Exponential convergence of recursive least squares with forgetting factor for multiple-output systems

Sven Brüggemann a, Robert R. Bitmead a

Abstract
We extend results of the recursive-least-squares-with-forgetting-factor identifier for single-input-single-output systems to the multiple-output case by deriving the corresponding minimized objective function and by showing exponential convergence of the estimation error.

Key words: recursive least squares, adaptive control, MIMO systems

1 Introduction
This work revolves around the popular recursive-least-squares-with-forgetting-factor (RLSFF) estimation algorithm for multiple-output (MO) systems. The idea of a recursive formulation goes back to [4], where it is shown that the estimate converges to a value which minimizes the sum of output errors, whereby the outputs’ weights reduce exponentially with respect to their time of measurement; hence, the name forgetting factor. These results are established for single-input-single-output (SISO) systems and extended to MO systems in [8]. Since, the RLSFF algorithm has been proposed in many different variations, particularly for the SISO case (see e.g. [3], [13], [14]). Also, exponential convergence of the estimation error under the influence of the RLSFF identifier from [4] is shown in [9]. The latter work contains a deterministic analysis considering SISO systems without noise.

Despite the popularity of the RLSFF algorithm, to our best knowledge, a concise convergence result of the RLSFF for the MO case has not been published. Given the literature on adaptive control algorithms using a recursive-least-squares identifier in one way or the other, whose analysis is restricted to SISO systems (see e.g. [1], [2], [12], [15]), especially within the framework of model predictive control (see e.g. [5], [6], [7], [10]), and the recent article [8] dealing with MO systems, we are convinced that SISO-equivalent results for the MO case are of interest and facilitate further research in adaptive control; especially in the area of adaptive model predictive control, which has attracted relatively little attention [11].

Our note therefore closes the gap between the scalar and the MO case by extending the results from [8] and [9]. Under the usual assumption of persistence of excitation, we show that the RLSFF algorithm for MO systems minimizes a criterion for which the multiple outputs are weighted not only by their time of measurement (through the forgetting factor), but also among each other (via a user-defined weighting matrix). Furthermore, we demonstrate that the estimation error converges exponentially to the origin.

2 Main results
Consider the system

\[ y_{k+1} = \psi_k^T \theta, \]

where \( y_k \in \mathbb{R}^n \) is the measurable output, \( \psi_k \in \mathbb{R}^{m \times n} \) the regressor matrix, respectively at time \( k \), and \( \theta \in \mathbb{R}^m \) the constant parameter vector. Suppose that the parameter vector is uncertain and, thus, to be estimated. Toward this end, consider the following RLSFF algorithm.

\[ \hat{\theta}_{k+1} = \hat{\theta}_k + P_{k-1} \psi_k D_k^{-1} \left( y_{k+1} - \psi_k^T \hat{\theta}_k \right), \]

Email addresses: abruengge@eng.ucsd.edu (Sven Brüggemann), rbitmead@eng.ucsd.edu (Robert R. Bitmead).
where \( D_k = \lambda T + \psi_k^T P_{k-1} \psi_k \) with \( T = T^T > 0 \in \mathbb{R}^{n \times n} \) and constant forgetting factor \( \lambda \in (0, 1) \), and

\[
P_{k+1} = \lambda^{-1} \left( I - P_k \psi_{k+1} D_{k+1}^{-1} \psi_{k+1}^T \right) P_k, \tag{3}
\]
or equivalently,

\[
P_{k-1}^{-1} = \lambda P_{k-2}^{-1} + \psi_{k-1} T^{-1} \psi_{k-1}^T.
\tag{4}
\]

where \( P_{-1} \in \mathbb{R}^{m \times m} \) is symmetric positive definite. Define persistence of excitation as follows.

**Definition 1** The matrix sequence \( \{ \psi_k \} \) is said to be persistently exciting (PE) if for some constant \( S \) and all \( j \) there exist positive constants \( \alpha \) and \( \beta \) such that

\[
0 < \alpha I \leq \sum_{i=j}^{j+S} \psi_i \psi_i^T \leq \beta I < \infty.
\]

The following lemma is akin to the result in [8].

