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,

$$x \approx \mathbf{Q}_x \mathbf{\Lambda}_x^{1/2} x'$$

$$y \approx \mathbf{Q}_y \mathbf{\Lambda}_y^{1/2} y'$$

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. 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 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 scale system corresponding to the large scales:

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

4 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,

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