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,

$$x \approx \mathbf{Q}_x \Lambda_x^{1/2} x'$$
$$y \approx \mathbf{Q}_y \Lambda_y^{1/2} y'$$

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

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.

3 Equivalence to standard KF

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

$$\mathbf{K} = \mathbf{P}\mathbf{H}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1},$$
$$x^a = x + \mathbf{K}(y - \mathbf{H}x),$$
$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}.$$

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

$$\mathbf{K}' = \mathbf{I_x}\mathbf{\Sigma}^T(\mathbf{\Sigma}\mathbf{I_x}\mathbf{\Sigma} + \mathbf{I_y})^{-1} = \mathbf{\Sigma}^T(\mathbf{\Sigma}\mathbf{\Sigma}^T + \mathbf{I_y})^{-1}$$
$$x^{a\prime} = x' + \mathbf{K}'(y' + \mathbf{\Sigma}x')$$
$$\mathbf{P}^{a\prime} = (\mathbf{I} - \mathbf{K}'\mathbf{\Sigma})\mathbf{P}'$$

For the sake of greater generality take

4 Transformation from coarse

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:

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

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

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

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

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

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

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 $R^{-1/2}HT_w^i$ from the singular value decomposition of R, H, and T_w^i?