**Lemma 2** Suppose \( \{ \psi_k \} \) is PE. Then, for \( k \geq S \), the algorithm in (2) and (3) converges to the value \( \theta \) which minimizes

\[
\sum_{i=1}^{k} \lambda^{k-i} |y_i - \psi_{i-1}^T \theta|_T^2.
\]

**PROOF.** The proof is analogous to that of [8, Theorem 2] and hence omitted for brevity.

We have thus presented the objective which is minimized by the RLSFF algorithm in (2) and (3). In contrast to the existing literature, we explicitly incorporate a weight for each respective output via the matrix \( T \). We now wish to obtain an exponentially stable estimation error

\[
\hat{\theta}_k = \theta - \hat{\theta}_k.
\]

The corresponding theorem as an extension of [9] follows.

**Theorem 3** Suppose \( \{ \psi_k \} \) is PE. Then, for any initial condition \( \hat{\theta}_0 \), the estimation error \( \hat{\theta}_k \) converges exponentially to \( \theta \), i.e. for any \( \theta_0 \) there exist \( \gamma > 0 \) such that for all \( k \geq S \)

\[
|\hat{\theta}_k|^2 \leq \gamma \lambda^k |\hat{\theta}_0|^2.
\]

**PROOF.** The proof is divided into two parts. Part I establishes a lower bound on \( P_k^{-1} \). This is then used in part II to show exponential stability of the estimation error.

**Part I:** Recollect that if \( B \) is symmetric, then for any matrix \( A \),

\[
ABA^T \geq \lambda_{\min}(B) A A^T.
\]

This follows by definition of a positive definite matrix;

\[
x^T (ABA^T - \lambda_{\min}(B) A A^T) x
\]

\[
= x^T A B A^T x - \lambda_{\min}(B) x^T A A^T x
\]

\[
= (A^T x) B (A^T x) - \lambda_{\min}(B) x^T A A^T x
\]

\[
\geq \lambda_{\min}(B) (A^T x) A^T x - \lambda_{\min}(B) x^T A A^T x
\]

\[
= 0,
\]

where we use the fact that \( \lambda_{\min}(B) |x|^2 \leq |x|_B^2 \). It follows that if \( \{ \psi_k \} \) is PE, then

\[
P_j^{-1} + \cdots + P_{j+S-1}^{-1} \geq \sum_{k=j}^{j+S} \psi_{k-1} T^{-1} \psi_{k-1}^T \]

\[
\geq \lambda_{\min}(T^{-1}) \alpha I
\]

for all \( k \geq S \). Following [9, Lemma 1] leads to the lower bound

\[
P_k^{-1} \geq \frac{\lambda_{\min}(T^{-1}) \alpha (\lambda^{-1} - 1)}{\lambda^{-(S+1)} - 1} > 0
\]

(5)

for all \( k \geq S \).

**Part II:** By (2), one can write recursively the estimation error as

\[
\hat{\theta}_{k+1} = (I - P_{k-1} \psi_k D_k^{-1} \psi_k^T) \hat{\theta}_k.
\]

Consider the Lyapunov function candidate

\[
W_k = \hat{\theta}_k^T P_{k-1}^{-1} \hat{\theta}_k.
\]

Then, using the recursions in (4) and (6) yields

\[
W_{k+1} - W_k = \hat{\theta}_{k+1}^T P_{k-1}^{-1} \hat{\theta}_{k+1} - \hat{\theta}_k^T P_{k-1}^{-1} \hat{\theta}_k
\]

\[
= \hat{\theta}_k^T \left[ (\lambda - 1) P_{k-1}^{-1} - \lambda \psi_k D_k^{-1} \psi_k^T + C \right] \hat{\theta}_k,
\]

