Kalman Filters

Guillaume Frèche

Version 1.0

Keywords: Kalman filters

In this document, we introduce a generalization of linear regression: **Kalman filters**, we derive their update equations and the corresponding algorithm, and we apply them on some examples.

1 Filter description

In the document about linear regression, we have introduced the following model:

$$\forall i \in [\![1, n]\!] \qquad y_i = \sum_{j=1}^{m} w_j x_{i,j} + e_i = \mathbf{x}_i^\mathsf{T} \mathbf{w}_m + e_i$$

where $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,m})^\mathsf{T} \in \mathbb{R}^m$ and $y_i \in \mathbb{R}$ are given or observed data, e_i is a realization of a zero-mean observation random noise $E \sim \mathcal{N}(0, \sigma_e^2)$ uncorrelated with data, and $\mathbf{w}_m = (w_1, \ldots, w_m)^\mathsf{T}$ is a vector of hidden weights that we want to determine. In this document, we replace $\mathbf{w}$ by $\boldsymbol{\theta}$, denoting some hidden state that we are trying to determine, and output $\mathbf{y}_k$ can now be a vector instead of a scalar. Thereby, we transform the linear regression equation into a so-called **observation equation**:

$$\mathbf{y}_k = \mathsf{H}_k \boldsymbol{\theta}_k + \mathbf{u}_k \tag{1}$$

where $\mathbf{u}_k$ is a zero-mean random noise representing the uncertainty of the observation measure. The generalization brought by Kalman filters is that hidden parameters can vary over time according to a **state equation**:

$$\boldsymbol{\theta}_{k+1} = \mathsf{F}_k \boldsymbol{\theta}_k + \mathbf{v}_k \tag{2}$$

where $\mathbf{v}_k$ is a zero-mean random noise representing the quality of the evolution model. Therefore, Kalman filters are entirely described by:

$$\begin{cases} \boldsymbol{\theta}_{k+1} & = \mathsf{F}_k \boldsymbol{\theta}_k + \mathbf{v}_k \\ \mathbf{y}_k & = \mathsf{H}_k \boldsymbol{\theta}_k + \mathbf{u}_k \end{cases}$$

Matrices F_k and H_k are fixed and given by the model. We suppose that $\mathbf{u}_k$ and $\mathbf{v}_k$ are uncorrelated second-order white noises with respective covariance matrices $\mathsf{R}_k^u = E\left(\mathbf{u}_k \mathbf{u}_k^\mathsf{T}\right)$ and $\mathsf{R}_k^v = E\left(\mathbf{v}_k \mathbf{v}_k^\mathsf{T}\right)$. The purpose of Kalman filters is to provide at any time $k \in \mathbb{N}^*$ an estimate $\widehat{\boldsymbol{\theta}}_k$ of hidden state variable $\boldsymbol{\theta}_k$. These filters work in two steps:

- a **prediction stage** where we estimate prior state variable $\widehat{\boldsymbol{\theta}}_{k+1|k}$ given previous observations $\mathbf{y}_1, \ldots, \mathbf{y}_k$;
- an **update stage** where we estimate posterior state variable $\widehat{\boldsymbol{\theta}}_{k+1|k+1}$, given the prior $\widehat{\boldsymbol{\theta}}_{k+1|k}$ and the new observation $\mathbf{y}_{k+1}$.

Remarks:

- There are also non-linear Kalman filters, for which matrices F_k and H_k are replaced by non-linear mappings. These filters are out of the scope of this document.
- The equations above describe Kalman filters with only evolving hidden states and no input or command. This case can be treated by adding a term $\mathsf{G}_k\mathbf{x}_k$ in Equation (2), where $\mathbf{x}_k$ denotes the input. The derivation in the next subsection can be adapted to this case.

