**EECS 225A Statistical Signal Processing** Lecture 15 - 03/12/2020

# Lecture 15: Recursive Least Squares Algorithm

*Lecturer: Jiantao Jiao* *Scribe: Alejandro Saldarriaga Fuertes*

The Recursive Least Squares (RLS) algorithm is a well-known adaptive filtering algorithm that efficiently update or "downdate" the least square estimate. We present the algorithm and its connections to Kalman filter in this lecture.

## 1 Recursive Least Squares [1, Section 2.6]

Let's consider

$$Y_i = \begin{pmatrix} Y(0) \\ Y(1) \\ \dots \\ Y(i) \end{pmatrix} = H_i X + v_i \tag{1}$$

where

$$H_i = \begin{pmatrix} h_0^* \\ h_1^* \\ \dots \\ h_i^* \end{pmatrix}, h_j \in \mathbb{C}^{n\times 1}$$

$$v_i = \begin{pmatrix} v(0) \\ v(1) \\ \dots \\ v(i) \end{pmatrix} \in \mathbb{C}^{(i+1)\times 1}$$

$$X \in \mathbb{C}^{n\times 1}$$

We also assume that

$$\langle X, X\rangle = \Pi_0 \tag{2}$$

$$\langle v_i, v_i\rangle = I_{i+1} \tag{3}$$

$$\langle X, v_i\rangle = 0 \tag{4}$$

The theory of linear estimation in lecture 2 yields an expression for such optimal estimator at step $i$:

$$\hat{X}_i = (\Pi_0^{-1} + H_i^* H_i)^{-1} H_i^* Y_i \tag{5}$$

Suppose we already obtained the estimator $\hat{X}_{i-1}$, and a new observation $Y(i)$ arrived. How can we update the estimate to $\hat{X}_i$? Can we use $\hat{X}_{i-1}$ to reduce the computational burden? Legendre and Gauss found an algebraic trick to provide a fast recursive way to compute $\hat{X}_i$ from $\hat{X}_{i-1}$ with $O(n^2)$ complexity. One can make it even faster [2].

Before explaining the heart of RLS algorithm, we need the following lemma.

**Lemma 1.** *(Sherman–Morrison–Woodbury formula) Let $A \in \mathbb{C}^{n\times n}$, $C \in \mathbb{C}^{m\times m}$, $B \in \mathbb{C}^{n\times m}$ and $D \in \mathbb{C}^{m\times n}$. Then, if all inverse operations below are well defined,*

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1} \tag{6}$$

**Proof** The straight-forward approach yields

$$
\begin{aligned}
&(A+BCD)(A^{-1}-A^{-1}B(C^{-1}+DA^{-1}B)^{-1}DA^{-1}) \\
&= I_n - B(C^{-1}+DA^{-1}B)^{-1}DA^{-1} + BCDA^{-1} - BCDA^{-1}B(C^{-1}+DA^{-1}B)^{-1}DA^{-1} \\
&= I_n + BCDA^{-1} - BC(C^{-1}(C^{-1}+DA^{-1}B)^{-1} + DA^{-1}B(C^{-1}+DA^{-1}B)^{-1})DA^{-1} \\
&= I_n + BCDA^{-1} - BC((C^{-1}+DA^{-1}B)(C^{-1}+DA^{-1}B)^{-1})DA^{-1} \\
&= I_n
\end{aligned}
$$

□

If we define

$$P_i = (\Pi_0^{-1} + H_i^* H_i)^{-1},$$

then (5) becomes

$$\hat{X}_i = P_i H_i^* Y_i.$$

**Theorem 2.** *We have the following recurrence for all $i \geq 0$:*

$$P_{-1} = \Pi_0 \tag{7}$$

$$P_i = P_{i-1} - \frac{P_{i-1} h_i h_i^* P_{i-1}}{1 + h_i^* P_{i-1} h_i} \tag{8}$$

**Proof** We notice that

$$H_i^* H_i = H_{i-1}^* H_{i-1} + h_i h_i^*$$

Therefore,

$$P_i = (P_{i-1}^{-1} + h_i h_i^*)^{-1}$$

and the Sherman–Morrison–Woodbury applied to this case ($m = 1$) yields the desired result. □

The complexity of each step is in $O(n^2)$ so the total complexity after $i$ observations is $O(in^2)$, which is linear in the number of observations.

