

Simulation

Yukai Yang

Metric

I choose the Frobenius distance between two matrices on the Stiefel manifold, which is

$$\begin{aligned} d(X, Y) &= \|X - Y\|^2 = \text{tr}\{(X - Y)'(X - Y)\} \\ &= \text{tr}\{X'X + Y'Y - X'Y - Y'X\} \\ &= 2r - 2\text{tr}\{X'Y\} \end{aligned} \tag{1}$$

due to the fact that $X'X = I_r$ and $Y'Y = I_r$. Thus, $d(X, Y) \in [0, 4r]$, and when $X = Y$, $d(X, Y) = 0$, when $X = -Y$ (the furthest case), $d(X, Y) = 4r$.

What to report?

We simulate the data from the model one and two based on different settings. Then we do the filtering and collect U_t matrices, see Proposition 5.1-5.2. At time t , U_{t-1} is the modal orientation of the predictive distribution and U_t is the modal orientation of the updated distribution of the latent process.

Next step we can compute the distances $d(U_t, \alpha_t)$ (model 1) and $d(U_t, \beta_t)$ (model 2). These numbers should be close to zero. But notice that they are not supposed to approach zero with the increase of the sample size T ! The fact is that the filtering algorithms only give the conditional a posteriori distributions for the latent processes, and the modal orientation U_t is not supposed to be consistent to the true ones. Like Kalman filter, they are just something like “conditional expectations” and the corresponding “conditional variance” is not shrinking.

We can also try different initial values for the recursive filtering algorithms to see if the filtering allows for a wrong initial value (to what extent). The expectation is that, provided a wrong initial value, the sequence of U_t should go back to its true value soon.

Simulation Design for model one

We do not try different sample size T as they are irrelevant to the accuracy of the estimates of the latent process due to Proposition 5.1-5.2.

The dimension of the dependent variable will vary $p \in \{2, 10, 20\}$.

The rank number will be $r \in \{1, 2\}$.

The covariance matrix of the errors will be $\Omega = \rho I_p$, where $\rho \in \{1, 0.1\}$.

The initial value for the filtering takes the value α_0 (the true value) and $-\alpha_0$.

The simulation model takes the form (1) and (16) without Bz_t .

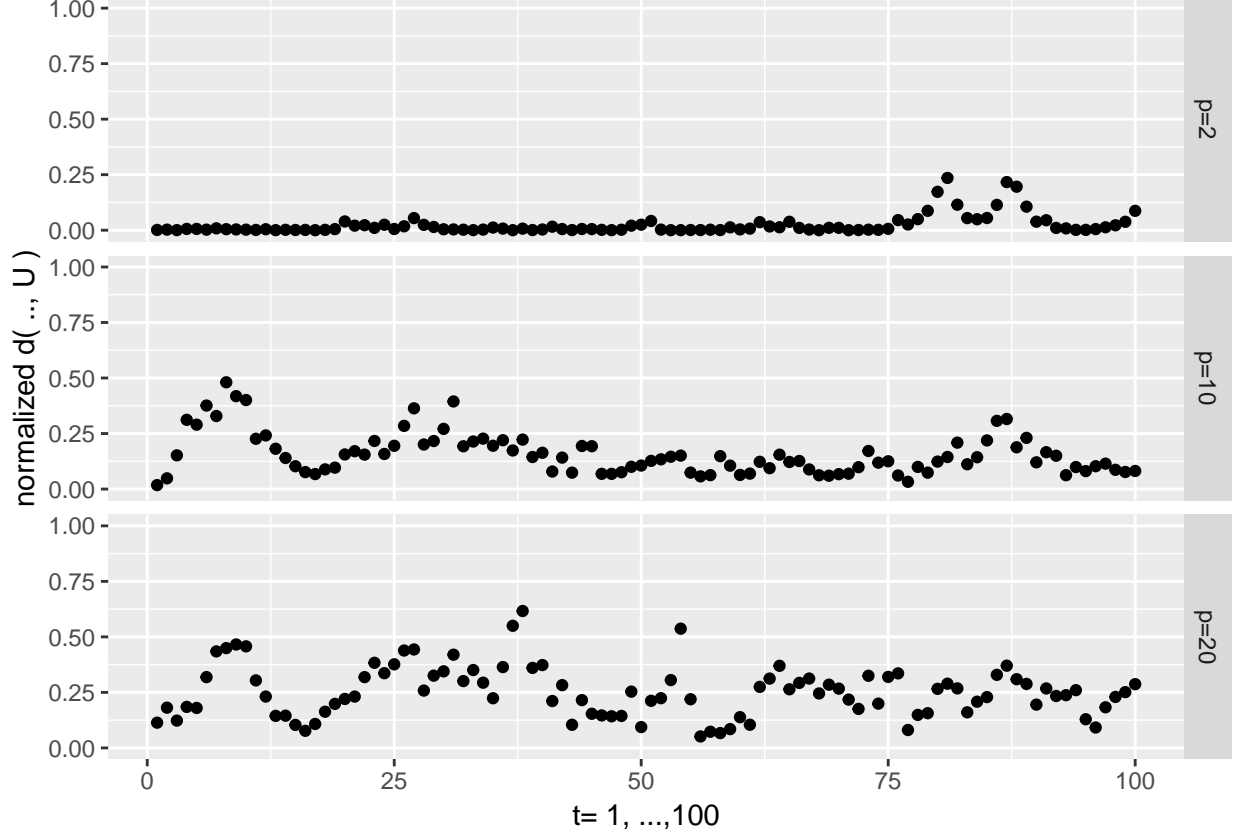
$\beta = (1, -1, \dots)'/\sqrt{q_1}$ where $q_1 = 3$, $\alpha_0 = (1, -1, \dots)'/\sqrt{p}$

The explanatory variable x_t is sampled from independent standard normal.

