_x} - \mathbf{Q}_x\mathbf{Q}_x^T)\mu = \mathbf{Q}_x\mathbf{Q}_x^T\mu + \mu - \mathbf{Q}_x\mathbf{Q}_x^T\mu = \mu.$$

We can therefore always recover the parts of μ that are removed because of the representation of μ in the column space of $\mathbf{Q}_x$.

3 Multi-scale optimal transform

Suppose we aga

4 Two ensembles

Suppose that we have two ensembles representing the state x. One ensemble, $\mathbf{X}_c$ will have many members, but will have a lower resolution and only represent large scale structures of the state. The other ensemble, $\mathbf{X}$ will have fewer members, but will have a finer resolution and therefore be able to represent fine scale structures of the state. We then have,

$$\dim \mathbf{X}_c = N_{x_c}, N_{e_c}$$

and

$$\dim \mathbf{X} = N_x, N_e$$

where $N_{x_c} < N_x$ and $N_{e_c} > N_e$.

Taking the singular value decomposition of the ensemble of coarse perturbations, $\mathbf{X}_c^*$, will give us the eigenvalues and eigenvectors of the sample covariance matrix derived from $\mathbf{X}_c$.

$$\mathbf{X}_c^* = \left(\mathbf{X}_c - \frac{1}{N}\bar{\mathbf{X}}_c \mathbf{1}_c\right) / \sqrt{N_{e_c} - 1}$$

$$\mathbf{Q}_c \mathbf{\Lambda}_c^{1/2} \mathbf{V}_{temp}^T = \mathbf{X}_c^*$$

$$\mathbf{Q}_c \mathbf{\Lambda}_c \mathbf{Q}_c^T = (\mathbf{X}_c^*)(\mathbf{X}_c^*)^T$$

where $\mathbf{1}_c$ is a matrix of all ones of with the same dimension as $\mathbf{X}_c$ and $\bar{\mathbf{X}}_c$ is the sample mean of $\mathbf{X}_c$. Assuming that we know $\mathbf{Q}_y$ and $\mathbf{\Lambda}_y$, we can then calculate $\mathbf{U}_c$ and $\mathbf{V}_c$,

$$\mathbf{U}_c \mathbf{\Sigma}_c \mathbf{V}_c^T = \mathbf{\Lambda}_y^{-1/2} \mathbf{Q}_y^T \mathbf{H} \mathbf{Q}_c \mathbf{\Lambda}_c^{1/2}.$$

This gives us the optimal linear transform, or as much of it as we deem appropriate to keep, for the coarse scale system. We can then use this to calculate the coarse parts of the optimal linear transform for the fine scale system. To do this, we can use $\mathbf{Q}_c$, $\mathbf{U}_c$, and $\mathbf{V}_c$, but must recalculate $\mathbf{\Lambda}_c$ and $\mathbf{\Sigma}_c$ for their fine state equivalents $\mathbf{\Lambda}_{fc}$ and $\mathbf{\Sigma}_{fc}$.

In order to calculate $\mathbf{\Lambda}_{fc}$ we must first interpolate the eigenvectors $\mathbf{Q}_c$ to the fine state space resulting in $\mathbf{Q}_{fc}$. The interpolated eigenvectors will need to be modified so that they form a orthonormal set in fine space. Once this is done, we can then calculate their corresponding eigenvalues by assuming that the interpolated eigenvectors are left singular values of the ensemble perturbations for $\mathbf{X}$, $\mathbf{X}^*$:

$$\lambda_{fc}[i]^{1/2} = \|q_{fc}[i]^T \mathbf{X}^*\|$$

where $\lambda_{fc}[i]$ is the eigenvalue of $\mathbf{X}^* (\mathbf{X}^*)^T$ corresponding to the eigenvector $q_{fc}[i]$. We can then be arrange all of the eigenvalues and vectors into the matrices $\mathbf{\Lambda}_{fc}$ and $\mathbf{Q}_{fc}$.

We can calculate the singular values by assuming that $\mathbf{U}_c$ and $\mathbf{V}_c$ have columns made of left and right singular vectors of $\mathbf{\Lambda}_y^{1/2}\mathbf{Q}_y^T\mathbf{H}\mathbf{Q}_{fc}\mathbf{\Lambda}_{fc}^{-1/2}$:

$$\sigma_{fc}[i] = u_c[i]^T\mathbf{\Lambda}_y^{1/2}\mathbf{Q}_y^T\mathbf{H}\mathbf{Q}_{fc}\mathbf{\Lambda}_{fc}^{-1/2}v_c[i]$$

where $\sigma_{fc}[i]$ is the singular value corresponding to the left and right singular vectors $u_c[i]$ and $v_c[i]$. These singular values can then be arranged in the matrix $\mathbf{\Sigma}_{fc}$.