2 Prediction and update equations derivation

First we need the two following definitions:

- The **prior innovation** $\widetilde{\boldsymbol{\theta}}_{k+1|k}$ is the difference between the actual state $\boldsymbol{\theta}_{k+1}$ and the prior estimate $\widehat{\boldsymbol{\theta}}_{k+1|k}$:
$$\widetilde{\boldsymbol{\theta}}_{k+1|k} = \boldsymbol{\theta}_{k+1} - \widehat{\boldsymbol{\theta}}_{k+1|k}$$
It is a random vector with covariance matrix $\mathsf{P}_{k+1|k} = E\left(\widetilde{\boldsymbol{\theta}}_{k+1|k}\widetilde{\boldsymbol{\theta}}_{k+1|k}^{\mathsf{T}}\right)$.
- The **posterior innovation** $\widetilde{\boldsymbol{\theta}}_{k+1|k+1}$ is the difference between the actual state $\boldsymbol{\theta}_{k+1}$ and the posterior estimate $\widehat{\boldsymbol{\theta}}_{k+1|k+1}$:
$$\widetilde{\boldsymbol{\theta}}_{k+1|k+1} = \boldsymbol{\theta}_{k+1} - \widehat{\boldsymbol{\theta}}_{k+1|k+1}$$
It is a random vector with covariance matrix $\mathsf{P}_{k+1|k+1} = E\left(\widetilde{\boldsymbol{\theta}}_{k+1|k+1}\widetilde{\boldsymbol{\theta}}_{k+1|k+1}^{\mathsf{T}}\right)$.

Now we define our estimates $\widehat{\boldsymbol{\theta}}_{k+1|k}$ and $\widehat{\boldsymbol{\theta}}_{k+1|k+1}$, and derive the corresponding innovation covariance matrices $\mathsf{P}_{k+1|k}$ and $\mathsf{P}_{k+1|k+1}$. Since we assumed that $E(\mathbf{v}_k) = 0$, based on Equation (2), we define the prior state estimate $\widehat{\boldsymbol{\theta}}_{k+1|k}$ as

$$\widehat{\boldsymbol{\theta}}_{k+1|k} = \mathsf{F}_k\widehat{\boldsymbol{\theta}}_{k|k} \tag{3}$$

The corresponding prior innovation is then

$$\widetilde{\boldsymbol{\theta}}_{k+1|k} = \boldsymbol{\theta}_{k+1} - \widehat{\boldsymbol{\theta}}_{k+1|k} = \mathsf{F}_k\boldsymbol{\theta}_k + \mathbf{v}_k - \mathsf{F}_k\widehat{\boldsymbol{\theta}}_{k|k} = \mathsf{F}_k\left(\boldsymbol{\theta}_k - \widehat{\boldsymbol{\theta}}_{k|k}\right) + \mathbf{v}_k = \mathsf{F}_k\widetilde{\boldsymbol{\theta}}_{k|k} + \mathbf{v}_k$$

Since posterior innovation $\widetilde{\boldsymbol{\theta}}_{k|k}$ and noise $\mathbf{v}_k$ are uncorrelated, we have

$$\mathsf{P}_{k+1|k} = \mathrm{cov}\left(\widetilde{\boldsymbol{\theta}}_{k+1|k}\right) = \mathrm{cov}\left(\mathsf{F}_k\widetilde{\boldsymbol{\theta}}_{k|k}\right) + \mathrm{cov}\left(\mathbf{v}_k\right) = \mathsf{F}_k\mathsf{P}_{k|k}\mathsf{F}_k^{\mathsf{T}} + \mathsf{R}_k^{v} \tag{4}$$

Since we assumed that $E(\mathbf{u}_k) = 0$, from prior state estimate $\widehat{\boldsymbol{\theta}}_{k+1|k}$ we can define a prior estimate $\widehat{\mathbf{y}}_{k+1|k} = H_{k+1}\widehat{\boldsymbol{\theta}}_{k+1|k}$ of the upcoming observation $\mathbf{y}_{k+1}$. To define the posterior state estimate $\widehat{\boldsymbol{\theta}}_{k+1|k+1}$, we inspire from the RLS weight update equation to write:

$$\widehat{\boldsymbol{\theta}}_{k+1|k+1} = \widehat{\boldsymbol{\theta}}_{k+1|k} + \mathsf{K}_{k+1}\left(\mathbf{y}_{k+1} - \widehat{\mathbf{y}}_{k+1|k}\right) = \widehat{\boldsymbol{\theta}}_{k+1|k} + \mathsf{K}_{k+1}\left(\mathbf{y}_{k+1} - \mathsf{H}_{k+1}\widehat{\boldsymbol{\theta}}_{k+1|k}\right) \tag{5}$$

where K_{k+1} is the **Kalman gain** that we are going to determine. The corresponding posterior innovation is then