Different p and r with low and high D

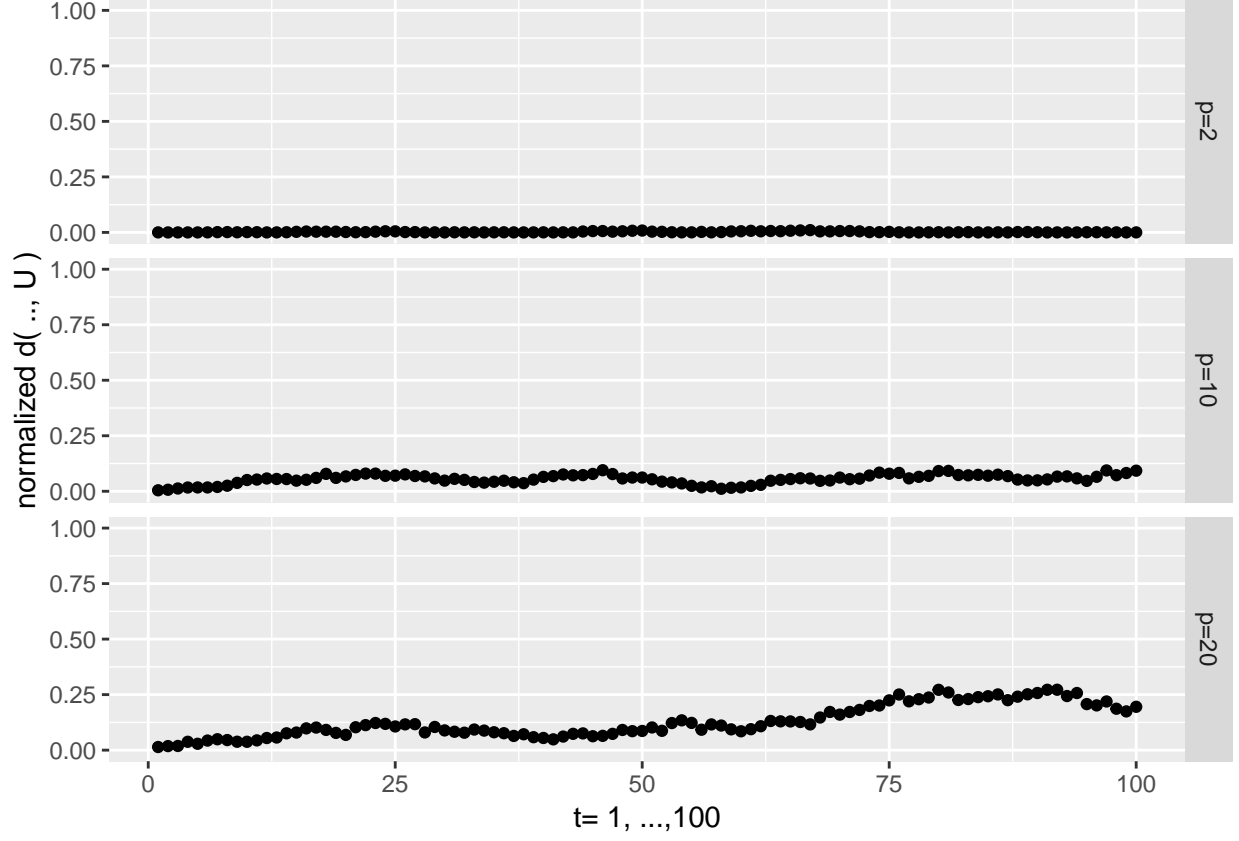
From Proposition 5.1, the dispersions of the predictive (a priori) and filtered (a posteriori) distributions of the latent processes are dominated by both the magnitude of the diagonals in matrix D and the pair of dimensions (p, r) .

In this experiment, we consider the setting $T = 100$, $r = 1$, and $\rho = 0.1$. For fixed $D = 50$, we compare the filtering results via different values of dimension p .



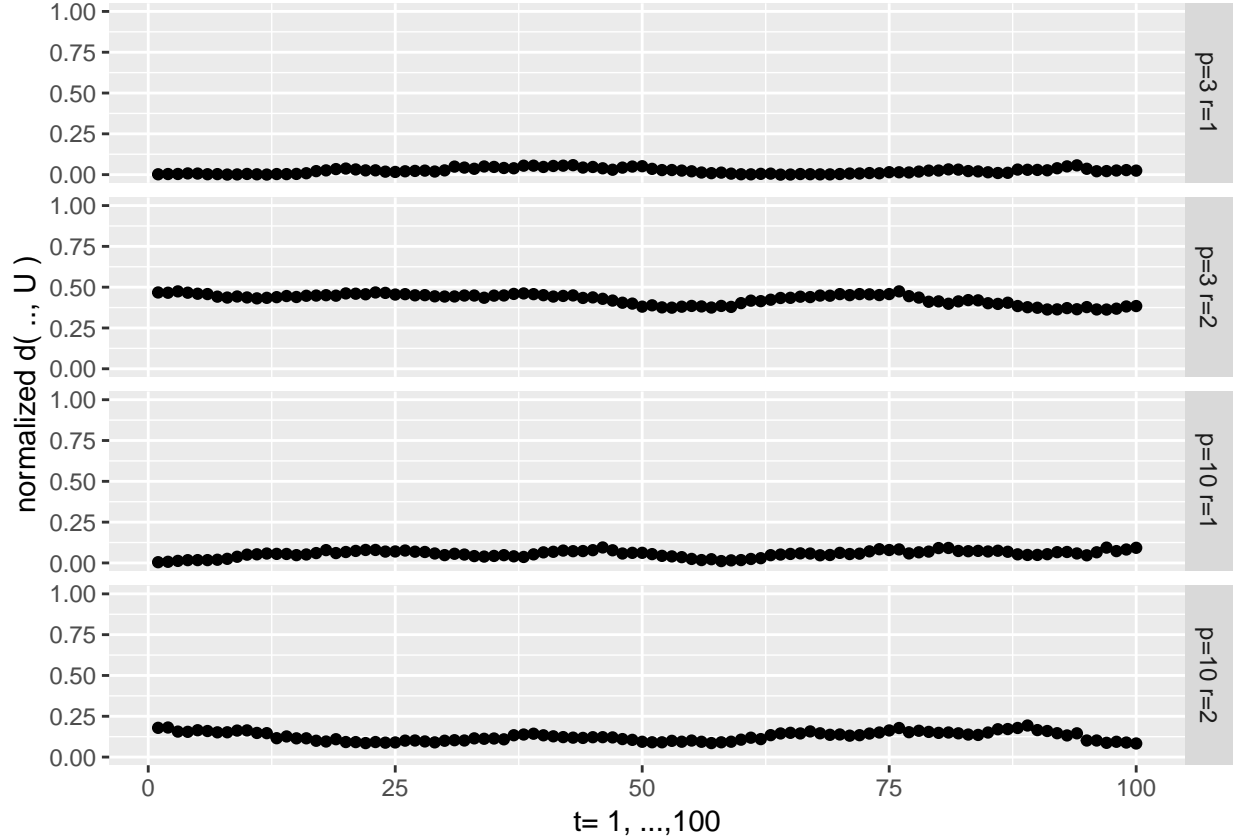
We see that, with the increase of the dimension, the results become more and more unstable. The results from the upper plot are acceptable as they are very close to zero, which implies that the filtered modal orientations are very close to the true values, and the corresponding dispersion is small enough.

