# Lecture Notes - 2018/10/22
# Recursive Least Squares Estimation

Team 4
Nick Heppert, 김동민, 김병수, 박민준

October 24, 2018

In this lecture we derived the recursive least squares estimation and discussed properties of the "$P_t$"-matrix. Eventually we briefly covered a Bayesian view of the problem.

## 1 Deriving of the Recursive Least Squares Estimation Update Rule

As in the previous lecture shown the estimation of the parameters is given by

$$\hat{\theta} = P_t B_t. \tag{1}$$

Our goal is now to continuously update the estimation based on new measurements. Therefore we investigate how to update the two factors $P_t$ and $B_t$ of the product.

For $B_t$ it is easy, we just remove the last term of the sum and write it separately

$$B_t = \sum_{i=1}^{t} y(i)\bar{u}(i) = \underbrace{\sum_{i=1}^{t-1} y(i)\bar{u}(i)}_{B_{t-1}} + y(t)\bar{u}(t) \tag{2}$$

$$= B_{t-1} + y(t)\bar{u}(t) \tag{3}$$

For the second factor in the product $P_t$ appending a new parameter is not as easy as for $B_t$. We can write $P_t^{-1}$ in a similar way as before

$$P_t^{-1} = \sum_{i=1}^{t} \bar{u}(i)\bar{u}(i)^T = P_{t-1}^{-1} + \bar{u}(t)\bar{u}^T(t) \tag{4}$$

But calculating the inverse of $P_t^{-1}$ to get $P_t$ in each step is rather time-consuming and not efficient. We therefore derive a closed-form solution for updating $P_t$ without the inverse.

Let's assume the formula from before and apply multiple operations:

$$
\begin{aligned}
P_t^{-1} &= & P_{t-1}^{-1} + \bar{u}(t)\bar{u}^T(t) & \quad |P_t \cdot & (5)\\
P_t P_t^{-1} &= & P_t P_{t-1}^{-1} + P_t\bar{u}(t)\bar{u}^T(t) & \quad | \cdot P_{t-1} & (6)\\
\underline{P_t P_t^{-1}} P_{t-1} &= & P_t \underline{P_{t-1}^{-1} P_{t-1}} + P_t\bar{u}(t)\bar{u}^T(t)P_{t-1} & \quad |\text{reducing} & (7)\\
P_{t-1} &= & P_t + P_t\bar{u}(t)\bar{u}^T(t)P_{t-1} & \quad | \cdot \bar{u}(t) & (8)\\
P_{t-1}\bar{u}(t) &= & P_t\bar{u}(t) + P_t\bar{u}(t)\bar{u}^T(t)P_{t-1}\bar{u}(t) & \quad |\text{factorizing} & (9)\\
P_{t-1}\bar{u}(t) &= & P_t\bar{u}(t)\left[1 + \bar{u}^T(t)P_{t-1}\bar{u}(t)\right] & \quad |/\left[1 + \bar{u}^T(t)P_{t-1}\bar{u}(t)\right] & (10)\\
P_t\bar{u}(t) &= & \tfrac{P_{t-1}\bar{u}(t)}{1+\bar{u}^T(t)P_{t-1}\bar{u}(t)} & \quad | \cdot \bar{u}^T(t)P_{t-1} & (11)\\
P_t\bar{u}(t)\bar{u}^T(t)P_{t-1} &= & \tfrac{P_{t-1}\bar{u}(t)\bar{u}^T(t)P_{t-1}}{1+\bar{u}^T(t)P_{t-1}\bar{u}(t)} & & (12)
\end{aligned}
$$

If we now reorder Eq. 8 to

$$P_t\bar{u}(t)\bar{u}^T(t)P_{t-1} = P_{t-1} - P_t \tag{13}$$

We can see they both share the same left hand side so we can continue deriving the update equation by setting Eq. 13 and Eq. 12 equal:

$$\frac{P_{t-1}\bar{u}(t)\bar{u}^T(t)P_{t-1}}{1+\bar{u}^T(t)P_{t-1}\bar{u}(t)} = P_{t-1} - P_t \tag{14}$$