Now that these quantities have been calculated, we can then form the optimal linear transforms and their right inverses for our fine system corresponding to the large scales:

$$\mathbf{T}_{xl} = \mathbf{V}_c^T\mathbf{\Lambda}_{fc}^{-1/2}\mathbf{Q}_{fc}^T$$

$$\mathbf{T}_{xl}^{-R} = \mathbf{Q}_{fc}\mathbf{\Lambda}_{fc}^{1/2}\mathbf{V}_c$$

$$\mathbf{T}_{yl} = \mathbf{U}_c^T\mathbf{\Lambda}_y^{-1/2}\mathbf{Q}_y^T$$

$$\mathbf{T}_{yl}^{-R} = \mathbf{Q}_y\mathbf{\Lambda}_y^{1/2}\mathbf{U}_c$$

We can also calculate the linear transform in our fine system corresponding to our short scales. Since the above transforms will assimilate the information corresponding to the long scales, we are able to apply localization that is appropriate for our short scales when calculating the following transforms. This process is very similar to the standard reduced rank optimal linear transform, accept that we must ensure that certain vectors are orthogonal to those used above.

We start by calculating and localizing the sample covariance based on our fine ensemble $\mathbf{X}$:

$$\mathbf{P}_{loc} = \mathbf{L_s} \circ \mathbf{X}^* \left(\mathbf{X}^*\right)^T$$

5 Equivalence to standard KF

Sticking with the above notation, we can calculate the standard Kalman filter update and the corresponding equations:

$$\mathbf{K} = \mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1},$$

$$\mu^a = \mu + \mathbf{K}(y - \mathbf{H}\mu),$$

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}.$$

Alternatively, in terms of our transformed equations, we have:

$$\tilde{\mathbf{K}} = \mathbf{I}_{N_{\lambda_x}}\mathbf{\Sigma}^T(\mathbf{\Sigma}\mathbf{I}_{N_{\lambda_x}}\mathbf{\Sigma}^T + \mathbf{I}_{N_{\lambda_y}})^{-1} = \mathbf{\Sigma}^T(\mathbf{\Sigma}\mathbf{\Sigma}^T + \mathbf{I}_{N_{\lambda_y}})^{-1}$$

$$\tilde{\mu}^a = \tilde{\mu} + \tilde{\mathbf{K}}(\tilde{y} + \mathbf{\Sigma}\tilde{\mu})$$

$$\tilde{\mathbf{P}}^a = (\mathbf{I}_{N_{\lambda_x}} - \tilde{\mathbf{K}}\mathbf{\Sigma})\tilde{\mathbf{P}}.$$

We can then convert $\tilde{x}^a$ and $\tilde{\mathbf{P}}^a$ back to the original x-space through,

$$\mu_1^a = \mathbf{T}_x^{-R}\tilde{\mu}^a$$

and

$$\mathbf{P}_1^a = \mathbf{T}_x^{-R}\tilde{\mathbf{P}}^a\left(\mathbf{T}_x^{-R}\right)^T.$$

Assuming that the ranks of $\mathbf{P}$ and $\mathbf{R}$ are N_{λ_x} and N_{λ_y} respectively,

$$\tilde{\mathbf{K}} = \mathbf{\Sigma}^T(\mathbf{\Sigma}\mathbf{\Sigma}^T + \mathbf{I}_{N_{\lambda_y}})^{-1}$$

$$\tilde{\mathbf{K}} = \mathbf{T}_x\mathbf{P}\mathbf{T}_x^T(\mathbf{T}_y\mathbf{H}\mathbf{T}_x^{-R})^T\left((\mathbf{T}_y\mathbf{H}\mathbf{T}_x^{-R})(\mathbf{T}_y\mathbf{H}\mathbf{T}_x^{-R})^T + \mathbf{T}_y\mathbf{R}\mathbf{T}_y^T\right)^{-1}$$

$$\tilde{\mathbf{K}} = \mathbf{T}_x\mathbf{P}\mathbf{T}_x^T\left(\mathbf{T}_x^{-R}\right)^T\mathbf{H}^T\mathbf{T}_y^T\left(\mathbf{T}_y\mathbf{H}\mathbf{T}_x^{-R}\left(\mathbf{T}_x^{-R}\right)^T\mathbf{H}^T\mathbf{T}_y^T + \mathbf{T}_y\mathbf{R}\mathbf{T}_y^T\right)^{-1}$$

$$\tilde{\mathbf{K}} = \mathbf{T}_x\mathbf{P}\mathbf{Q}_x\mathbf{Q}_x^T\mathbf{H}^T\mathbf{T}_y^T\left(\mathbf{T}_y\left(\mathbf{H}\mathbf{Q}_x\Lambda_x\mathbf{Q}_x^T\mathbf{H}^T + \mathbf{R}\right)\mathbf{T}_y^T\right)^{-1}$$

If we assume that μ is in the span of $\mathbf{Q}_x$, we have,

$$x_1^a = \mathbf{T}_x^{-R}\tilde{x}^a$$

$$x_1^a = \mathbf{T}_x^{-R}\left(\tilde{x} + \tilde{K}\right)$$

6 Transformation from coarse OLD

We have two ensembles representing the same system. The ensemble X is a coarse representation of the state, but has many ensemble members. The ensemble Z is a fine representation of the sates, but has few ensemble members.