Note that using this formula, we also obtain a recursive pattern for the estimates

$$
\begin{aligned}
\hat{X}_i &= P_i H_i^* Y_i \\
&= (P_{i-1} - \frac{P_{i-1} h_i h_i^* P_{i-1}}{1 + h_i^* P_{i-1} h_i})(H_{i-1}^* Y_{i-1} + h_i Y(i)) \\
&= \hat{X}_{i-1} - \frac{P_{i-1} h_i h_i^* X_{i-1}}{1 + h_i^* P_{i-1} h_i} + P_{i-1} h_i Y(i) - \frac{P_{i-1} h_i h_i^* P_{i-1} h_i Y(i)}{1 + h_i^* P_{i-1} h_i}
\end{aligned}
$$

which yields

$$\hat{X}_i = \hat{X}_{i-1} + \frac{P_{i-1} h_i}{1 + h_i^* P_{i-1} h_i}(Y(i) - h_i^* \hat{X}_{i-1}) \tag{9}$$

with

$$\hat{X}_{-1} = 0. \tag{10}$$

This expression can be mirrored with a Kalman filter where the term $Y(i) - h_i^* \hat{X}_{i-1}$ represents the innovation of the sequence. More formally, we can define the following state-space model

$$\begin{aligned}
X_{j+1} &= X_j, X_0 = X \\
Y_j &= h_j^* X_j + v(j) \\
\langle X_0, X_0 \rangle &= \Pi_0 \\
\langle v(i), v(j) \rangle &= \delta_{ij} \\
\langle X_0, v(i) \rangle &= 0, i \geq 0
\end{aligned} \tag{11}$$

At time $i$, the observation model in (11) is equivalent to that in (1), and the RLS solution is in fact exactly the same as the Kalman filter solution for (11).

Hence, the RLS algorithm can be viewed as

1. a special case of Kalman filter under state-space model (11)
2. a recursive algorithm to solve the optimal linear estimator given model (1)
3. a recursive algorithm to solve the deterministic least squares problem

$$\min_X (X^* \Pi_0^{-1} X + \|Y_i - H_i X\|^2)$$

One way to connect the deterministic optimization with the stochastic optimization problem is through the Gaussian trick. We would assume that $X \sim \mathcal{N}(0, \Pi_0), v_i \sim \mathcal{N}(0, I_{i+1})$, and $X$ is independent of $v_i$. Then, we know that the maximum a posterior estimate of $X$ is given by the deterministic optimization problem, but we also know it is the optimal linear estimator due to the Gauss–Markov theorem.

## 2 Downdating Least Squares [1, Section 2.7]

Now we consider the opposite problem, which is trying to compute a linear estimator by forgetting some observations. In other words, we want to find the best estimator of $X$ given $Y_{1:i}$ which is the vector $Y_i$ without the first element $Y(0)$.

Here again, theory provides a closed expression for that optimal linear estimator $\hat{X}_{1:i}$:

$$\hat{X}_{1:i} = (\Pi_0^{-1} + H_{1:i}^* H_{1:i})^{-1} H_{1:i}^* Y_{1:i} \tag{12}$$

Similarly to the RLS algorithm, we can derive a backward recursive equation that translates the operation of forgetting the first observation.