Now we consider the cases when $D = 500$.

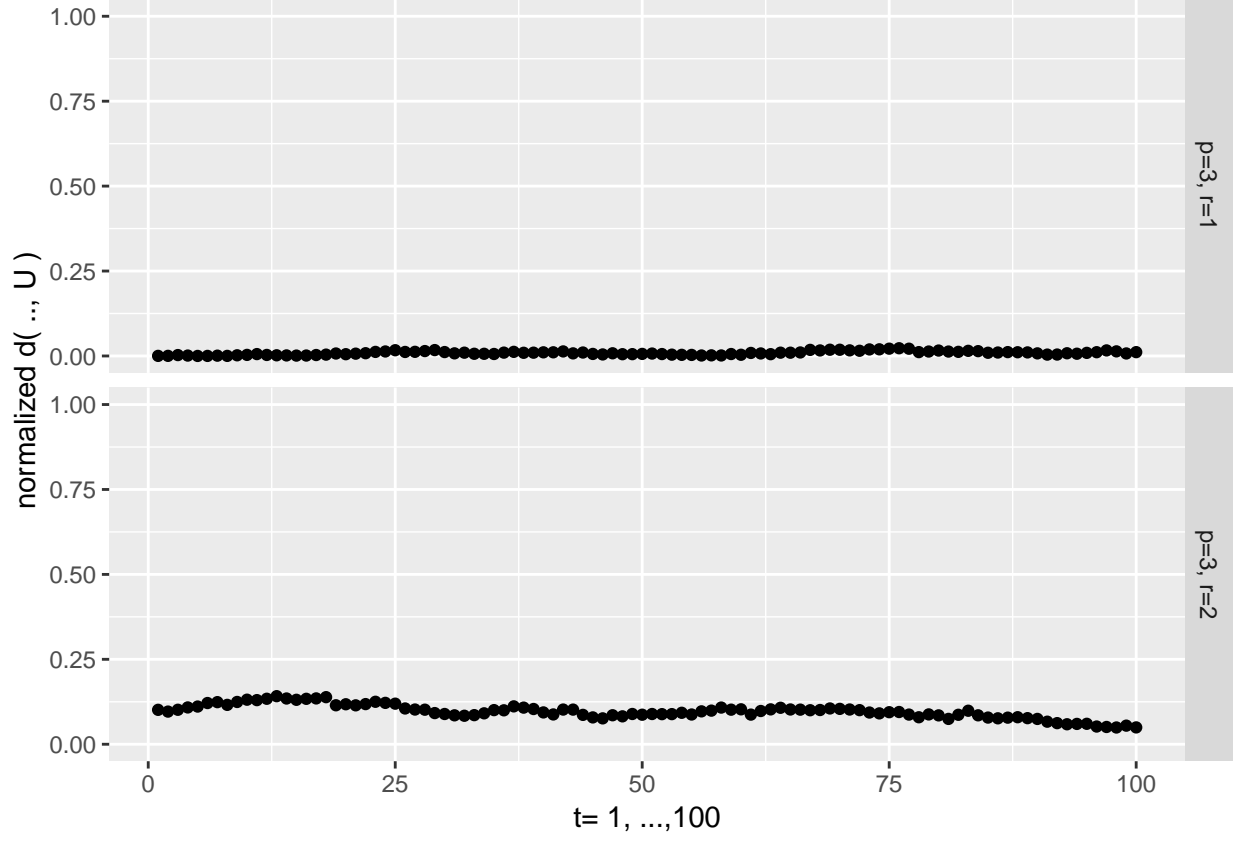


The results in the middle and lower plots show that modal orientations get closer to the true values as the dispersion becomes smaller.

Then we turn to investigate the impact of the rank r on the filtered distribution. In the following we use the setting $T = 100$, $\rho = 0.1$ and $D = 500$.



We see that the results in the case that $p = 3$ and $r = 2$ are not satisfactory as they look very stable around somewhere far from the true values, even though $r = 2$ is small. The difference between p and r plays the role but not the magnitude of r . It can be seen that, when the dimensions of the orthonormal matrix move towards a square matrix, the Laplace method fails to approximate the integral as there is a clear bias with small dispersion. Similar phenomena appear in nonlinear extended Kalman filter methods when the finite order Taylor expansion cannot reliably capture the nonlinearity. The following case with $D = 800$ shows that a bigger concentration D is the remedy for this.