A simple reorder yields the final update rule

$$P_t = P_{t-1} - \frac{P_{t-1}\bar{u}(t)\bar{u}^T(t)P_{t-1}}{1+\bar{u}^T(t)P_{t-1}\bar{u}(t)}. \tag{15}$$

Next we continue by choosing an parameter update rule similar to gradient descent:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + K_t\left[y(t) - \bar{u}^T(t)\hat{\theta}(t-1)\right] \tag{16}$$

Where $K_t$ is a type of gain for correcting the prediction error. As previously the parameters can be estimates respectively by

$$\hat{\theta}(t) = P_t B_t, \qquad \hat{\theta}(t-1) = P_{t-1}B_{t-1} \tag{17}$$

If we now plug these in Eq. 16

$$\hat{\theta}(t) - \hat{\theta}(t-1) = P_t B_t - P_{t-1}B_{t-1} \tag{18}$$

derive a formula for $K_t$.

$$
\begin{aligned}
\hat{\theta}(t) - \hat{\theta}(t-1) = & \quad \left[P_{t-1} - \frac{P_{t-1}\bar{u}(t)\bar{u}^T(t)P_{t-1}}{1+\bar{u}^T(t)P_{t-1}\bar{u}(t)}\right]\left[B_{t-1} + y(t)\bar{u}(t)\right] - P_{t-1}B_{t-1}\\
= \quad \ldots = & \quad \underbrace{\frac{P_{t-1\bar{u}(t)}}{1+\bar{u}^T(t)P_{t-1}\bar{u}(t)}}_{K_t}\left[y(t) - \bar{u}^T(t)\hat{\theta}(t-1)\right]
\end{aligned}
$$

*In class the professor did not derive this formula. In the appendix section B we show how to derive it even with a forgetting regularizer.*

Finally, The full update rule has this form:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{P_{t-1}\bar{u}(t)}{1+\bar{u}^T(t)P_{t-1}\bar{u}(t)}\left[y(t)) - \bar{u}^T(t)\hat{\theta}(t-1)\right] \tag{19}$$

The initial conditions are given by:

| | |
|---|---|
| $\hat{\theta}(0)$ | arbitrary |
| $P_0$ | identity matrix |
| $B_0$ | identity matrix |

## 2 Discussion about the $P$-Matrix

Abstractly seen the $P$-matrix has a similar structure to a co-variance matrix.

$$P^{-1} = \sum_{i=1}^{t} \bar{u}(i)\bar{u}^T(i) \qquad \frac{1}{N}\sum_{i=1}^{N}(\bar{x}(i) - \bar{m}_x)(\bar{x}(i) - \bar{m}_x)^T \tag{20}$$

We reshape $P_t^{-1}$ to

$$P_t^{-1} = \sum_{i=1}^{t} \bar{u}(i)\bar{u}^T(i) = \sqcup_t \sqcup_t^T, \text{where } \sqcup_t = \begin{bmatrix} | & & | \\ \bar{u}(1) & \dots & \bar{u}(t) \\ | & & | \end{bmatrix} \in \mathbb{R}^{m \times t} \tag{21}$$

where $m$ is the the size of $u \in \mathbb{R}^m$ and $t$ the number of time steps.

If we now look at the the eigenvalues of $P_t^{-1}$ we can say a bigger eigenvalue responds to a more well-traveled direction, i.e. a direction in which we acquired more diverse samples. Vice versa a smaller eigenvalue corresponds to a less explored direction and therefore might be preferred to explore more.

For this discussed case we assumed the eigenvalues of $P_t^{-1}$, so actually the inverse of $P_t$. If we construct the eigenvalues matrix for $P_t^{-1}$ in this manner

$$D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_m \end{bmatrix} \tag{22}$$

the corresponding matrix for $P_t$ has the form

$$D^{-1} = \begin{bmatrix} 1/\lambda_1 & & 0 \\ & \ddots & \\ 0 & & 1/\lambda_m \end{bmatrix}. \tag{23}$$

Therefore the desired directions are the opposite than before.

