EL2320 - Applied Estimation
Lab I
Alexandros Filotheou

1 Part I

1.1 Question 1

Since

$$\hat{x}_t = g(u_t, x_{t-1}) + \epsilon_t \tag{1}$$

and

$$z_t = h(x_t) + \delta_t \tag{2}$$

u_t models a part of the dynamics between x_{t-1} and x_t. So, it is an action that carries information about the change of state in the environment. For example, u_t can be the setting of the desired temperature in a room.

On the other hand, x_t is the environmental state at time t and depends only on the state at the previous timestep x_{t-1} and the control u_t at time t. As an example, x_t can be the actual temperature of a room.

z_t is the measurement on x_t and depends only on the environmental state x_t at time t. For example z_t can be the reading on a thermometer, when trying to estimate the true temperature of a room.

1.2 Question 2

The uncertainty in the belief during a Kalman filter update step is given by

$$\Sigma_t = (I - K_t C_t)\bar{\Sigma}_t \tag{3}$$

where

$$K_t = \bar{\Sigma}_t C_t^T (C_t \bar{\Sigma}_t C_t^T + Q_t)^{-1} \tag{4}$$

If we plug equation 4 into equation 3 and use equation the Sherman/Morrison formula, then

$$\Sigma_t = \bar{\Sigma}_t - \bar{\Sigma}_t C_t^T (C_t \bar{\Sigma}_t C_t^T + Q_t)^{-1} C_t \bar{\Sigma}_t = (\bar{\Sigma}_t^{-1} + C_t^T Q_t^{-1} C_t)^{-1} \tag{5}$$

where we know that $C_t^T Q_t^{-1} C_t$ is positive semi-definite. Hence, the uncertainty in the belief during an update cannot increase.

1.3 Question 3

From lecture's 5 notes:

$$\mu_t = \bar{\mu}_t + K_t(z_t - C_t\bar{\mu}_t) = (I - K_t C_t)\bar{\mu}_t + K_t z_t = (I - W)\bar{\mu}_t + W\mu_z \tag{6}$$

where $W = \Sigma_t C_t^T Q_t^{-1} C_t$ and $C_t\mu_z = (z_t - \bar{z}_t + C_t\bar{\mu}_t)$. Hence, essentially the relation between Σ_t and Q_t decides the weighting between the measurements and the belief.

1.4 Question 4

If Q is large, then the Kalman Gain will be small and the filter would take more time to converge.

1.5 Question 5

For the measurements to have an increased effect, the Kalman Gain should be large, since $\mu_t = \bar{\mu}_t + K_t(z_t - C_t\bar{\mu}_t)$. For the Kalman Gain to be large, Q_t, that is, the uncertainty regarding the measurements should be small.

1.6 Question 6

The belief uncertainty during prediction is given by equation 7:

$$\bar{\Sigma}_t = A_t\Sigma_{t-1}A_t^T + R_t \tag{7}$$

If $A_t \geq I$ then $A_t\Sigma_{t-1}A_t^T$ increases, and noise (R_t) is added. Hence, in general, $\bar{\Sigma}_t$ increases but it depends on A_t and R_t.

1.7 Question 7

1.8 Question 8

If there is no a priori knowledge about the distribution, then the mean of the Kalman Filter can be the MLE of the Gaussian belief. If there is a priori knowledge about the distribution, then a Kalman Filter can be a MAP if the distribution is Gaussian.

1.9 Question 9

The EKF revokes the assumption that a state is a linear function of its previous state, and applies the KF by approximating the non-linear function that is tangent to g at the mean of the posterior. Furthermore, $A_t x_{t-1} + B_t u_t$ is replaced by $g(u_t, x_{t-1})$, $C_t x_t$ by $h(x_t)$, A_t by G_t and C_t by H_t.

1.10 Question 10

No. It depends on the the local nonlinearity of the g function.

1.11 Question 11

We could increase the model uncertainties by increasing the covariances Q_t and R_t.

1.12 Question 12

2 Part II

2.1 Question 1

In this case $\epsilon_k : 2 \times 1$, δ_k : scalar. In the general case, ϵ_k will be $N \times 1$, where N is the number of state variables, and δ_k will be $M \times 1$, where M is the number of variables the KF tries to estimate.
A scalar Gaussian is characterized by a mean value μ and a variance σ^2. A white Gaussian has $\mu = 0$. In the general case, $\boldsymbol{\mu}$ is a single-column matrix and the scalar variance is replaced by a covariance matrix $\boldsymbol{\Sigma}$. In this case, a white Gaussian has $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma}$ is a diagonal matrix because the noise in each state variable is independent of one another.

2.2 Question 2

Variable	Usage
x	The true state of the system.
$\hat{x}$	The estimate of the true state of the system by the KF.
P	Estimate error covariance matrix.
G	Identity matrix for dimensionality consistency.
D	Identity matrix for dimensionality consistency. Scalar here.
Q	Measurement noise variance.
R	Process noise covariance matrix.
$wStdP$	The actual (simulated) standard deviation of the noise in position.
$wStdV$	The actual (simulated) standard deviation of the noise in velocity.
$vStd$	The actual (simulated) standard deviation of the noise in position estimation.
u	Control signal, the acceleration.
PP	Estimate error covariances over time.