$$\begin{aligned}
\widetilde{\boldsymbol{\theta}}_{k+1|k+1} &= \boldsymbol{\theta}_{k+1} - \widehat{\boldsymbol{\theta}}_{k+1|k+1} = \boldsymbol{\theta}_{k+1} - \widehat{\boldsymbol{\theta}}_{k+1|k} - \mathsf{K}_{k+1}\left(\mathsf{H}_k\boldsymbol{\theta}_k + \mathbf{u}_k - \mathsf{H}_{k+1}\widehat{\boldsymbol{\theta}}_{k+1|k}\right) \\
&= \boldsymbol{\theta}_{k+1} - \widehat{\boldsymbol{\theta}}_{k+1|k} - \mathsf{K}_{k+1}\left(\mathsf{H}_k\left(\boldsymbol{\theta}_k - \widehat{\boldsymbol{\theta}}_{k+1|k}\right) + \mathbf{u}_k\right) \\
&= \left(\mathsf{I} - \mathsf{K}_{k+1}\mathsf{H}_k\right)\widetilde{\boldsymbol{\theta}}_{k+1|k} - \mathsf{K}_{k+1}\mathbf{u}_k
\end{aligned}$$

Since the prior innovation and the measurement noise $\mathbf{u}_k$ are uncorrelated, we can write :

$$\begin{aligned}\mathsf{P}_{k+1|k+1} &= \operatorname{cov}\left(\widetilde{\boldsymbol{\theta}}_{k+1|k+1}\right) = \operatorname{cov}\left(\left(\mathsf{I}-\mathsf{K}_{k+1}\mathsf{H}_k\right)\widetilde{\boldsymbol{\theta}}_{k+1|k}\right) + \operatorname{cov}\left(\mathsf{K}_{k+1}\mathbf{u}_k\right)\\ &= \left(\mathsf{I}-\mathsf{K}_{k+1}\mathsf{H}_{k+1}\right)\mathsf{P}_{k+1|k}\left(\mathsf{I}-\mathsf{K}_{k+1}\mathsf{H}_{k+1}\right)^{\mathsf{T}} + \mathsf{K}_{k+1}\mathsf{R}_k^u\mathsf{K}_{k+1}^{\mathsf{T}}\end{aligned}$$

Hence posterior covariance matrix $\mathsf{P}_{k+1|k+1}$ depends on Kalman gain K_{k+1}, and we need to define the latter one so that it minimizes the following error criterion: the expected ℓ_2 norm of innovation $\widetilde{\boldsymbol{\theta}}_{k+1|k+1}$. This expected norm is related to matrix $\mathsf{P}_{k+1|k+1}$ by its trace:

$$\begin{aligned}E\left(\left\|\widetilde{\boldsymbol{\theta}}_{k+1|k+1}\right\|^2\right) &= E\left(\widetilde{\boldsymbol{\theta}}_{k+1|k+1}^{\mathsf{T}}\widetilde{\boldsymbol{\theta}}_{k+1|k+1}\right) = E\left(\operatorname{tr}\left(\widetilde{\boldsymbol{\theta}}_{k+1|k+1}^{\mathsf{T}}\widetilde{\boldsymbol{\theta}}_{k+1|k+1}\right)\right) = E\left(\operatorname{tr}\left(\widetilde{\boldsymbol{\theta}}_{k+1|k+1}\widetilde{\boldsymbol{\theta}}_{k+1|k+1}^{\mathsf{T}}\right)\right)\\ &= \operatorname{tr}\left(E\left(\widetilde{\boldsymbol{\theta}}_{k+1|k+1}\widetilde{\boldsymbol{\theta}}_{k+1|k+1}^{\mathsf{T}}\right)\right) = \operatorname{tr}\left(\mathsf{P}_{k+1|k+1}\right) = \xi_{k+1}(\mathsf{K}_{k+1})\end{aligned}$$

Before computing the derivative of criterion ξ_{k+1} with respect to K_{k+1}, let us develop the expression of $\mathsf{P}_{k+1|k+1}$:

$$\mathsf{P}_{k+1|k+1} = \mathsf{P}_{k+1|k} - \mathsf{P}_{k+1|k}\mathsf{H}_{k+1}^{\mathsf{T}}\mathsf{K}_{k+1}^{\mathsf{T}} - \mathsf{K}_{k+1}\mathsf{H}_{k+1}\mathsf{P}_{k+1|k} + \mathsf{K}_{k+1}\mathsf{H}_{k+1}\mathsf{P}_{k+1|k}\mathsf{H}_{k+1}^{\mathsf{T}}\mathsf{K}_{k+1}^{\mathsf{T}} + \mathsf{K}_{k+1}\mathsf{R}_k^u\mathsf{K}_{k+1}^{\mathsf{T}}$$