## 3 Bayesian View of Recursive Least Squares Estimation

Assuming we will minimizes this

$$J_t(\bar{\theta}) = \underbrace{\tfrac{1}{2}\textstyle\sum_{i=1}^{t}\left[y(i) - \bar{u}^T(i)\bar{\theta}\right]^2}_{\text{Only data, square estimation}} + \underbrace{\frac{1}{2}(\bar{\theta} - \bar{\theta}_0)P_0^{-1}(\bar{\theta} - \bar{\theta}_0)}_{\text{Prior/Regulization; Weighted squared distance from } \bar{\theta}_0}$$

extended cost function we will experience a different behavior due to the fact we added a second term. As before the minimum is found by setting the derivative

$$\frac{dJ}{d\theta} = 0 \tag{24}$$

to zero.

As $t \to \infty$ the square estimation gets bigger as the sum is not normalized. For small $t$ the parameters $\bar{\theta}$ are pulled towards the prior $\bar{\theta}_0$ especially if the eigenvalues of $P_0^{-1}$ are large. On the other side if the eigenvalues of $P_0^{-1}$ are small, $\bar{\theta}$ tends to change more quickly. So one can see the initial $P_0$ as a level of confidence in the initial/prior parameter $\bar{\theta}_0$.

# 4 Appendix

In the following appendix we will provide you with some additional information.

# A Extend to "forget" older parts

One possible extension of this current algorithm is to extend it by adding a forgetting regularizer $\lambda$.

For $B_t$ it is easy to integrate in Eq. 3:

$$B_t = \lambda B_{t-1} + y(t)\bar{u}(t) \tag{25}$$

The integration in Eq. 4 is in a similar style:

$$P_t^{-1} = \lambda P_{t-1}^{-1} + \bar{u}(t)\bar{u}^T(t) \tag{26}$$

We could now redo all the derivation from before but for the sake of readability we only present the final update rule for $P_t$ by:

$$P_t = \frac{1}{\lambda}P_{t-1} - \frac{P_{t-1}\bar{u}(t)\bar{u}^T(t)\frac{1}{\lambda}P_{t-1}}{\lambda + \bar{u}^T(t)P_{t-1}\bar{u}(t)}. \tag{27}$$

# B Derivation of $K_t$ with Forgetting Regularizer

Assuming the same update rule Eq. 16 as before we will derive the formula for $K_t$ with the updated version for $P_t$ as well as $B_t$ from section A.

$$\hat{\theta}(t) - \hat{\theta}(t-1) = \left[\frac{1}{\lambda}P_{t-1} - \frac{P_{t-1}\bar{u}(t)\bar{u}^T(t)\frac{1}{\lambda}P_{t-1}}{\lambda + \bar{u}^T(t)P_{t-1}\bar{u}(t)}\right][\lambda B_{t-1} + y(t)\bar{u}(t)] - P_{t-1}B_{t-1} \tag{28}$$

First, we expand the equation in each separate term and simplify.

$$\begin{aligned}
\hat{\theta}(t) - \hat{\theta}(t-1) &= \frac{1}{\lambda}P_{t-1}\lambda B_{t-1} - \frac{P_{t-1}\bar{u}(t)\bar{u}^T(t)\frac{1}{\lambda}P_{t-1}\lambda B_{t-1}}{\lambda + \bar{u}^T(t)P_{t-1}\bar{u}(t)} && (29)\\
&\quad + \frac{1}{\lambda}P_{t-1}y(t)\bar{u}(t) - \frac{P_{t-1}\bar{u}(t)\bar{u}^T(t)\frac{1}{\lambda}P_{t-1}y(t)\bar{u}(t)}{\lambda + \bar{u}^T(t)P_{t-1}\bar{u}(t)} - P_{t-1}B_{t-1} && (30)\\
\dots &= \underbrace{-\frac{P_{t-1}\bar{u}(t)\bar{u}^T(t)P_{t-1}B_{t-1}}{\lambda + \bar{u}^T(t)P_{t-1}\bar{u}(t)}}_{=A} && (31)\\
&\quad + \frac{1}{\lambda}P_{t-1}y(t)\bar{u}(t) - \underbrace{\frac{P_{t-1}\bar{u}(t)\bar{u}^T(t)\frac{1}{\lambda}P_{t-1}y(t)\bar{u}(t)}{\lambda + \bar{u}^T(t)P_{t-1}\bar{u}(t)}}_{=B} && (32)\\
& && (33)
\end{aligned}$$