$$\dim X = N_x, N_{ex}$$

$$\dim Z = N_z, N_{ez}$$

In order to find the eigenvalues and eigenvectors of the sample correlation of X, we take the singular value decomposition of $\tilde{X} = (N_{ex}-1)^{-1/2}(X-\bar{X})/\operatorname{sd}(X)$, where $\operatorname{sd}(X)$ is the standard deviation of each element of X.

$$U_x S_x V_x^T = \tilde{X}$$

This means that,

$$U_x S_z^2 U_x^T = \hat{C}_x$$

where $\hat{C}_x$ is the sample correlation of X.

We want to then use U_x, or its leading columns, to estimate the leading U_z. To do this, we interpolate U_x to the z space and then use QR factorization to ensure the interpolated u_x's are orthonormal:

$$U_x = \operatorname{interp}(U_x)$$

$$U_{xi}, R = \mathrm{QR}(U_{xi}).$$

We then must choose how many of the columns of U_{xi} should be used. The best way to make this choice is unclear. A few ways that I am considering are:

1. Keep U_{xi} based on the cumulative sum of their corresponding eigenvalues.
2. Keep U_{xi} based on the rate of change of their corresponding eigenvalues.
3. Keep U_{xi} based on some measure of the length scale produced by the low rank approximation of $\hat{C}_x$ and the residual.
4. Keep U_{xi} based on how orthogonal the interpolated vector is in z space before orthogonalization.

It should be noted that keeping all columns can be detrimental to the assimilation process. If some of the columns of U_{xi} are representing small scale structures, then in the following steps they will still be used to represent some of these small scale structures.

After choosing which columns of U_{xi} to keep, we must determine what eigenvalues they should have. We do this by taking $\lambda = u_{xi}^T \hat{C}_z u_{xi}$ as the eigenvalue of uxi for our approximation of C_z. Alternatively, we can take $\lambda = (\tilde{Z}^T u_{xi})^T (\tilde{Z}^T u_{xi})$ where $\tilde{Z}$ is defined similarly as $\tilde{X}$. This will give us the leading eigenvectors and an approximation of their eigenvalues of C_Z.

These eigenvectors and eigenvalues represent the large scale structure of the problem. To then find the small scales of the problem, we take

$$C_z^{\perp} = \hat{C}_z - C_z^{\parallel}$$

where

$$C_z^{\parallel} = U_{xi} \Lambda_{xi} U_{xi}^T$$

and Λ_{xi} is the diagonal matrix with the λs described above on the diagonal. We can then localize $C_z^{\perp}$ to get at the small scales that are represented in Z.

We must choose how to localize $C_z^{\perp}$. One reasonable expectation is that the scales in $C_z^{\perp}$ will be shorter or similar to N_z/N_x. This choice will also be affected by the choice of how many columns of U_{xi} to keep. Once the localization matrix L is chosen, we can then generate our localized correlation matrix as:

$$C_z^{loc} = C_z^{\parallel} + L \circ C_Z^{\perp}$$

and the corresponding covariance matrix as:

$$P_z^{loc} = D_z C_z^{loc} D_z.$$

where D_z is the diagonal matrix with sample standard deviations of Z on the diagonal.

We can then find the leading eigenvectors and eigenvalues of C_z^{loc}:

$$Q \Lambda Q^T = C_z^{loc}$$

and use them to transform the z variable. First, we must calculate a whitening transformation

$$T_w = \Lambda^{-1/2} Q^T$$

and its right inverse

$$T_w^i = Q\Lambda^{1/2}$$

and the singular value decomposition

$$U\Sigma V^T = R^{-1/2} H T_w^i.$$

We can then define transformations for both z and y

$$T_z = V^T T_w$$

$$T_y = U^T R^{-1/2}$$

We then know $z^* = T_z z$ and $y^* = T_y y$ have identity covariance matrices and

$$y^* = \Sigma z^* + \epsilon^*$$

$$U^T R^{-1/2} y = \Sigma V^T \Lambda^{-1/2} Q^T z + \epsilon^*$$

$$y = R^{1/2} U \Sigma V^T \Lambda^{-1/2} Q^T z + \epsilon$$

$$y = R^{1/2} R^{-1/2} H Q \Lambda^{1/2} \Lambda^{-1/2} Q^T z + \epsilon$$

$$y = Hz + \epsilon$$

The question is: can I do this better? Specifically:

1. Can I calculate the eigenvectors and values of C^{loc} without generating $C^{\parallel}$ or $C^{\perp}$ explicitly?

2. Can I calculate the eigenvectors and values of P from those of C without generating C?

3. Can I calculate the singular vectors and values of $R^{-1/2} H T_w^i$ from the singular value decomposition of R, H, and T_w^i?