Therefore

$$\begin{aligned}\frac{\partial\xi_{k+1}}{\partial\mathsf{K}_{k+1}} &= \frac{\partial}{\partial\mathsf{K}_{k+1}}\operatorname{tr}\left(-\mathsf{P}_{k+1|k}\mathsf{H}_{k+1}^{\mathsf{T}}\mathsf{K}_{k+1}^{\mathsf{T}} - \mathsf{K}_{k+1}\mathsf{H}_{k+1}\mathsf{P}_{k+1|k} + \mathsf{K}_{k+1}\left(\mathsf{H}_{k+1}\mathsf{P}_{k+1|k}\mathsf{H}_{k+1}^{\mathsf{T}} + \mathsf{R}_k^u\right)\mathsf{K}_{k+1}^{\mathsf{T}}\right)\\ &= -2\left(\mathsf{H}_{k+1}\mathsf{P}_{k+1|k}\right)^{\mathsf{T}} + 2\mathsf{K}_{k+1}\left(\mathsf{H}_{k+1}\mathsf{P}_{k+1|k}\mathsf{H}_{k+1}^{\mathsf{T}} + \mathsf{R}_k^u\right)\end{aligned}$$

Since the optimal gain corresponds to $\dfrac{\partial\xi_{k+1}}{\partial\mathsf{K}_{k+1}} = 0$, we get:

$$\mathsf{K}_{k+1} = \mathsf{P}_{k+1|k}\mathsf{H}_{k+1}^{\mathsf{T}}\left(\mathsf{H}_{k+1}\mathsf{P}_{k+1|k}\mathsf{H}_{k+1}^{\mathsf{T}} + \mathsf{R}_k^u\right)^{-1} \tag{6}$$

If we replace the expression of the optimal gain in $\mathsf{P}_{k+1|k+1}$, we get:

$$\mathsf{P}_{k+1|k+1} = \mathsf{P}_{k+1|k} - \mathsf{K}_{k+1}\left(\mathsf{H}_{k+1}\mathsf{P}_{k+1|k}\mathsf{H}_{k+1}^{\mathsf{T}} + \mathsf{R}_k^u\right)\mathsf{K}_{k+1}^{\mathsf{T}} = \mathsf{P}_{k+1|k} - \mathsf{K}_{k+1}\mathsf{H}_{k+1}\mathsf{P}_{k+1|k}$$

which finally yields to:

$$\mathsf{P}_{k+1|k+1} = \left(\mathsf{I}-\mathsf{K}_{k+1}\mathsf{H}_{k+1}\right)\mathsf{P}_{k+1|k} \tag{7}$$

Finally we simply initialize the algorithm with $\mathsf{P}_{0|0} = 0$, as we have done no estimation yet. The initialization of $\widehat{\boldsymbol{\theta}}_{k|k}$ depends on how much information we have about the system at the beginning of the estimation process. We wrap up equations (3), (4), (5), (6) and (7) into the following algorithm.

3 Application: trajectory tracking

Imagine a mobile object which can only moves horizontally in one direction. We discretize time with step T. We denote $x_k = x(kT)$ the position of the mobile, $\dot{x}_k = \dot{x}(kT)$ its speed and $\ddot{x}_k = \ddot{x}(kT)$ its acceleration. This object is initially at position $x_0 = 0$ with speed $\dot{x}_0 = v_0$. All vertical forces compensate and only a random acceleration $a_k \sim \mathcal{N}(0, \sigma_a^2)$ is applied in the motion direction. The state vector that we want to estimate is $\boldsymbol{\theta}_k = \begin{pmatrix} x_k \\ \dot{x}_k \end{pmatrix}$. By the second law of dynamics, $\ddot{x}_k = a_k$.