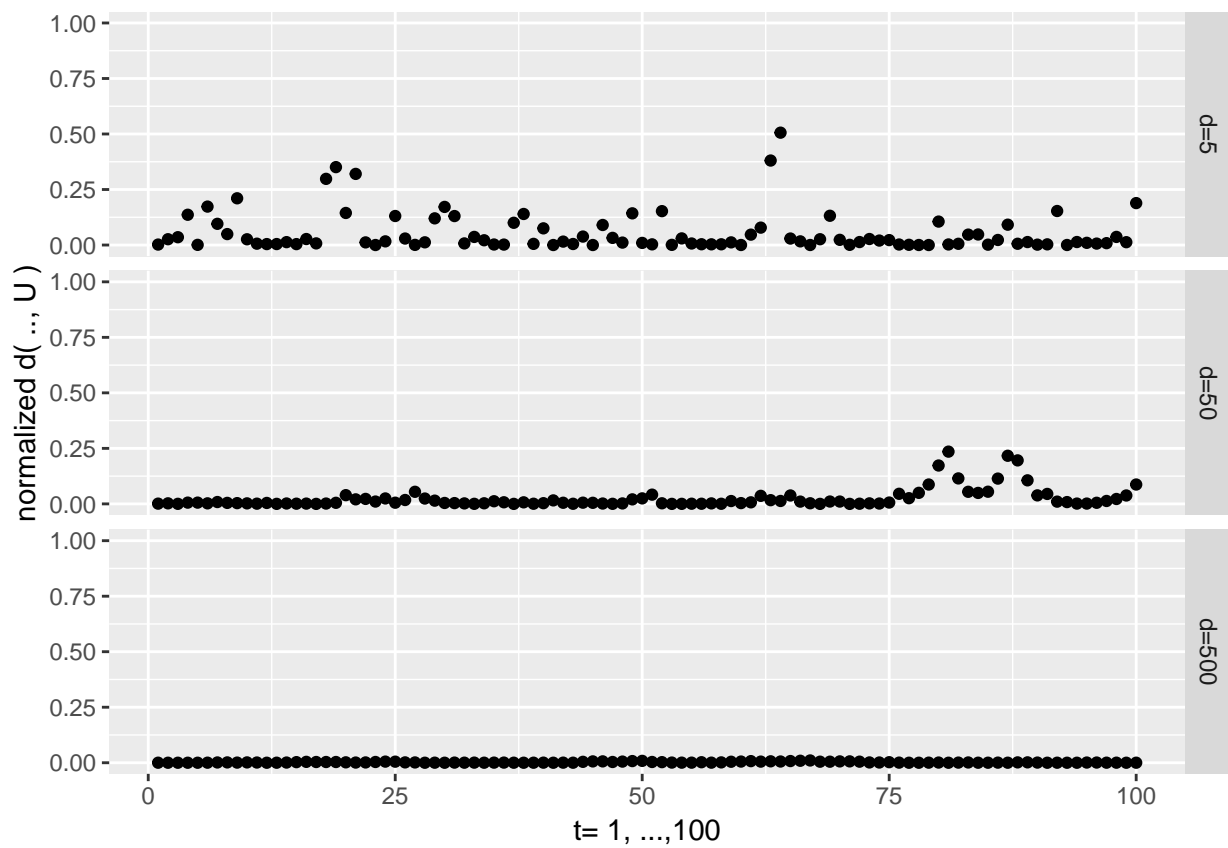
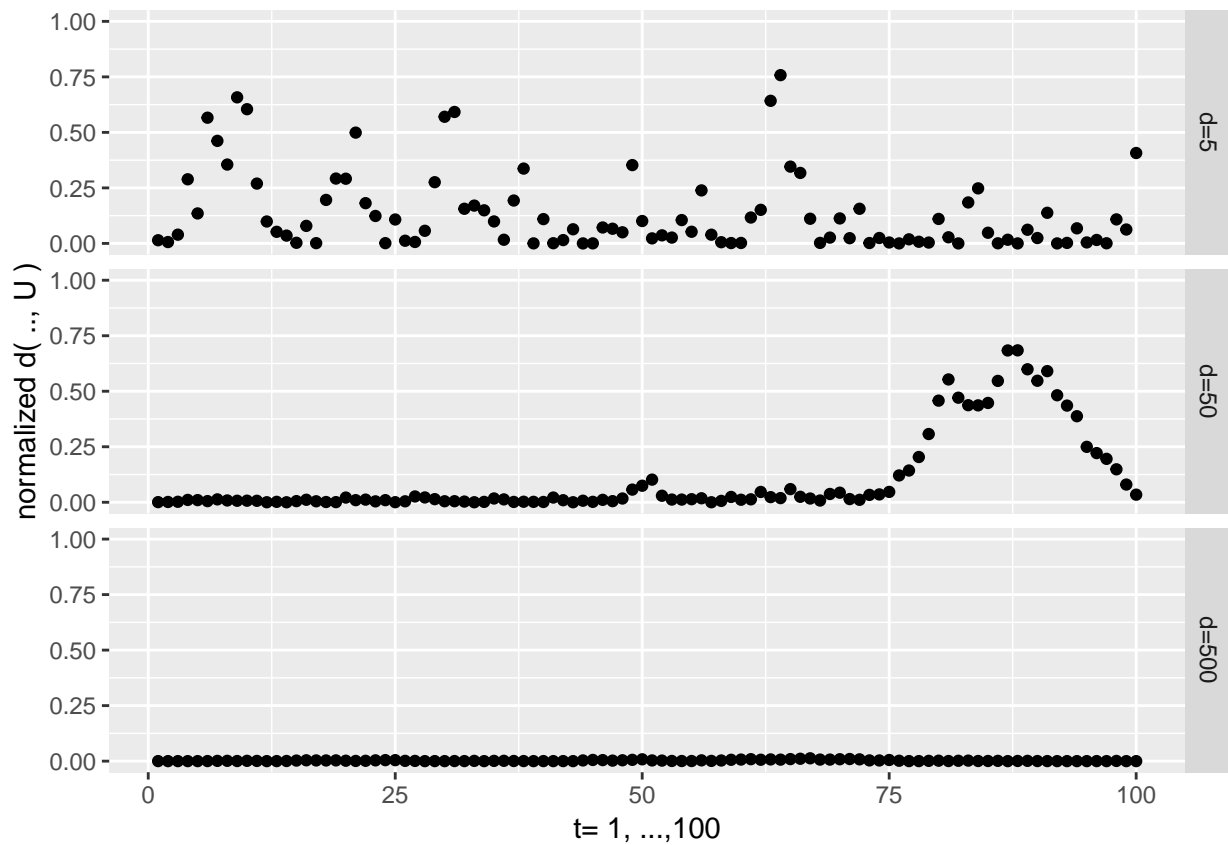


The conclusion is that low dimeion p , large distance between p and r , and big concentration D improve the accuracy of the filtering algorithm.

The covariance matrix of the errors and the concentration

From Proposition 5.1, we see that the filtered density may be dominated by the second term $C'_t \alpha_t$, or in other words, tends to be more like a matrix Langevin distribution, if \mathbf{D} goes to infinity and $\mathbf{\Omega}$ goes to zeros. In the following, we see how the pair of \mathbf{D} and $\mathbf{\Omega}$ affect the filtering results.