2.3 Question 3

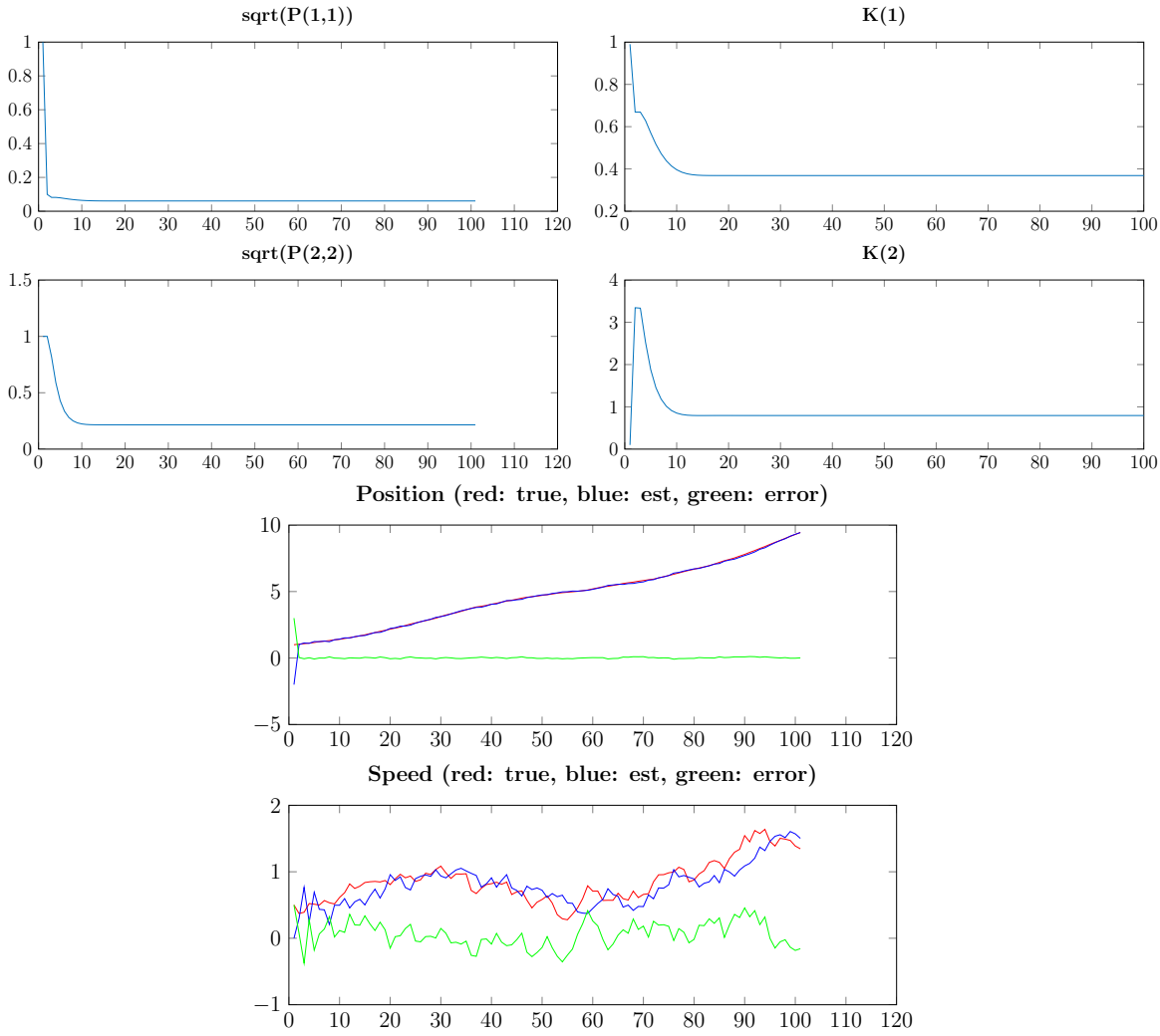

Figure 1: Estimation error, covariance and Kalman gain. Q, R default.

2.3.1 One change in the default parameters at a time

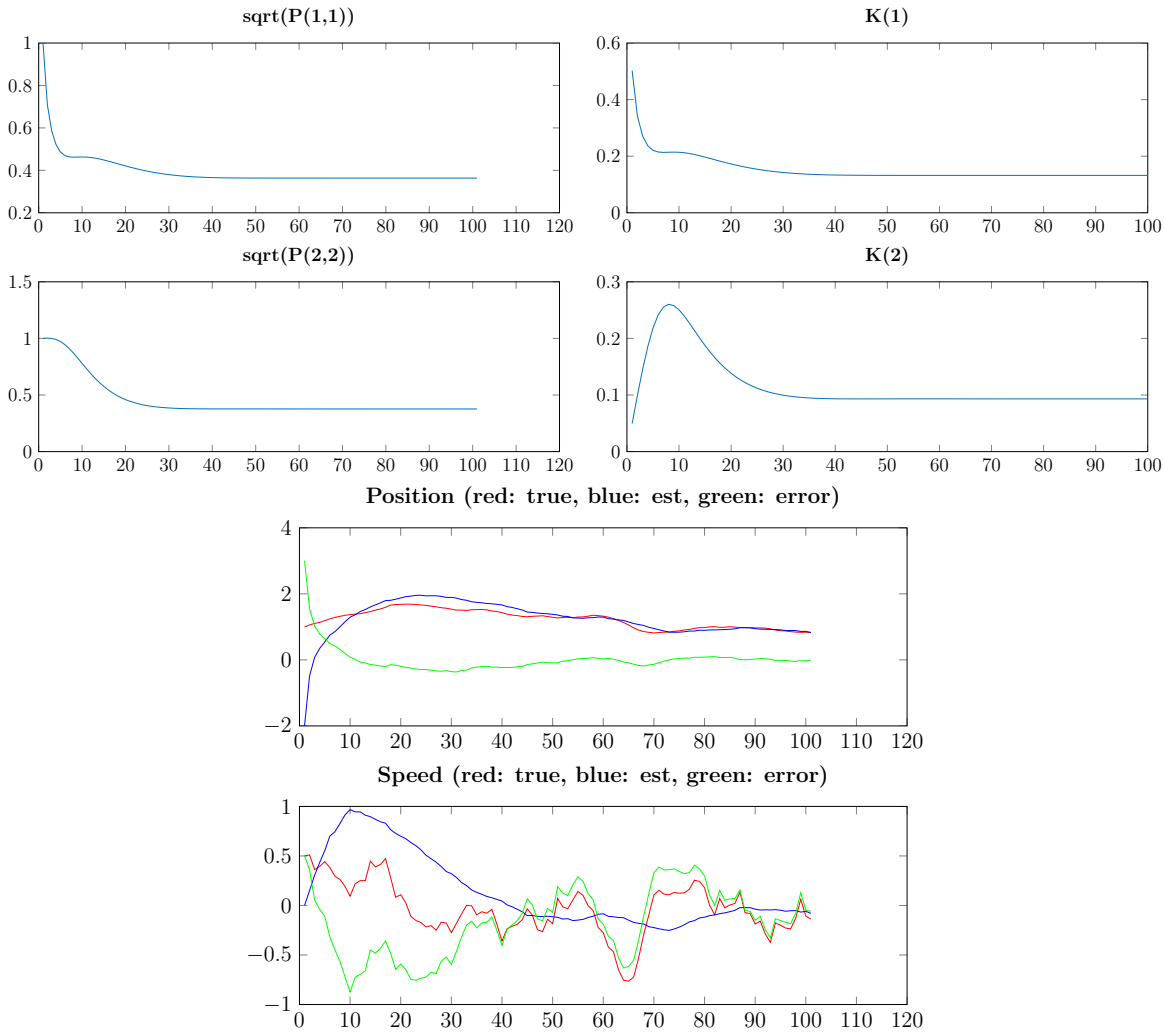

Figure 2: $Q \times 100, R$ default.