Algorithm 1 Kalman filter

procedure KALMAN-FILTER($\mathbf{y}$, F, R^v, H, R^u)
  Input $\mathbf{y}_k$, F_k, R^v_k, H_k and R^u_k for $k \in [\![1, n]\!]$
  $\mathrm{P}_{0|0} = 0$
  for $k \in [\![0, n]\!]$ **do**
    Update covariance matrices and Kalman gain
    $\mathrm{P}_{k+1|k} \leftarrow \mathrm{F}_k \mathrm{P}_{k|k} \mathrm{F}_k^{\mathrm{T}} + \mathrm{R}^v_k$
    $\mathrm{K}_{k+1} \leftarrow \mathrm{P}_{k+1|k} \mathrm{H}_{k+1}^{\mathrm{T}} \left(\mathrm{H}_{k+1} \mathrm{P}_{k+1|k} \mathrm{H}_{k+1}^{\mathrm{T}} + \mathrm{R}^u_k \right)^{-1}$
    $\mathrm{P}_{k+1|k+1} \leftarrow (\mathrm{I} - \mathrm{K}_{k+1} \mathrm{H}_{k+1}) \mathrm{P}_{k+1|k}$
    Update state variables
    $\widehat{\boldsymbol{\theta}}_{k+1|k} \leftarrow \mathrm{F}_k \widehat{\boldsymbol{\theta}}_{k|k}$
    $\widehat{\boldsymbol{\theta}}_{k+1|k+1} \leftarrow \widehat{\boldsymbol{\theta}}_{k+1|k} + \mathrm{K}_{k+1} \left(\mathbf{y}_{k+1} - \mathrm{H}_{k+1} \widehat{\boldsymbol{\theta}}_{k+1|k} \right)$
  end for
  Return $\hat{\mathbf{x}}$
end procedure

Taylor series of $x(t)$ and $\dot{x}(t)$ give:

$$
\begin{aligned}
x_{k+1} &= x((k+1)T) = x(kT) + T\dot{x}(kT) + \frac{T^2}{2}\ddot{x}(kT) + o(T^2) = x_k + T\dot{x}_k + \frac{T^2}{2} a_k + o(T^2) \\
\dot{x}_{k+1} &= \dot{x}((k+1)T) = \dot{x}(kT) + T\ddot{x}(kT) + o(T) = \dot{x}_k + T a_k + o(T)
\end{aligned}
$$

We can approximate these expressions matricially:

$$
\boldsymbol{\theta}_{k+1} = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} \boldsymbol{\theta}_k + \begin{pmatrix} \frac{T^2}{2} \\ T \end{pmatrix} a_k
$$

which gives:

$$
\mathrm{F}_k = \mathrm{F} = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} \qquad \mathbf{v}_k = \begin{pmatrix} \frac{T^2}{2} \\ T \end{pmatrix} a_k \qquad \mathrm{R}^v_k = \mathrm{R}^v = \sigma_a^2 \begin{pmatrix} \frac{T^4}{4} & \frac{T^3}{2} \\ \frac{T^3}{2} & T^2 \end{pmatrix}
$$

The observation variable y_k is simply a noisy observation of position x_k, i.e.

$$
y_k = \begin{pmatrix} 1 & 0 \end{pmatrix} \boldsymbol{\theta}_k + u_k
$$

where $u_k \sim \mathcal{N}(0, \sigma^2)$ is observation noise. Thus we have

$$
\mathrm{H}_k = \mathrm{H} = \begin{pmatrix} 1 & 0 \end{pmatrix} \qquad \mathrm{R}^u_k = \mathrm{R}^u = \sigma^2
$$

For the initialization, the initial position $x_0 = 0$ is known. The initial speed v_0 is unknown but we assume it to be a realization of a zero-mean random variable, thus we will set our initial speed estimation to 0. The initial posterior state estimate is then $\hat{\mathbf{x}}_{0|0} = (0, 0)^{\mathrm{T}}$.

Figure 1 displays the evolution over time of the true position, the noisy observation of the position and the position estimated by the Kalman filter. Figure 2 shows the evolution of the true speed and the speed estimated by the Kalman filter. Finally, Figure 3 presents the error deviation of the observed and estimated positions from the true position.