(8)

where

\[
C = \psi_k \left[ T^{-1} - D_k^{-1} \psi_k^T P_{k-1} \psi_k T^{-1} - \lambda D_k^{-1} \right.
\]

\[
- T^{-1} \psi_k^T P_{k-1} \psi_k D_k^{-1} + \lambda D_k^{-1} \psi_k^T P_{k-1} \psi_k D_k^{-1}
\]

\[
+ D_k^{-1} \psi_k^T P_{k-1} \psi_k T^{-1} \psi_k^T P_{k-1} \psi_k D_k^{-1} \right] \psi_k^T.
\]
To this end, multiply the inner term of both sides respectively and obtain

\[
\begin{align*}
I - T^{1/2} D_{k-1}^{-1} \psi_k^T P_{k-1} \psi_k T^{-1/2} - \lambda T^{1/2} D_{k}^{-1} T^{1/2} \\
- T^{-1/2} \psi_k^T P_{k-1} \psi_k D_{k}^{-1} T^{1/2} \\
+ \lambda T^{1/2} D_k^{-1} \psi_k^T P_{k-1} \psi_k D_{k}^{-1} T^{1/2} \\
+ T^{1/2} D_k^{-1} \psi_k^T P_{k-1} \psi_k T^{-1/2} P_{k-1} \psi_k D_k^{-1} T^{1/2}.
\end{align*}
\]

(9)

By defining

\[
\begin{align*}
\tilde{\psi} &= \psi_k T^{-1/2} \\
\tilde{D} &= T^{-1/2} D_k T^{-1/2} = \lambda I + \tilde{\psi}^T P_{k-1} \tilde{\psi}
\end{align*}
\]

one has that

\[
T^{1/2} D_k^{-1} = (D_k T^{-1/2})^{-1} = (\lambda T^{1/2} + \psi_k^T P_{k-1} \psi_k T^{-1/2})^{-1}
\]

\[
= \left( T^{1/2} (\lambda + T^{-1/2} \psi_k^T P_{k-1} \psi_k T^{-1/2}) \right)^{-1}
\]

\[
= \left( T^{1/2} (\lambda + \tilde{\psi}^T P_{k-1} \tilde{\psi}) \right)^{-1}
\]

\[
= \tilde{D}^{-1} T^{-1/2},
\]

so that for (9) it follows:

\[
\begin{align*}
I - \tilde{D}^{-1} \tilde{\psi}^T P_{k-1} \tilde{\psi} - \lambda \tilde{D}^{-1} \\
- \tilde{\psi}^T P_{k-1} \tilde{\psi} \tilde{D}^{-1} + \lambda \tilde{D}^{-1} \tilde{\psi}^T P_{k-1} \tilde{\psi} \tilde{D}^{-1} \\
+ \tilde{D}^{-1} \tilde{\psi}^T P_{k-1} \tilde{\psi} \tilde{D}^{-1}.
\end{align*}
\]

Observe that this can be reformulated as

\[
I - \tilde{D}^{-1}
\begin{pmatrix}
\tilde{\psi}^T P_{k-1} \tilde{\psi} + I
\end{pmatrix}
\]

\[
+ \tilde{D}^{-1}
\begin{pmatrix}
- \tilde{\psi}^T P_{k-1} \tilde{\psi} \tilde{D}^{-1} + \lambda \tilde{D}^{-1} \tilde{\psi}^T P_{k-1} \tilde{\psi} \tilde{D}^{-1}
\end{pmatrix}
\]

which is clearly zero and, thus, so is \( C \).

Therefore, the difference related to the Lyapunov function candidate in (8)

\[
W_{k+1} - W_k = \tilde{\theta}_k^T \left[ (\lambda - 1) P_{k-1}^{-1} - \lambda \psi_k D_k^{-1} \psi_k^T \right] \tilde{\theta}_k
\]

\[
\leq (\lambda - 1) \tilde{\theta}_k^T P_{k-1}^{-1} \tilde{\theta}_k
\]

\[
= (\lambda - 1) W_k,
\]

so that

\[
W_{k+1} \leq \lambda W_k \leq \lambda^{k+1} W_0 = \lambda^{k+1} \tilde{\theta}_0^T P_{k-1}^{-1} \tilde{\theta}_0.
\]