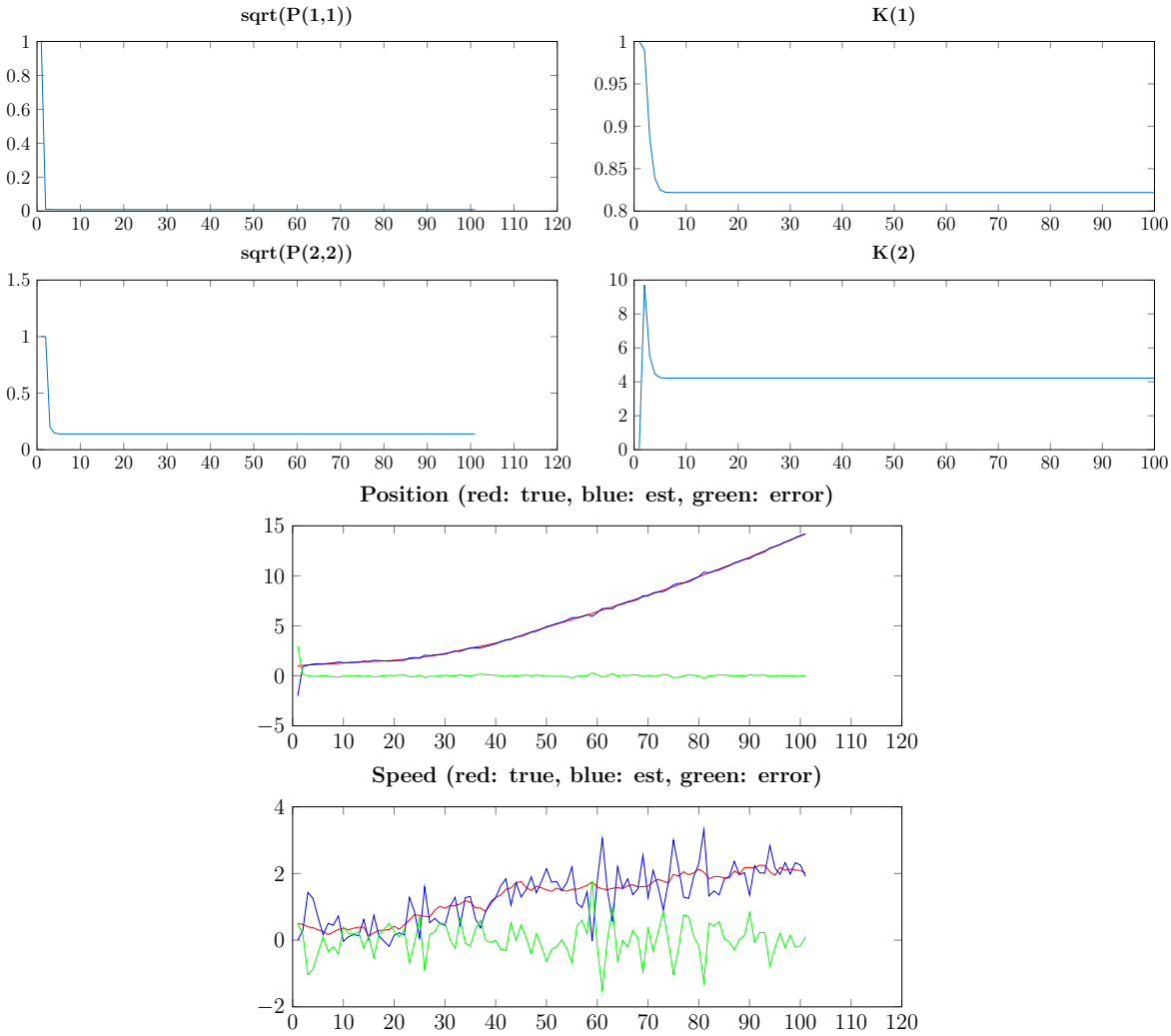

Figure 3: $Q/100, R$ default.

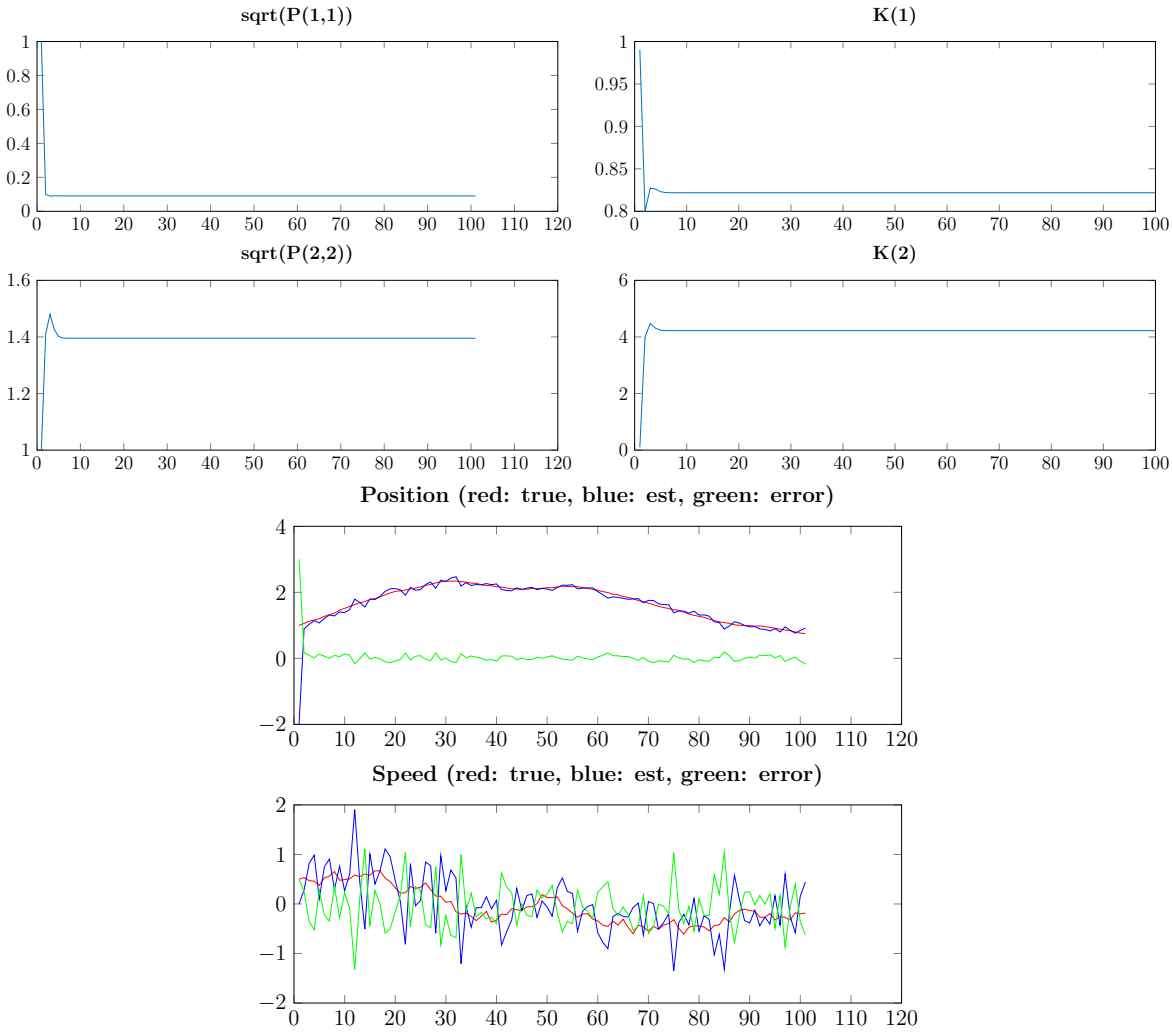

Figure 4: Q default, $R \times 100$.