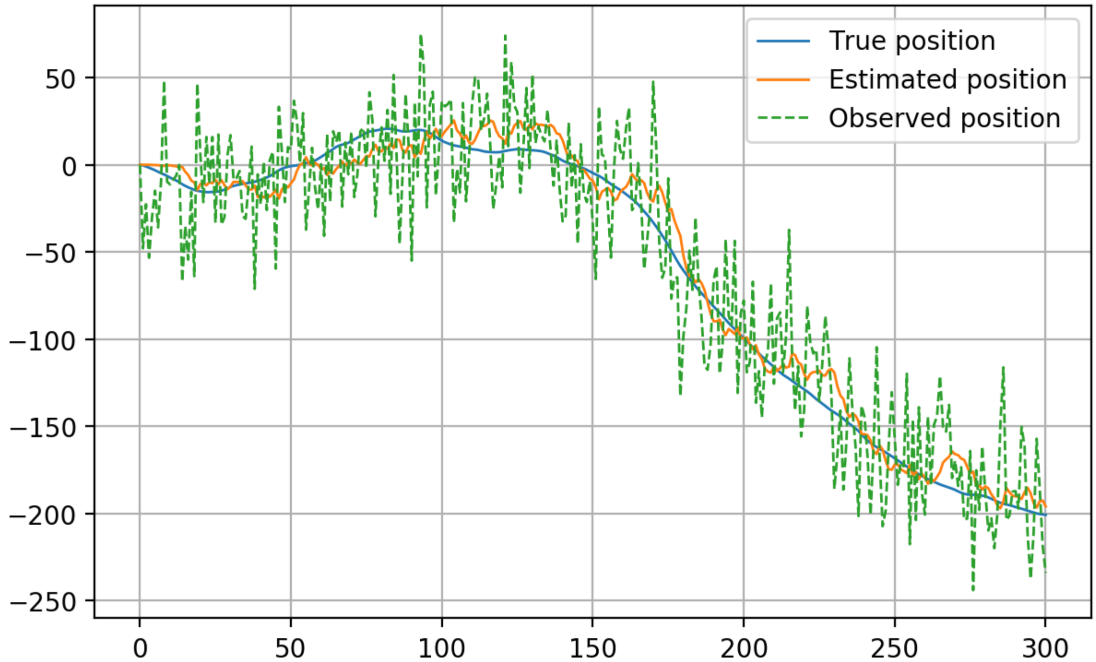

Figure 1: True, observed and estimated positions

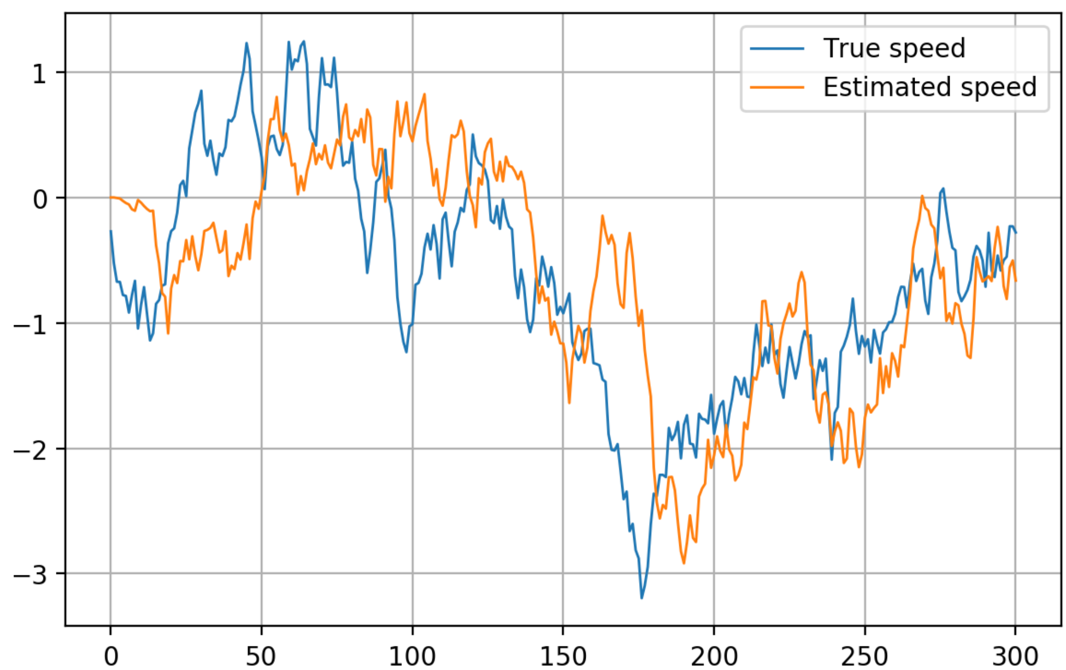

Figure 2: True and estimated speeds

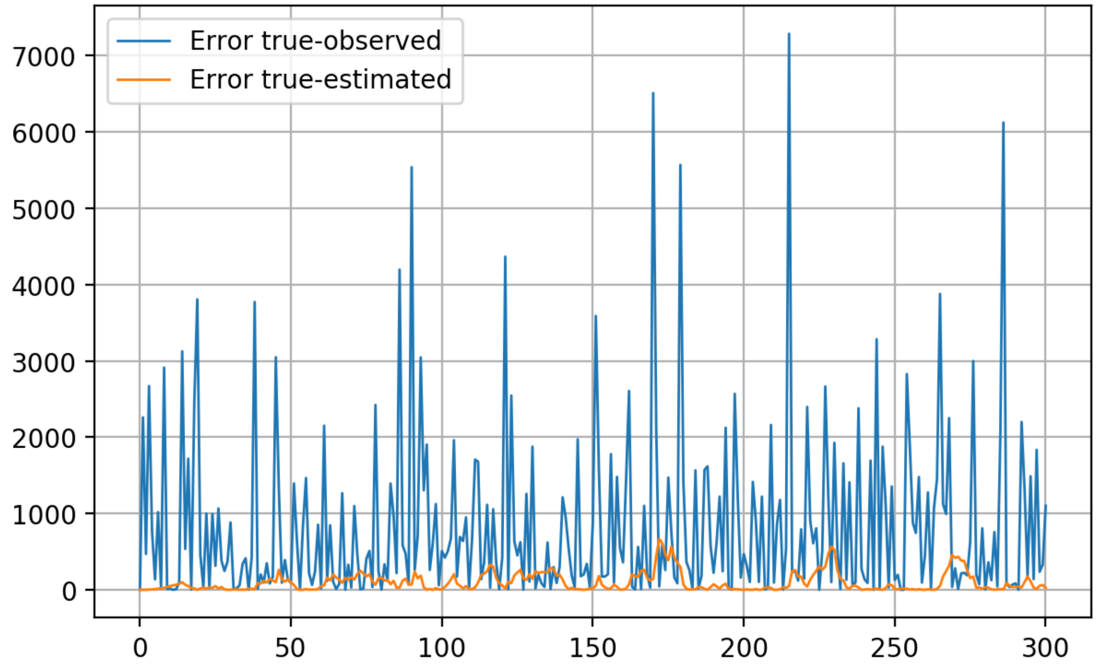

Figure 3: Error deviations of observed and estimated positions

4 Linear regression as a special case of Kalman filter

Kalman filters bring the addition of state evolution to linear regression. Therefore, the latter one can be seen as a special case of Kalman filter with a constant hidden state, the weights $\mathbf{w} = \boldsymbol{\theta}$ that we are trying to estimate. The corresponding prediction and update equations are then:

$$\mathbf{w}_{k+1} = \mathrm{I}_m \mathbf{w}_k = \mathbf{w}_k \qquad \text{and} \qquad y_k = \mathbf{x}_k^\mathrm{T} \mathbf{w}_k + e_k$$