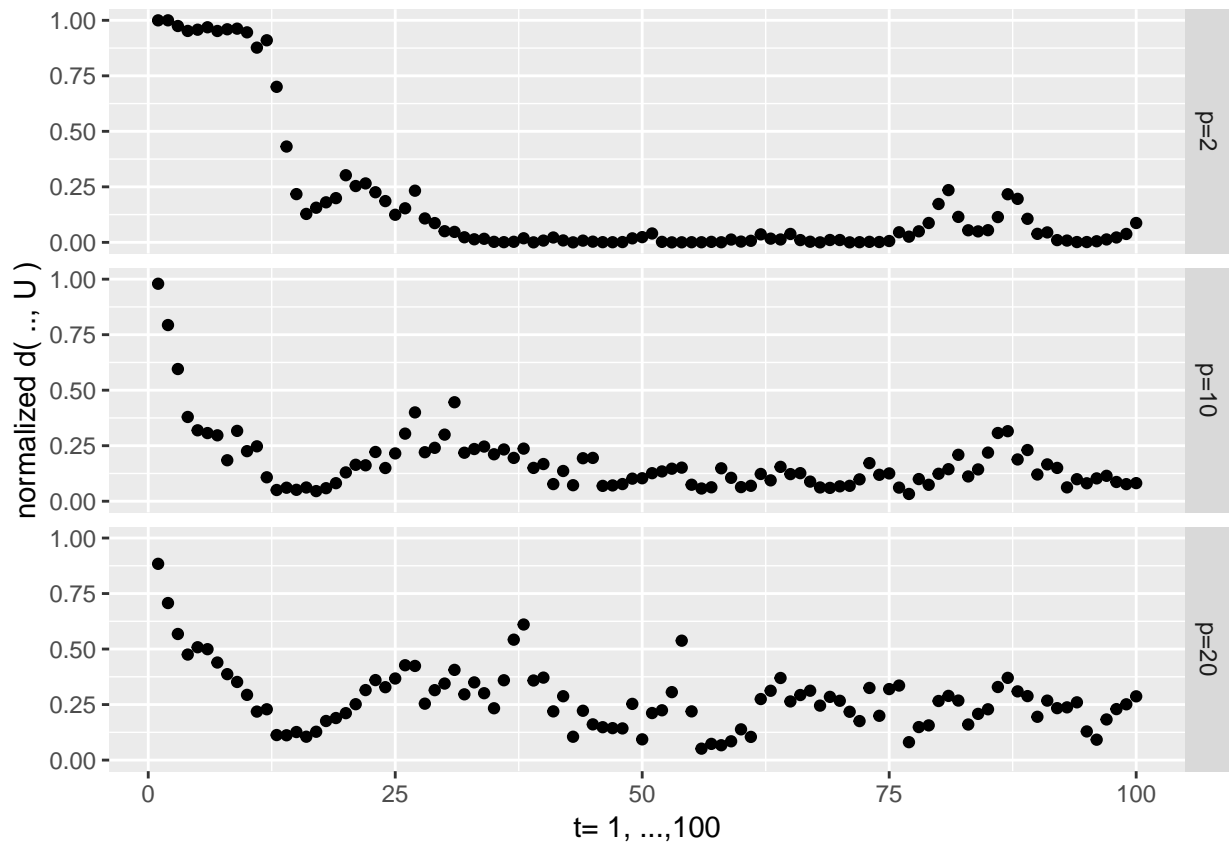
The basic setting is $T = 100$, $p = 2$, $r = 1$. The first three plots are for the case $\rho = 1$, the second three plots are for $\rho = .1$.



We see that even for a small $D = 5$, the results can be satisfactory with a small covariance matrix of the errors.

Wrong initial values

Now we consider the setting $T = 100$, $r = 1$, $\rho = 0.1$ and $D = 50$. In the following experiment, we look into the effect of wrongly specified initial values for the filtering algorithm. We choose the initial value $\mathbf{U}_0 = -\boldsymbol{\alpha}_0$ where $\boldsymbol{\alpha}_0$ is the true value, which cannot be worse as $-\boldsymbol{\alpha}_0$ is the furthest point away from $\boldsymbol{\alpha}_0$.



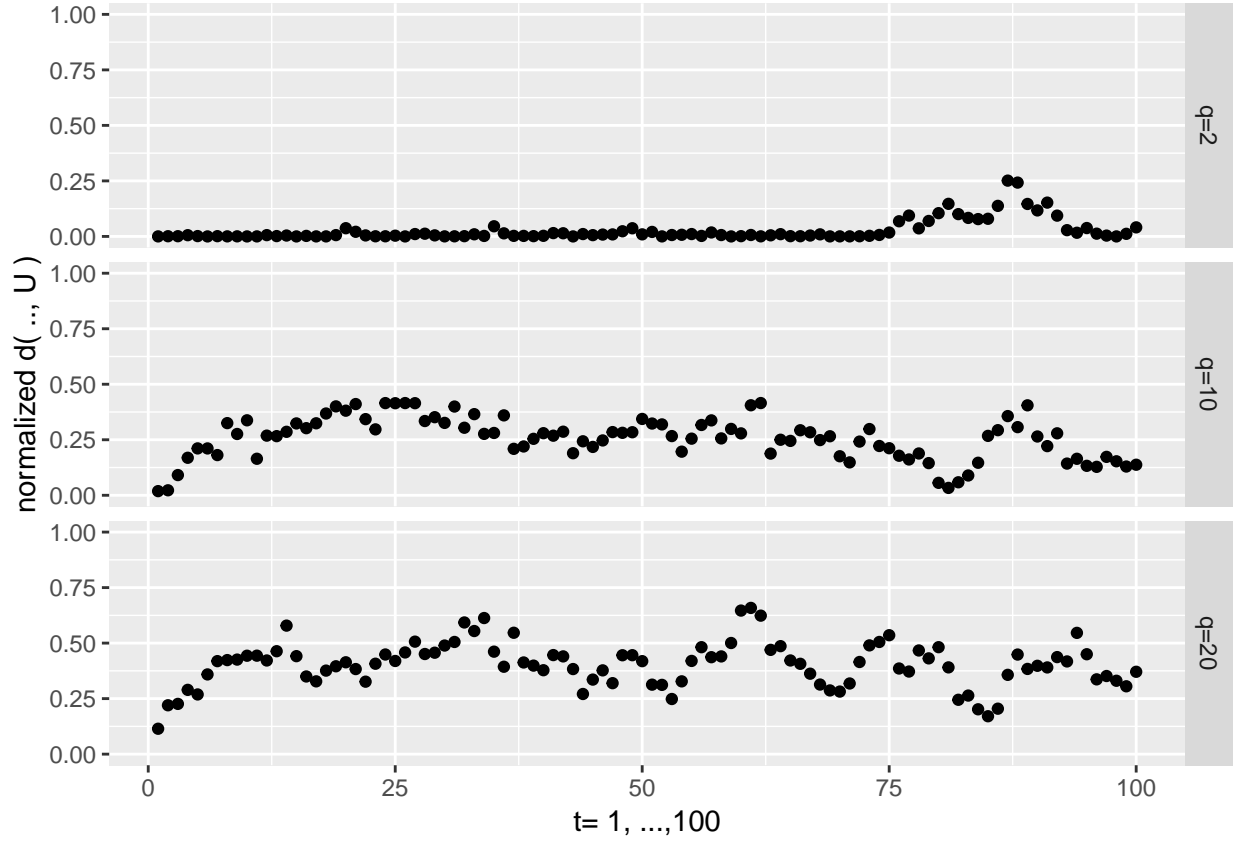
We see that all of the three modal orientations go back to the correct position though the lowest plot is not stable due to the high dimension.