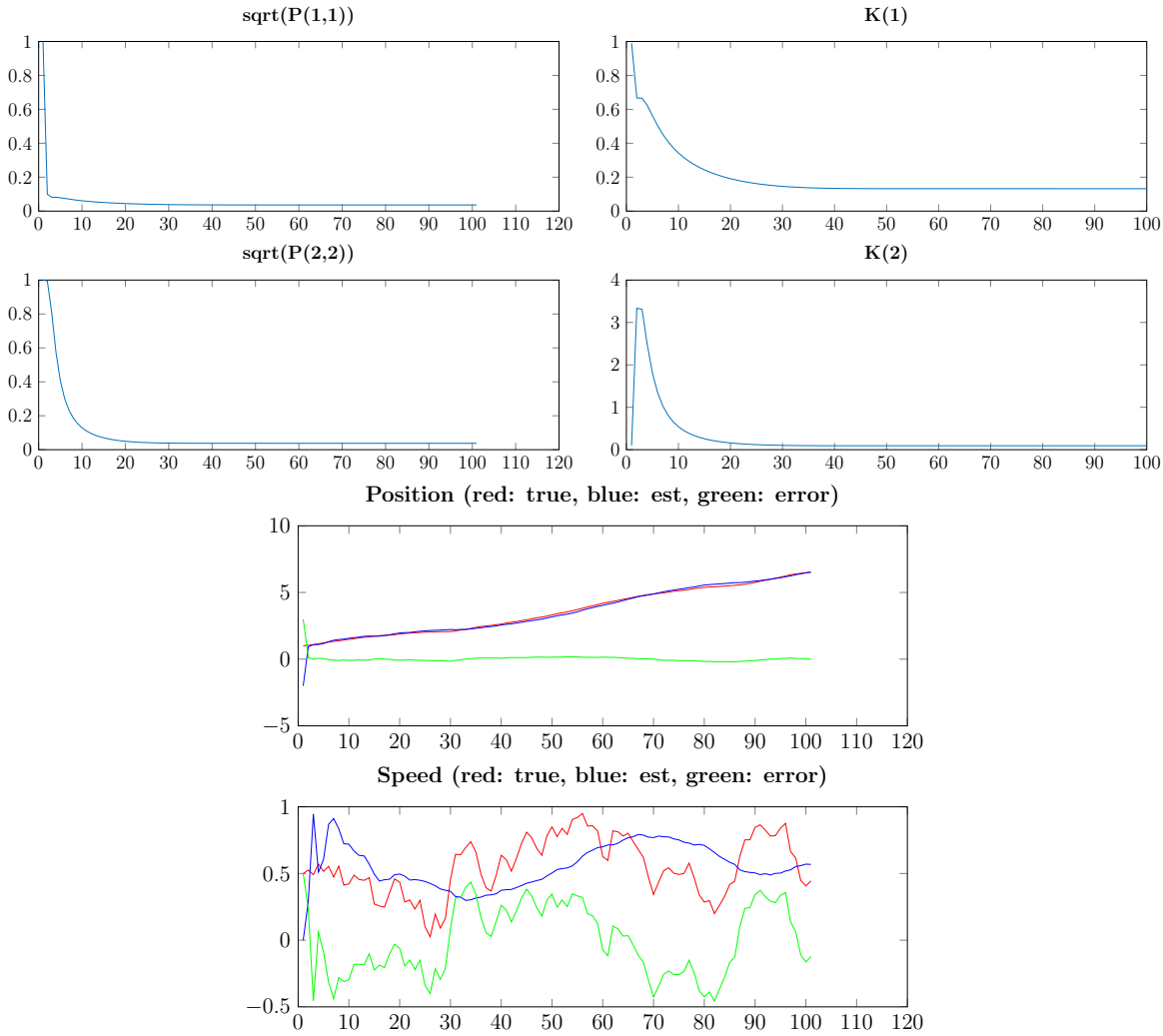

Figure 5: Q default, $R/100$.