where $\mathrm{F}_k = \mathrm{I}_m$, $\mathbf{v}_k = 0$, $\mathrm{H}_k = \mathbf{x}_k^\mathrm{T}$ and $u_k = e_k$. Then, innovation covariance matrices and Kalman gain can be written:

$$\mathrm{P}_{k+1|k} = \mathrm{F}_k \mathrm{P}_{k|k} \mathrm{F}_k^T + \mathrm{R}_k^v = \mathrm{P}_{k|k}$$

$$\mathrm{K}_{k+1} = \mathrm{P}_{k+1|k} \mathrm{H}_{k+1}^T \left(\mathrm{H}_{k+1} \mathrm{P}_{k+1|k} \mathrm{H}_{k+1}^T + \mathrm{R}_k^u \right)^{-1} = \frac{\mathrm{P}_{k|k} \mathbf{x}_{k+1}}{\mathbf{x}_{k+1}^T \mathrm{P}_{k|k} \mathbf{x}_{k+1} + \sigma_e^2}$$

$$\mathrm{P}_{k+1|k+1} = \left(\mathrm{I}_m - \mathrm{K}_{k+1} \mathbf{x}_{k+1}^T \right) \mathrm{P}_{k|k}$$

Therefore, we retrieve RLS Equations by substituting $\mathrm{P}_{k|k} \leftrightarrow \mathrm{P}_k$ and $\mathrm{K}_{k+1} \leftrightarrow \mathbf{g}_{k+1}$.