**Theorem 3.** *Let*

$$P_i = (\Pi_0^{-1} + H_i^* H_i)^{-1}$$

*and*

$$P_{1:i} = (\Pi_0^{-1} + H_{1:i}^* H_{1:i})^{-1}$$

*such that*

$$\hat{X}_i = P_i H_i^* Y_i$$

*and*

$$\hat{X}_{1:i} = P_{1:i} H_{1:i}^* Y_{1:i}$$

*Then, we have the following equality*

$$P_{1:i} = P_i - \frac{P_i h_0 h_0^* P_i}{-1 + h_0^* P_i h_0} \tag{13}$$

**Proof** It suffices to see that

$$H_{1:i}^* H_{1:i} = H_i^* H_i - h_0 h_0^*$$

which, by the same reasoning as in the theorem 2, using the Sherman-Morrison-Woodbury formula, yields the desired result. $\square$

We also obtain a recursive formula similar to (9)

$$\hat{X}_{1:i} = \hat{X}_i + \frac{P_i h_0}{-1 + h_0^* P_i h_0}(Y(0) - h_0^* \hat{X}_i) \tag{14}$$

However, this time we can't find the analoguous state space model from which the Kalman filter derives the expression, as it would imply that the noise $v(i)$ has a negative definite covariance matrix ($\langle v(i), v(j)\rangle = -\delta_{ij}$). It was shown in [3] that the Kalman filter can be generalized to estimation in indefinite metric spaces.

# 3 Robust Least Squares

Now, we are interested in the optimization problem

$$\min_X \|Y - HX\|^2$$

where the matrix $H$ may be corrupted, ie we don't have access to $H$ but to a possibly biased version of it

$$\tilde{H} = H + \Delta H$$

In order to work around that inconvenience, the Total Least Squares [4] method adds a preliminary step, which is finding an optimal pair $[\hat{H}, \hat{Y}]$ that minimizes the following criterion

$$\min_{\hat{Y} \in \mathsf{span}(H)} \|[H, Y] - [\hat{H}, \hat{Y}]\|_F^2 \tag{15}$$

where $\|A\|_F^2 = \sum_{i,j} |A_{i,j}|^2$ is the Frobenius norm in the matrix space. Once those estimates are found, we then solve the initial optimization problem with the parameters $\hat{Y}$ and $\hat{H}$.

**Theorem 4.** *Assume $H$ is full column rank, and let $\sigma_n$ be its smallest singular value. Assume also that $[H, Y]$ is full rank and let $\overline{\sigma}_{n+1}$ be its smallest singular value. If $\overline{\sigma}_{n+1} < \sigma_n$, then the TLS problem has a unique solution given by*

$$\hat{X} = (H^* H - \overline{\sigma}_{n+1}^2 I_n)^{-1} H^* Y \tag{16}$$

We also have the formulation

$$\hat{X} = \arg\min_X (\|Y - HX\|^2 - \sigma_{n+1}^2 \|X\|^2) \tag{17}$$

For more on robustifying least squares, we refer the readers to [5, 6].

# References


[1] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear estimation.* Prentice Hall, 2000, no. BOOK.

[2] D. T. Slock and T. Kailath, "Numerically stable fast transversal filters for recursive least-squares adaptive filtering," in *Numerical Linear Algebra, Digital Signal Processing and Parallel Algorithms.* Springer, 1991, pp. 605–615.

[3] B. Hassibi, A. H. Sayed, and T. Kailath, *Indefinite-Quadratic Estimation and Control: A Unified Approach to $H^2$ and $H^\infty$ Theories.* Siam, 1999, vol. 16.

[4] P. P. de Groen, "An introduction to total least squares," *arXiv preprint math/9805076*, 1998.

[5] A. H. Sayed and V. H. Nascimento, "Design criteria for uncertain models with structured and unstructured uncertainties," in *Robustness in identification and control.* Springer, 1999, pp. 159–173.

[6] A. H. Sayed, "A framework for state-space estimation with uncertain models," *IEEE Transactions on Automatic Control*, vol. 46, no. 7, pp. 998–1013, 2001.