2.3.2 Change in the default parameters at the same time

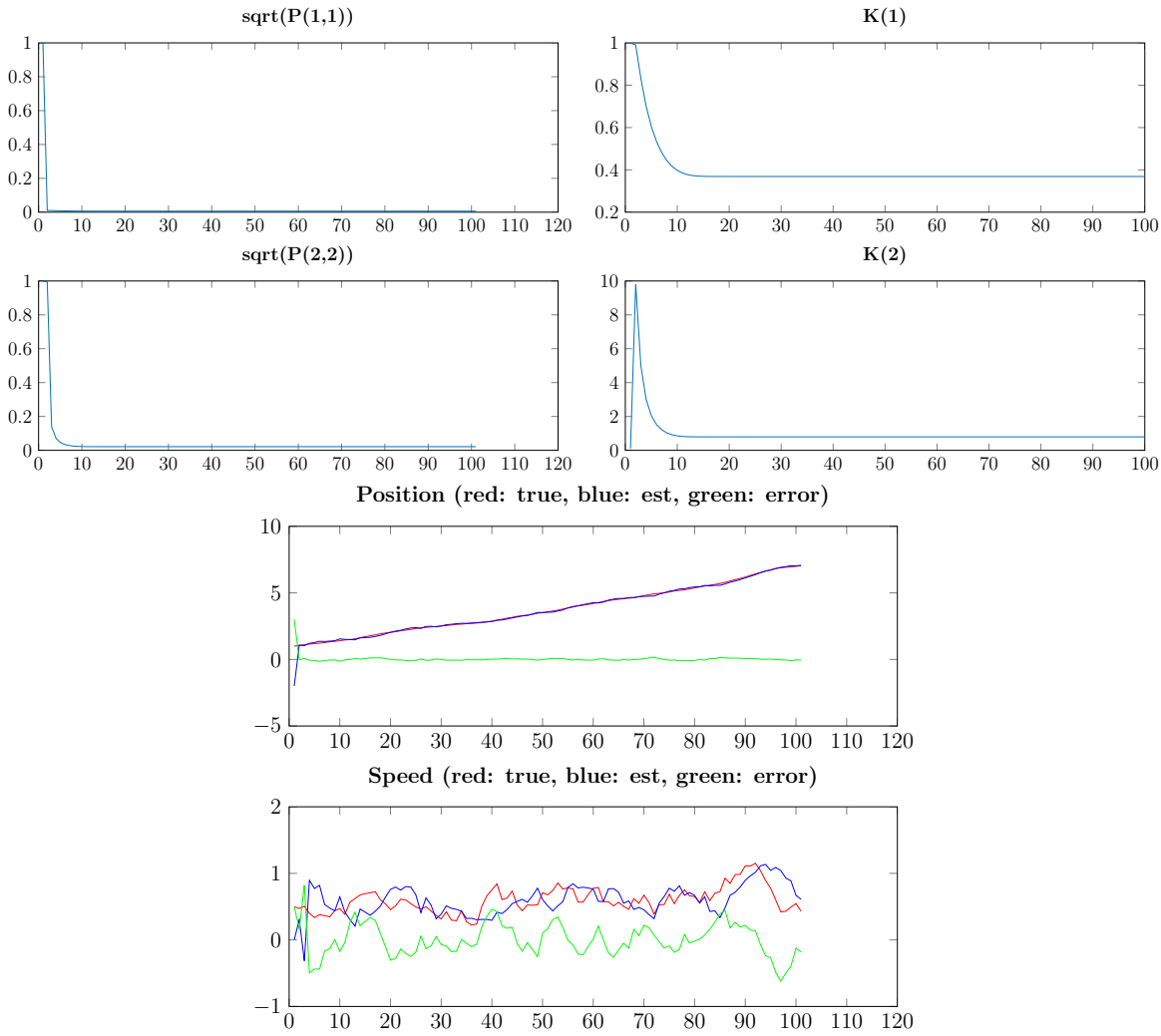

Figure 6: $Q/100, R/100$

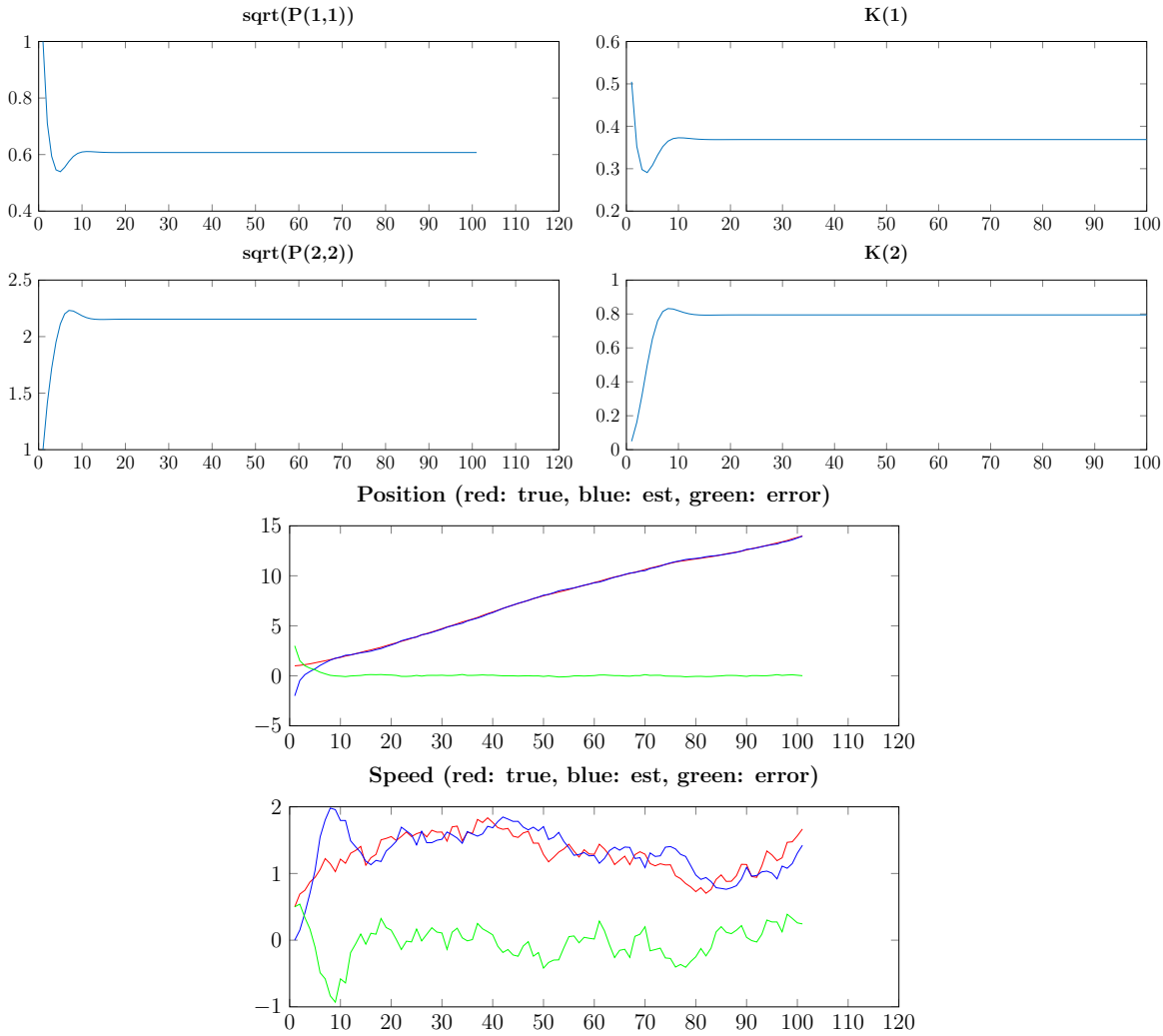

Figure 7: $Q \times 100, R \times 100$.

2.3.3 Conclusion

Comparing figure 2 with figure 1 we see that when the measurement noise increases, the covariance of the error in estimating the true position and velocity of the car increases, and the Kalman gain decreases, meaning that the measurements are trusted less and less. In theory, this means that the KF takes more time to converge, and here the plots verify this. In the opposite case, when the measurement noise decreases, we would expect that the effects are exactly the opposite. This is exactly the case, as it is illustrated in figure 3. The measurements are given a higher weight, as the Kalman gain suggests, since they are more reliable. Converging time decreases, along with the covariance of the estimation errors. Notably, a low measurement error variance causes the estimation of the true position and speed of the car to vary in a high degree, as opposed to the previous case, where the estimation was much smoother. With regard to process noise, when its variance increases we would expect the measurement model to be weighted more and more, since the estimation of the true state should be trusted less and less. This, of course, means a higher Kalman gain and a higher variance of the modelled estimation error. These remarks are verified in

comparing figures 2 and 4. Notably, the convergence time is not affected and, again, because the measurements are weighted more, the estimation of the position and speed of the car change quickly over time. In the opposite case, the KF's prediction is weighted more than the measurements, and this means that both the error covariance and the Kalman gains will be lower. Notably, as it can be seen in figure 5 the estimate is smoother, but it takes more time for the KF to converge. When both process and measurement noise are high, their modelled covariance is high, and the Kalman gains are low, since the measurements cannot be trusted. Notably, when both are low, the Kalman gains converge to the same values, but through different paths. In this case, convergence is quicker and the modelled covariance is, as expected, lower than having decreased the variance of each error separately.