Simulation Design for model two

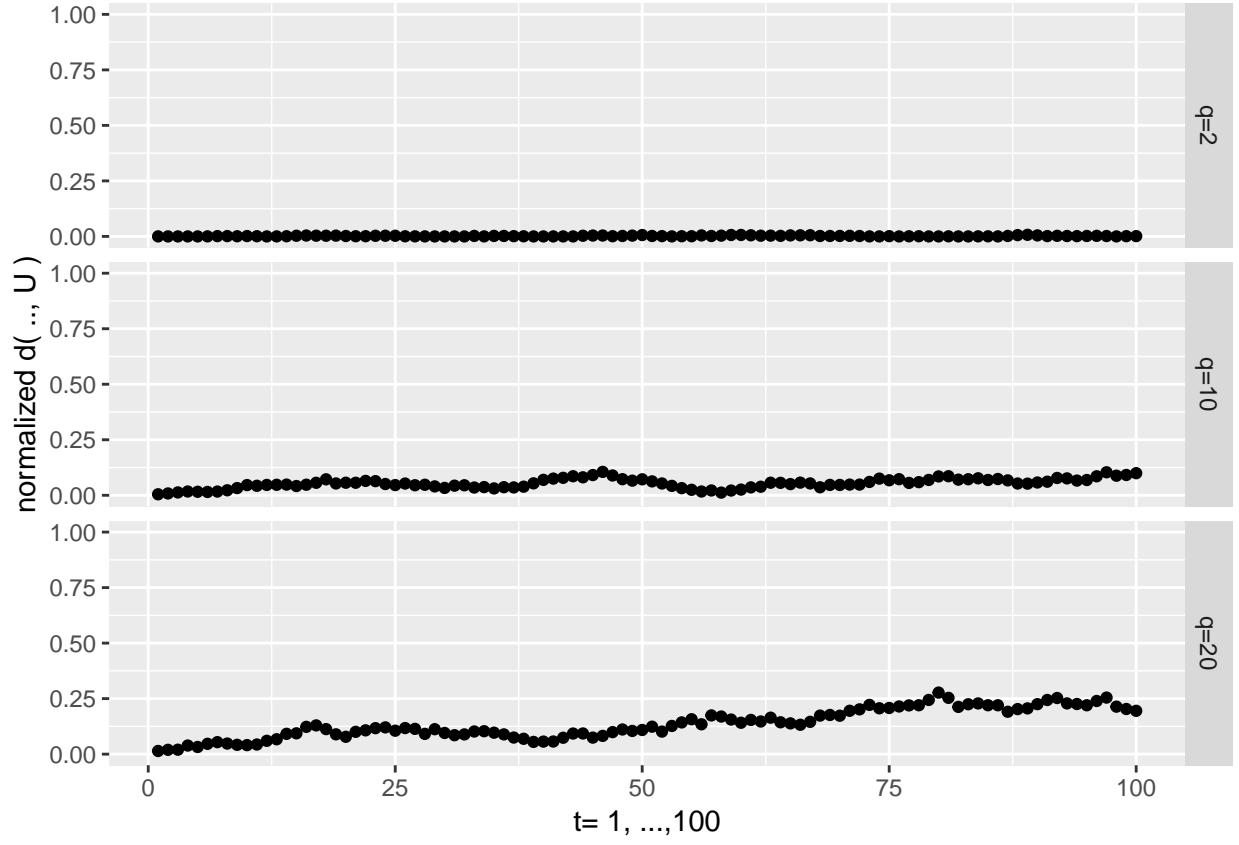
We use the same settings as those of model one, but now $p = 3$ and $q_1 \in \{2, 10, 20\}$.

Different q_1 and r with low and high D

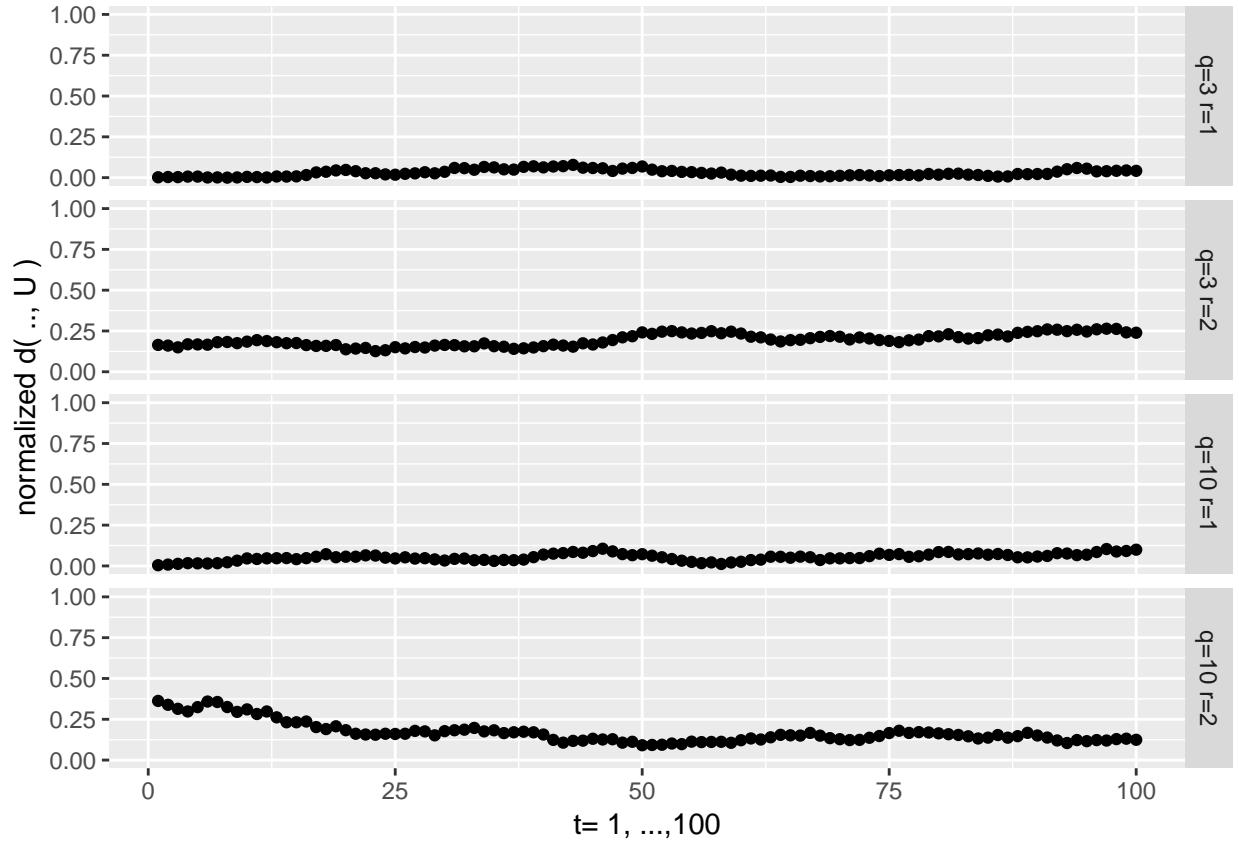
In this experiment, we consider the setting $T = 100$, $r = 1$, and $\rho = 0.1$. For fixed $D = 50$, we compare the filtering results via different values of dimension q_1 .