Finally, combining this inequality with the definition of \( W_k \) in (7) and the lower bound of \( P_{k-1}^{-1} \) in (5) leads to

\[
|\tilde{\theta}_k|^2 \leq \frac{\lambda^{-(S+1)} - 1}{\lambda_{\text{min}}(T-1) \alpha (\lambda-1) \lambda_{\text{max}} (P_{k-1}^{-1}) \lambda^k} \lambda |\tilde{\theta}_0|^2
\]

for all \( k \geq S \).

\[\blacksquare\]

**Remark 4** Note that exponential stability of the estimation error implies a bounded error in the presence of bounded additive noise. Thus, for additive bounded noise on system (1), the estimation error converges to a ball centered on the true parameter vector with a radius proportional to the bound on the disturbance.

### 3 Conclusion

We have shown that the RLSFF estimation algorithm for MO systems shares the properties of those for single output systems, i.e. it minimizes a similar cost function, where outputs are weighted among one another by a given matrix, and it induces exponential convergence of the estimate to the true parameter vector.

### References

[1] Karl Johan Åström and Björn Wittenmark. *Adaptive Control*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2nd edition, 1994.

[2] Chi Sing Leung, G. H. Young, J. Sum, and Wing-Kay Kan. On the regularization of forgetting recursive least square. *IEEE Transactions on Neural Networks*, 10(6):1482–1486, Nov 1999.

[3] S. Dasgupta and Yih-Fang Huang. Asymptotically convergent modified recursive least-squares with data-dependent updating and forgetting factor for systems with bounded noise. *IEEE Transactions on Information Theory*, 33(3):383–392, May 1987.

[4] Graham C. Goodwin and Robert L. Payne. *Dynamic system identification: experiment design and data analysis*. Academic Press New York, 1977.

[5] Tor Aksel N. Heirung, Bjarne Foss, and B. Erik Ydstie. MPC-based dual control with online experiment design. *Journal of Process Control*, 32:64 – 76, 2015.

[6] Tor Aksel N. Heirung, Tito L.M. Santos, and Ali Mesbah. Model predictive control with active learning for stochastic systems with structural model uncertainty: Online model discrimination. *Computers & Chemical Engineering*, 128:128 – 140, 2019.

[7] Tor Aksel N. Heirung, B. Erik Ydstie, and Bjarne Foss. Dual adaptive model predictive control. *Automatica*, 80:340 – 348, 2017.

[8] S. A. U. Islam and D. S. Bernstein. Recursive least squares for real-time implementation [lecture notes]. *IEEE Control Systems Magazine*, 39(3):82–85, June 2019.

[9] Richard M. Johnstone, C. Richard Johnson, Robert R. Bitmead, and Brian D.O. Anderson. Exponential convergence of recursive least squares with exponential forgetting factor. *Systems & Control Letters*, 2(2):77 – 82, 1982.
[10] Giancarlo Marafioti, Robert R. Bitmead, and Morten Hovd. Persistently exciting model predictive control. *International Journal of Adaptive Control and Signal Processing*, 28(6):536–552, 2014.

[11] David Q. Mayne. Model predictive control: Recent developments and future promise. *Automatica*, 50(12):2967–2986, 2014.

[12] C. Paleologu, J. Benesty, and S. Ciochină. A practical variable forgetting factor recursive least-squares algorithm. In *2014 11th International Symposium on Electronics and Telecommunications (ISETC)*, pages 1–4, Nov 2014.

[13] Mario E. Salgado, Graham C. Goodwin, and Richard H. Middleton. Modified least squares algorithm incorporating exponential resetting and forgetting. *International Journal of Control*, 47(2):477–491, 1988.

[14] Shu-Hung Leung and C. F. So. Gradient-based variable forgetting factor rls algorithm in time-varying environments. *IEEE Transactions on Signal Processing*, 53(8):3141–3150, Aug 2005.

[15] A. Vahidi, A. Stefanopoulou, and H. Peng. Recursive least squares with forgetting for online estimation of vehicle mass and road grade: theory and experiments. *Vehicle System Dynamics*, 43(1):31–55, 2005.