2.4 Question 4

2.4.1 Changing P_0

With respect to the initial conditions for the modelled covariance, if P is large, then the uncertainty about the true state is greater. This forces the Kalman gain to rise sharply, because now the measurements are to be trusted more in order to compensate for the KF's pessimistic estimation, before converging in approximately the same time as in the original case. However, when the initial P is small, the KF trusts the estimates more than the measurements, and the Kalman gains rise but in a slower manner than before. The time to converge in this case increases. In both cases the estimate error for the position is nearly unchanged, however with respect to the velocity, in the first case it varies heavily and in the second one it decreases. Figures 8 and 9 illustrate the effect of changing the initial conditions of P for $P_0' = P_0 \times 1000$ and $P_0' = P_0 \times 0.001$ respectively.

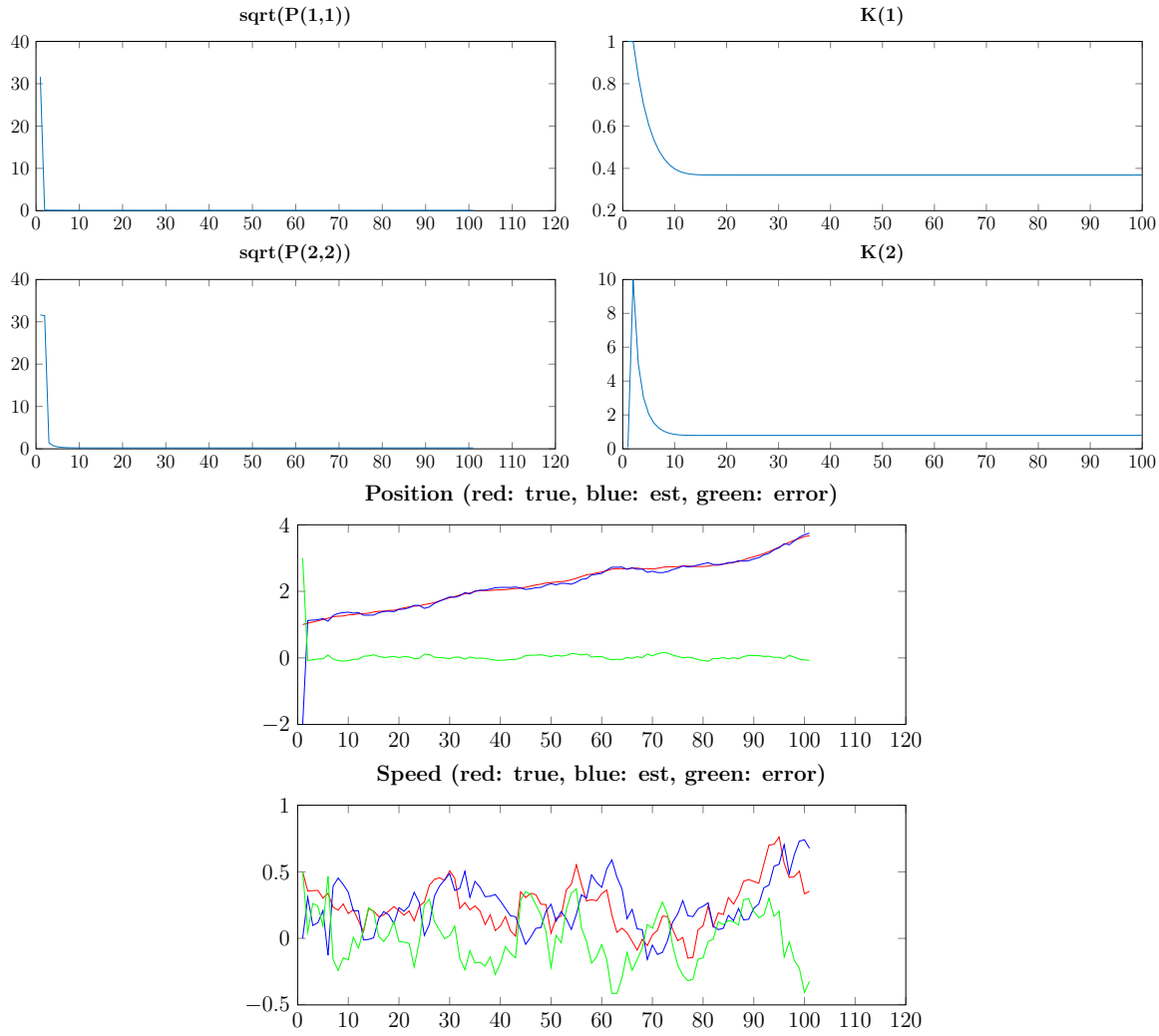

Figure 8: $P_0' = P_0 \times 1000$

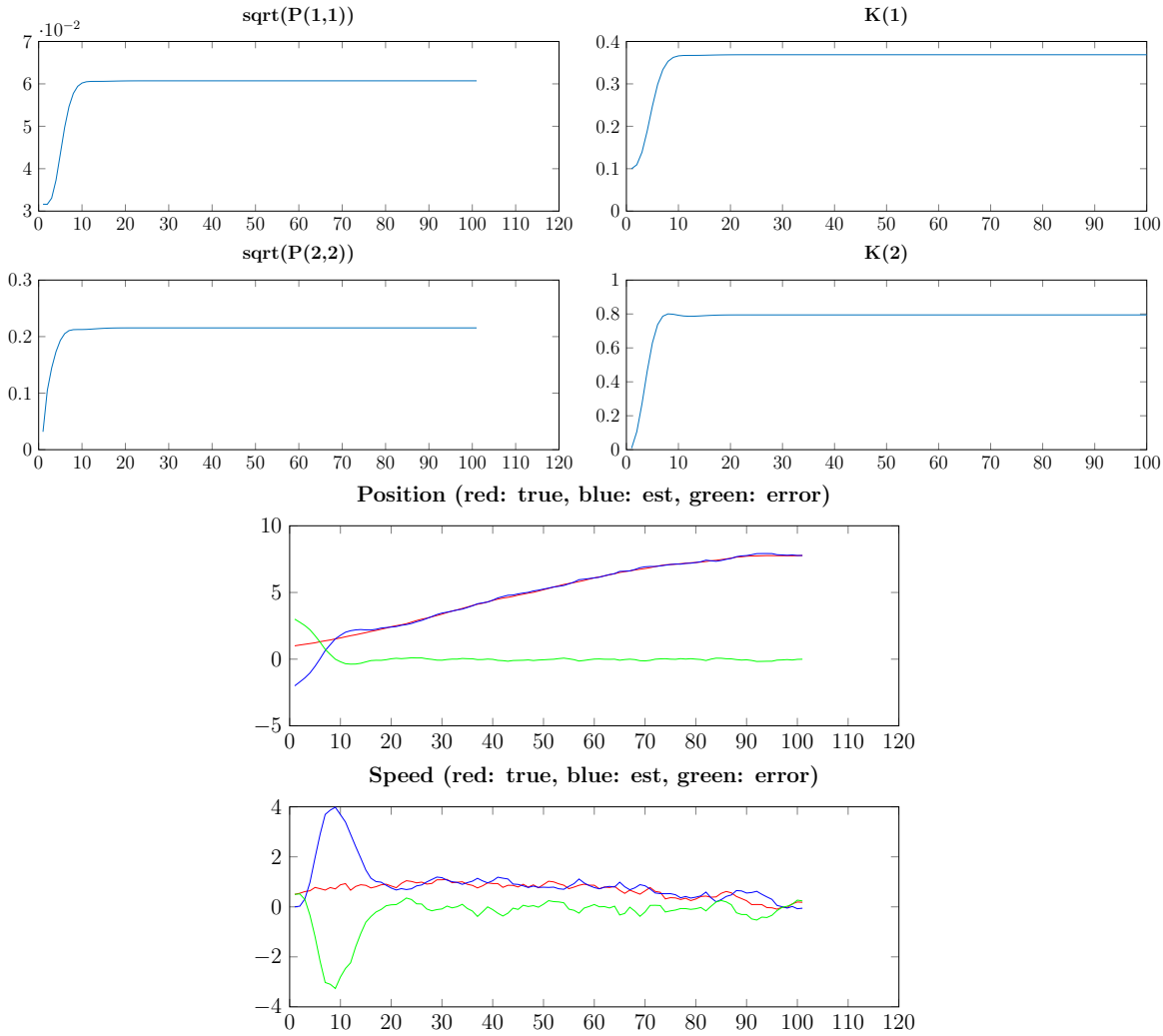

Figure 9: $P_0' = P_0 \times 0.001$

2.4.2 Changing $\hat{x}_0$

With regard to the initial estimate on the true state $\hat{x}$, it seems that the KF is robust. Changing it by $10^{\pm 3}$ has no effect in the rate of convergence, or the error of the estimates. This is reasonable, since changing the initial mean position and velocity of the car is irrelevant to the KF, since we assume linearity in the transitions between states.