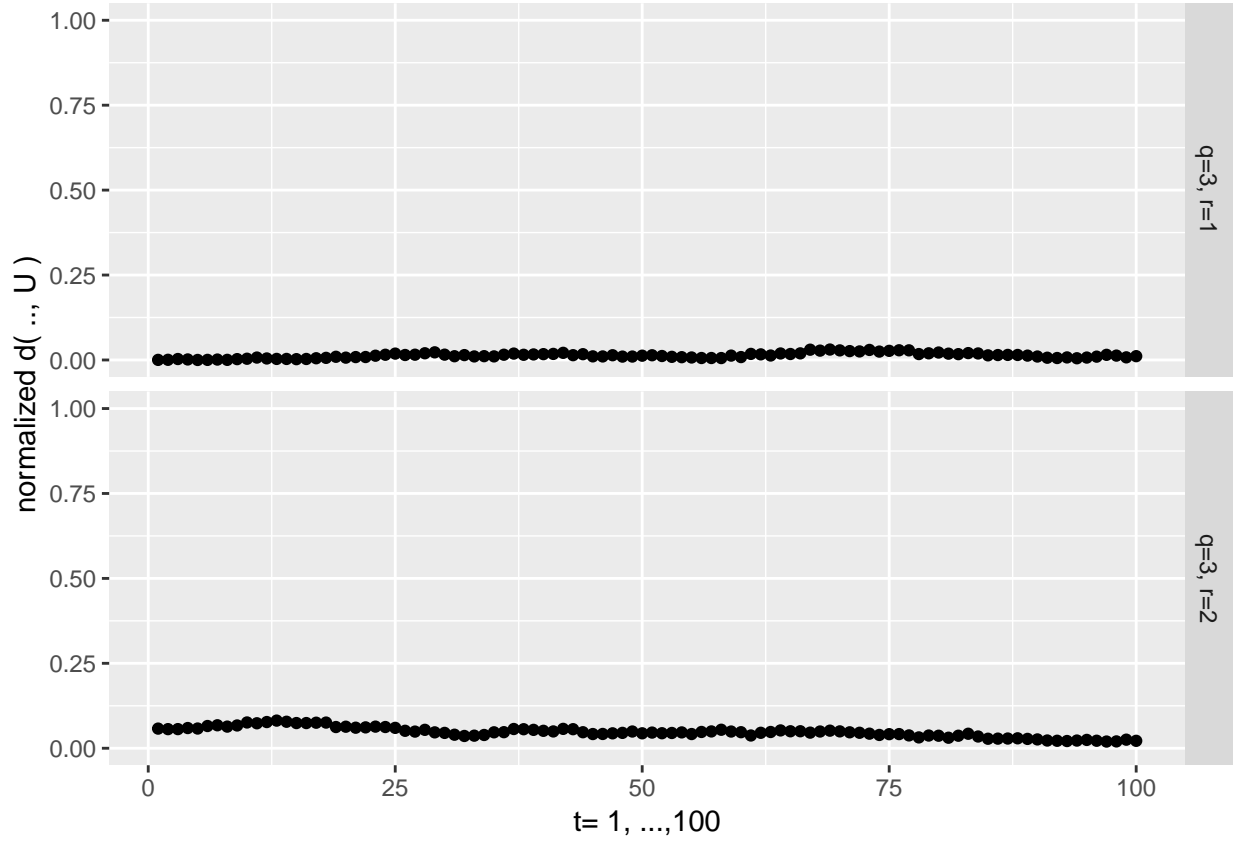
Then we consider the cases when $D = 500$.



Then we turn to investigate the impact of the rank r on the filtered distribution. In the following we use the setting $T = 100$, $\rho = 0.1$ and $D = 500$.