Second, we extend the term in the middle to the same denominator and expand it as well

$$\hat{\theta}(t) - \hat{\theta}(t-1) = -A + \frac{1}{\lambda} P_{t-1} y(t) \bar{u}(t) \frac{\lambda + \bar{u}^T(t) P_{t-1} \bar{u}(t)}{\lambda + \bar{u}^T(t) P_{t-1} \bar{u}(t)} - B \tag{34}$$

$$\ldots = -A + \frac{\frac{1}{\lambda} P_{t-1} y(t) \bar{u}(t) \left[\lambda + \bar{u}^T(t) P_{t-1} \bar{u}(t)\right]}{\lambda + \bar{u}^T(t) P_{t-1} \bar{u}(t)} - B \tag{35}$$

$$\ldots = -A + \frac{\frac{1}{\lambda} P_{t-1} y(t) \bar{u}(t) \lambda + \frac{1}{\lambda} P_{t-1} y(t) \bar{u}(t) \bar{u}^T(t) P_{t-1} \bar{u}(t)}{\lambda + \bar{u}^T(t) P_{t-1} \bar{u}(t)} - B \tag{36}$$

$$\ldots = -A + \frac{P_{t-1} y(t) \bar{u}(t) + \frac{1}{\lambda} P_{t-1} y(t) \bar{u}(t) \bar{u}^T(t) P_{t-1} \bar{u}(t)}{\lambda + \bar{u}^T(t) P_{t-1} \bar{u}(t)} - B \tag{37}$$

Third, we remove all terms that cancel each other

$$\hat{\theta}(t) - \hat{\theta}(t-1) = -\frac{P_{t-1} \bar{u}(t) \bar{u}^T(t) P_{t-1} B_{t-1}}{\lambda + \bar{u}^T(t) P_{t-1} \bar{u}(t)} \tag{38}$$

$$+\frac{P_{t-1} y(t) \bar{u}(t) + \frac{1}{\lambda} P_{t-1} y(t) \bar{u}(t) \bar{u}^T(t) P_{t-1} \bar{u}(t)}{\lambda + \bar{u}^T(t) P_{t-1} \bar{u}(t)} \tag{39}$$

$$-\frac{P_{t-1} \bar{u}(t) \bar{u}^T(t) \frac{1}{\lambda} P_{t-1} y(t) \bar{u}(t)}{\lambda + \bar{u}^T(t) P_{t-1} \bar{u}(t)} \tag{40}$$

$$\ldots = -\frac{P_{t-1} \bar{u}(t) \bar{u}^T(t) P_{t-1} B_{t-1} + P_{t-1} y(t) \bar{u}(t)}{\lambda + \bar{u}^T(t) P_{t-1} \bar{u}(t)} \tag{41}$$

Fourth, we factorize the nominator again and reorder it

$$\hat{\theta}(t) - \hat{\theta}(t-1) = \frac{P_{t-1} \bar{u}(t) \left[-\bar{u}^T(t) P_{t-1} B_{t-1} + y(t)\right]}{\lambda + \bar{u}^T(t) P_{t-1} \bar{u}(t)} \tag{42}$$

$$\ldots = \frac{P_{t-1} \bar{u}(t)}{\lambda + \bar{u}^T(t) P_{t-1} \bar{u}(t)} \left[y(t) - \bar{u}^T(t) P_{t-1} B_{t-1}\right] \tag{43}$$

Eventually we retrieve the final form by replacing $P_{t-1} B_{t-1}$ with the previous estimation $\hat{\theta}(t-1)$:

$$\hat{\theta}(t) - \hat{\theta}(t-1) = \underbrace{\frac{P_{t-1} \bar{u}(t)}{\lambda + \bar{u}^T(t) P_{t-1} \bar{u}(t)}}_{K_t} \left[y(t) - \bar{u}^T(t) \hat{\theta}(t-1)\right] \tag{44}$$

Therefore the update rule with the forgetting regularization as this form:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{P_{t-1} \bar{u}(t)}{\lambda + \bar{u}^T(t) P_{t-1} \bar{u}(t)} \left[y(t)) - \bar{u}^T(t) \hat{\theta}(t-1)\right] \tag{45}$$