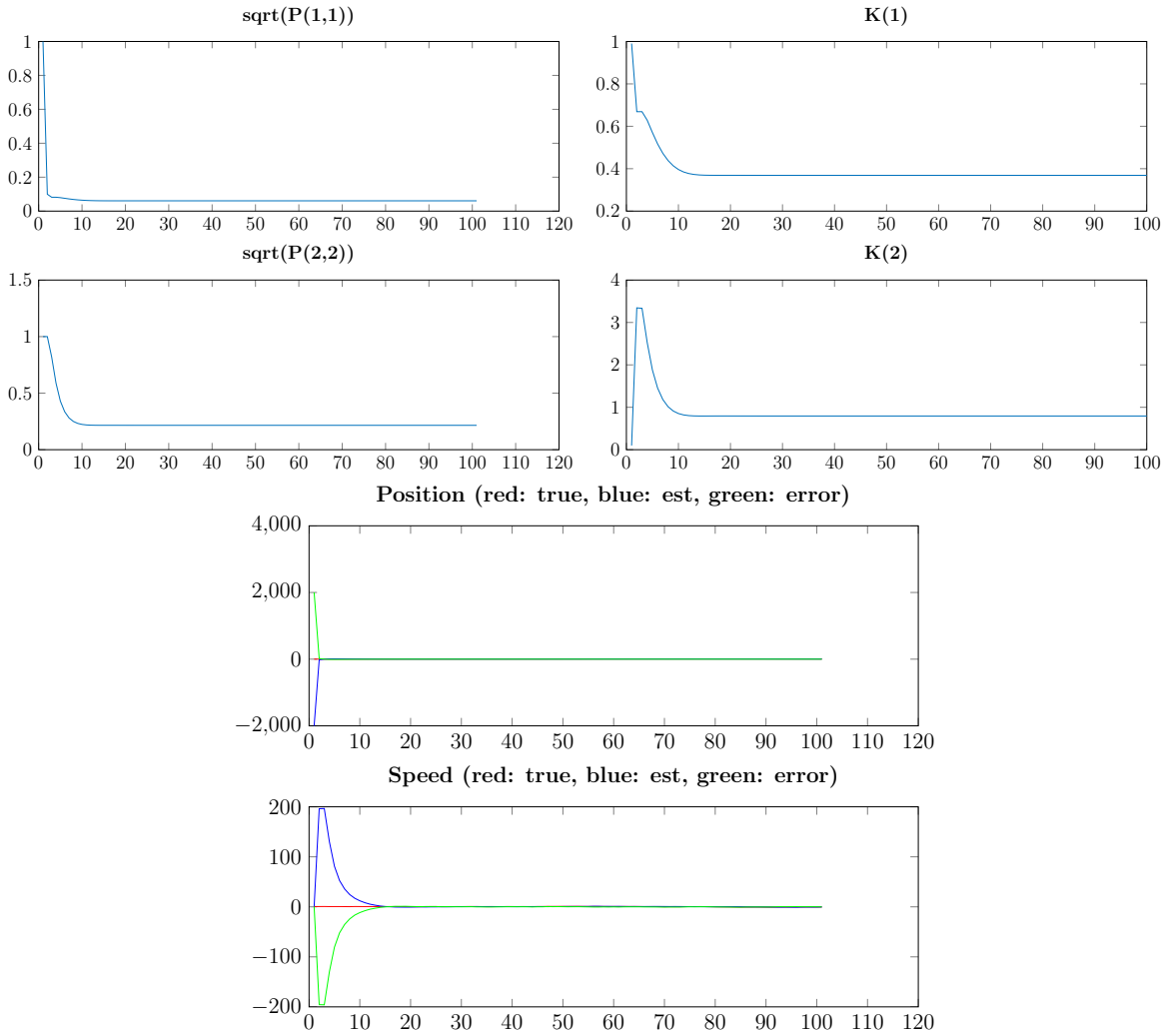

Figure 10: $\hat{x}'_0 = \hat{x}_0 \times 1000$

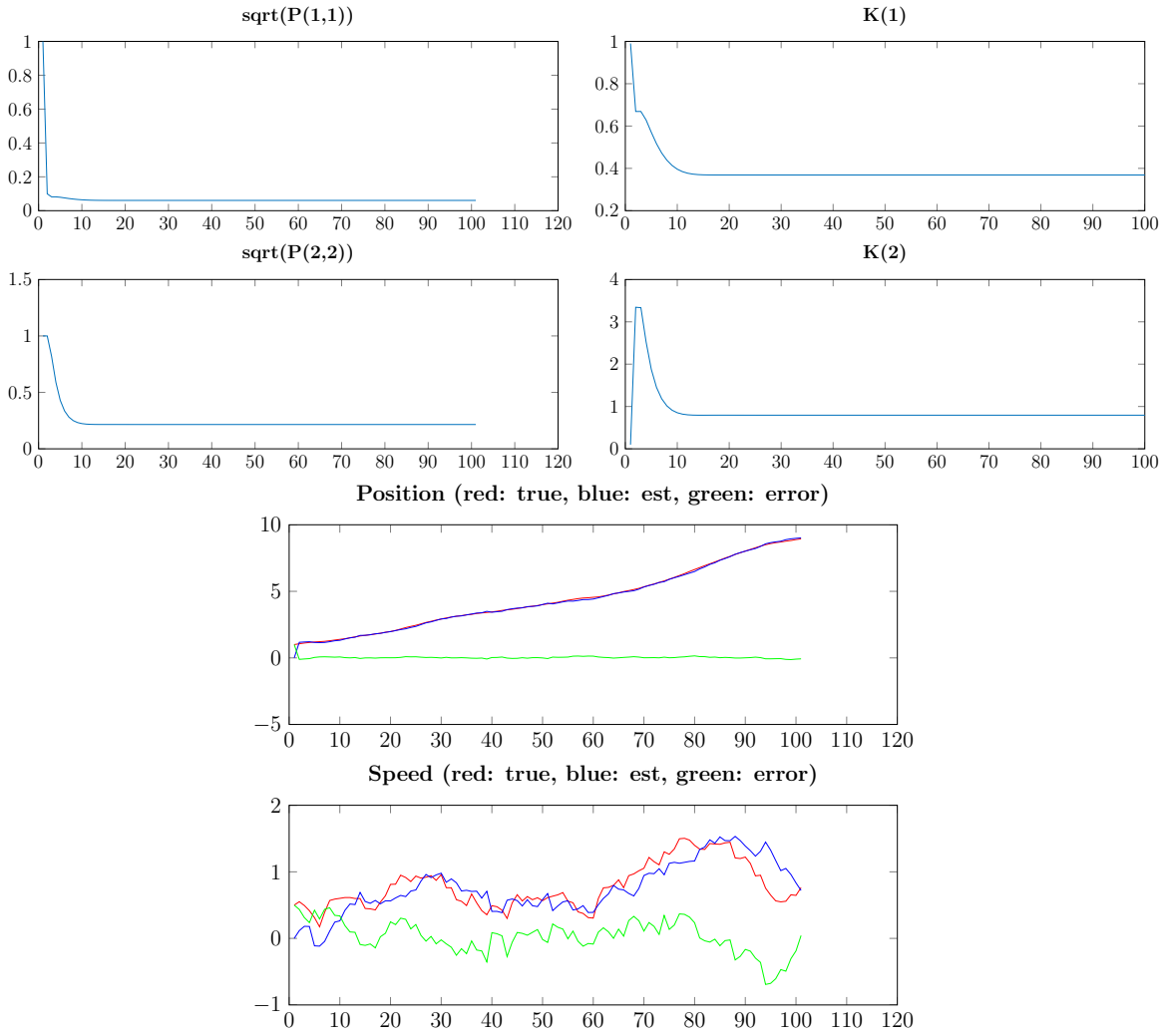

Figure 11: $\hat{x}'_0 = \hat{x}_0 \times 0.001$

3 Part III

3.1 Question 5

In equation 3 of the instructions, the prediction step is given by equation 8, while the update step is given by 9.

$$\overline{bel}(x_t) = p(x_t|u_{1:t}, z_{1:t-1}, \bar{x}_0, M) = \int p(x_t|u_t, x_{t-1}) bel(x_{t-1}) dx_{t-1} \tag{8}$$

$$bel(x_t) = p(x_t|u_{1:t}, z_{1:t}, \bar{x}_0, M) = \eta p(z_t|x_t, M) \overline{bel}(x_t) \tag{9}$$

Since the prediction step at time t refers to the belief about the state at time t before incorporating the measurements z_t, with regard to equation 2 in the instructions, the prediction step is given by factor in 10, while the update step is given by the factor in 11.

$$\int p(x_t|u_t, x_{t-1}) p(x_{t-1}|z_{1:t-1}, u_{1:t-1}, \bar{x}_0, M)) dx_{t-1} \tag{10}$$

$$\eta p(z_t|x_t, M) \int p(x_t|u_t, x_{t-1}) p(x_{t-1}|z_{1:t-1}, u_{1:t-1}, \bar{x}_0, M)) dx_{t-1} \tag{11}$$

3.2 Question 6

The measurements are independent of each other if-f the measurement noise is white Gaussian noise, which is in fact assumed here, as the value of the Jacobian H depends only on the value of $\hat{x}_t$ at time t. Then, the covariance matrix of the measurement noise distribution will be diagonal and all measurements will be uncorrelated with each other.

3.3 Question 7

Since δ_M is a probability, $0 \leq \delta_M \leq 1$. The inverse χ^2 is monotonously ascending, hence a larger value for δ_M will result in a larger λ_M threshold. The larger the threshold, the larger the Mahalanobis distance $D_M = (\overline{\nu}_t^i)^T (H_{t,j} \bar{\Sigma}_t (H_{t,j})^T + Q)(\overline{\nu}_t^i)$, which means that the smaller the uncertainty in measurement Q is. Hence, a large value for δ_M will result in having more confidence on the measurements and, hence, less outlier rejections. Conversely, a low value for δ_M will result in more rejections. As for the threshold λ_M, it follows