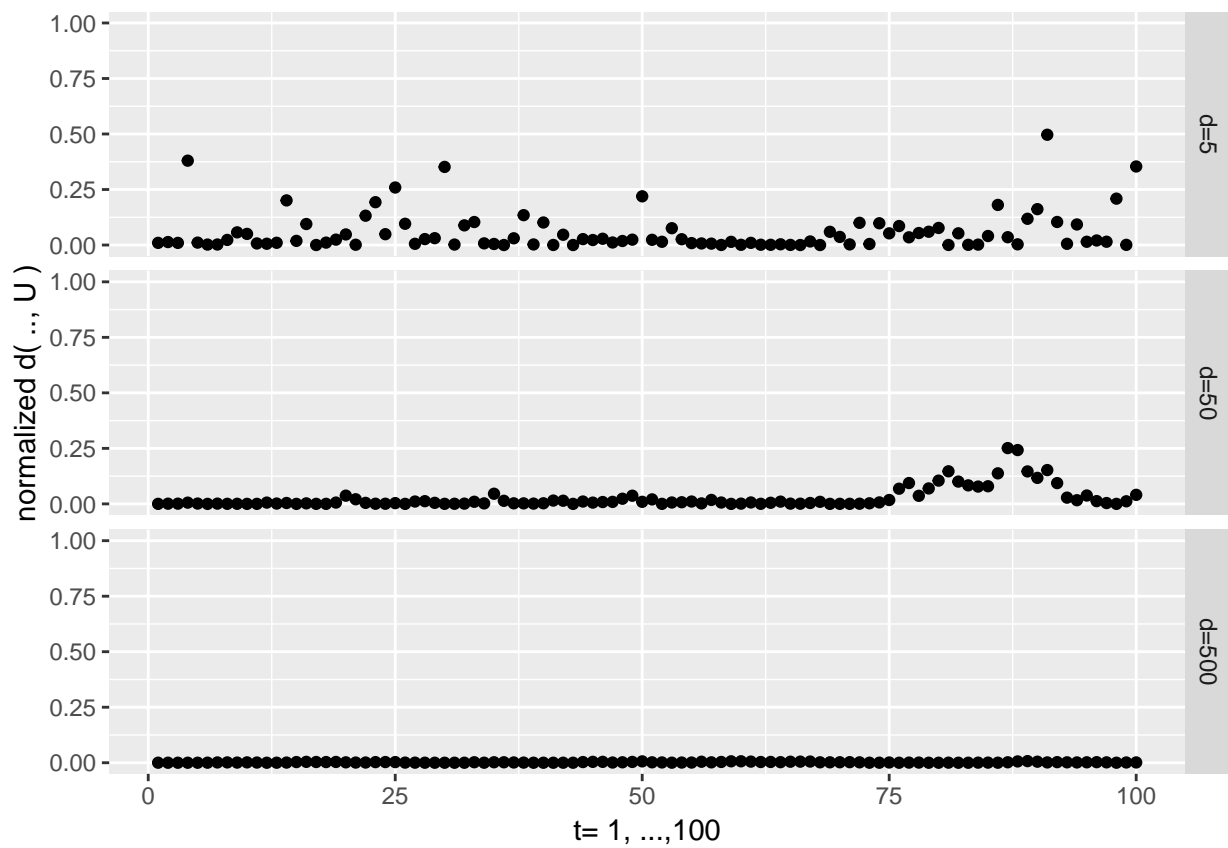
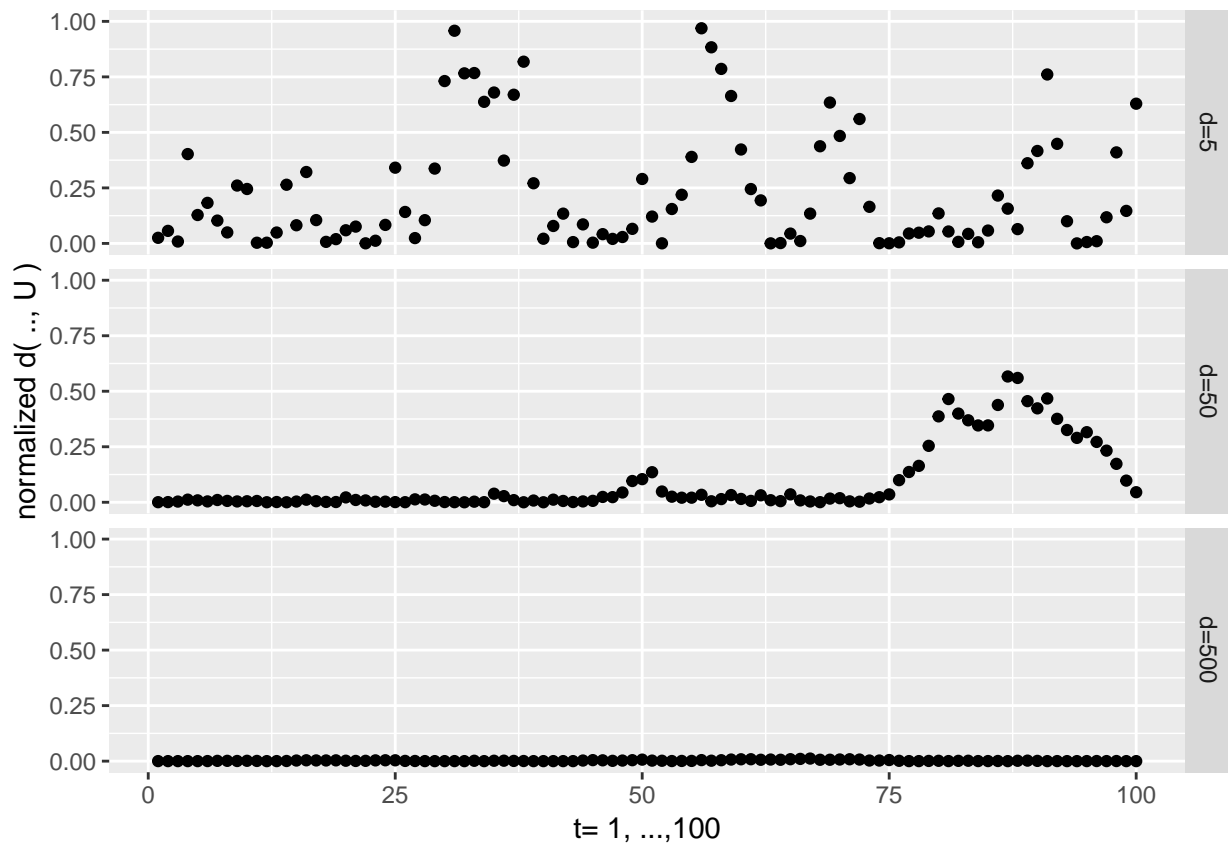


The following case with $D = 800$ shows that a bigger concentration D is the remedy for this.



The covariance matrix of the errors and the concentration

The basic setting is $T = 100$, $q_1 = 2$, $r = 1$. The first three plots are for the case $\rho = 1$, the second three plots are for $\rho = .1$.



Wrong initial values

Now we consider the setting $T = 100$, $r = 1$, $\rho = 0.1$ and $D = 50$. In the following experiment, we look into the effect of wrongly specified initial values for the filtering algorithm. We choose the initial value $\mathbf{U}_0 = -\beta_0$ where β_0 is the true value, which cannot be worse as $-\beta_0$ is the furthest point away from β_0 .