the same reasoning. If we know that the measurements are reliable, then Q is small, and we can utilize a higher threshold. On the other hand, if the measurements are known to be spurious, then a lower λ_M threshold should be used, so as to take into account only the most reliable ones, which are the ones closest to the centroid of the Mahalanobis ellipse.

3.4 Question 8

If the first measurement to be incorporated is noisy, then the innovation will be non-zero and the estimated mean for that measurement will be shifted erroneously. The covariance $\bar{\Sigma}$ will be incorrectly reduced, which will result in a higher Kalman gain in the next timestep.

Then, $S_{t,j}$ will be reduced, and the threshold on the Mahalanobis distance will increase, resulting in possibly valid measurements being rejected as outliers.

3.5 Question 9

In Alg 4 of the instructions, the first three assignments are not dependant on the observation i, however each one is computed $|z_t|\ times|M|$ times, where $|z_t|$ is the number of measured parameters and $|M|$ the number of landmarks. Hence, these computations are redundant and could be made beforehand in a separate loop, so as to reduce $|z_t|\ times|M|$ to just $|M|$.

3.6 Question 10

Matrix $\mathbf{h_t}$ is of 2×1 dimensions. Hence, the measurement matrix $\mathbf{z_t}$ will also be of 2×1 dimensions. The same applies for the innovation $\bar{\nu}_t^k$. $(\bar{\nu}_t^k)^T$ is of 1×2 dimensions, and n of those vectors are stored in $(\bar{\nu}_t)^T$. Hence, $\bar{\nu}_t$ is of $2n \times 1$ dimensions, where n is the number of inliers. Analogously, $\bar{H}_t$ is of $2n \times 3$ dimensions. In the sequential update method, $\bar{\nu}_t$ and $\bar{H}_t$ have 2×1 and 2×3 dimensions respectively. This is to be expected, since in the batch update method all the observations are grouped together, whereas in the sequential one, each observation is treated separately. The higher dimensionality in the former case will result in more computational costs than in the latter.