Recursive Least Squares with Variable-Direction Forgetting

Compensating for the loss of persistency

Ankit Goel, Adam L. Bruce, and Dennis S. Bernstein

POC: A. Goel (ankgoel@umich.edu)

The ability to estimate parameters depends on two things, namely, identifiability [1], which is ability to distinguish distinct parameters, and persistent excitation, which refers to the spectral content of the signals needed to ensure convergence of the parameter estimates to the true parameter values [2]–[4]. Roughly speaking, the level of persistency must be commensurate with the number of unknown parameters. For example, a harmonic input has two-dimensional persistency and thus can be used to identify two parameters, whereas white noise is sufficiently persistent for identifying an arbitrary number of parameters. Within the context of adaptive control, persistent excitation is needed to avoid bursting [5]; recent research has focused on relaxing these requirements [6]–[8].

Under persistent excitation, a key issue in practice is the rate of convergence, especially under changing conditions. For example, the parameters of a system may change abruptly, and the goal is to ensure fast convergence to the modified parameter values. In this case, it turns out that the rate of convergence depends on the ability to forget past parameters and incorporate new information. As discussed in "Summary," the ability to accommodate new information depends on the ability to forget; the ability to forget is thus crucial to the ability to learn. This paradox is widely recognized, and effective forgetting is of intense interest in machine learning [9]–[12].

In the first half of the present article, classical forgetting within the context of recursive least squares (RLS) is considered. In the classical RLS formulation [13]–[16], a constant forgetting factor $\lambda \in (0, 1]$ can be set by the user. However, it often occurs in practice that the performance of RLS is extremely sensitive to the choice of λ, and suitable values in the range 0.99 to 0.9999 are typically found by trial-and-error testing. This difficulty has motivated extensions of classical RLS in the form of variable-rate forgetting [17]–[23], constant trace adjustment, covariance resetting, and covariance modification [24], [25].

In the second half of this article, *variable-direction forgetting* (VDF), a technique that complements variable-rate forgetting is considered. Direction-dependent forgetting has been widely studied within the context of recursive least squares [26]–[32]. In the absence of persistent excitation, new information is confined to a limited number of directions. The goal of VDF is thus to determine these directions and thereby constrain forgetting to the directions in which new information is available. VDF allows RLS to operate without divergence during periods of loss of persistency.

The goal of this tutorial article is to investigate the effect of forgetting within the context of RLS in order to motivate the need for VDF. With this motivation in mind, the article develops and illustrates RLS with VDF. The presentation is intended for graduate students who may wish to understand and apply this technique to system identification for modeling and adaptive control. Table 1 and 2 summarizes the results and examples in this article. Some of the content in this article appeared in preliminary form in [33].

Although, in practical applications, all sensor measurements are corrupted by noise, the effect of sensor noise is not considered in this article in order to focus on the loss of persistency. Alternative interpretations of RLS in the special case of zero-mean, white sensor noise are presented in "RLS as a One-Step Optimal Predictor" and "RLS as a Maximum Likelihood Estimator".

Recursive Least Squares

Consider the model

$$y_k = \phi_k \theta, \tag{1}$$

where, for all $k \geq 0$, $y_k \in \mathbb{R}^p$ is the measurement, $\phi_k \in \mathbb{R}^{p \times n}$ is the regressor matrix, and $\theta \in \mathbb{R}^n$ is the vector of unknown parameters. The goal is to estimate θ as new data become available. One approach to this problem is to minimize the quadratic cost function

$$J_k(\hat{\theta}) \triangleq \sum_{i=0}^{k} \lambda^{k-i} (y_i - \phi_i \hat{\theta})^{\mathrm{T}} (y_i - \phi_i \hat{\theta}) + \lambda^{k+1} (\hat{\theta} - \theta_0)^{\mathrm{T}} R (\hat{\theta} - \theta_0), \tag{2}$$

where $\lambda \in (0, 1]$ is the *forgetting factor*, $R \in \mathbb{R}^{n \times n}$ is positive definite, and $\theta_0 \in \mathbb{R}^n$ is the initial estimate of θ. The forgetting factor applies higher weighting to more recent data, thereby enhancing the ability of RLS to use incoming data to estimate time-varying parameters. The following result is *recursive least squares*.

Theorem 1: For all $k \geq 0$, let $\phi_k \in \mathbb{R}^{p \times n}$ and $y_k \in \mathbb{R}^p$, let $R \in \mathbb{R}^{n \times n}$ be positive definite, and define $P_0 \triangleq R^{-1}$, $\theta_0 \in \mathbb{R}^n$, and $\lambda \in (0, 1]$. Furthermore, for all $k \geq 0$, denote the minimizer

TABLE 1: Summary of definitions and results in this article.

Definition 1	Persistently exciting regressor
Definition 2	Lyapunov stable equilibrium
Definition 3	Uniformly Lyapunov stable equilibrium
Definition 4	Globally aymptotically stable equilibrium
Definition 5	Uniformly globally geometrically stable equilibrium
Theorem 1-2	Recursive least squares (RLS)
Theorem 3-5	Lyapunov stability theorems
Theorem 6	Lyapunov analysis of RLS for $\lambda \in (0,1)$
Theorem 7	Stability analysis of RLS for $\lambda \in (0,1]$ based on θ_k
Theorem S1	A Quadratic Cost Function for Variable-Direction RLS
Proposition 1	Recursive update of P_k^{-1} with uniform-direction forgetting
Proposition 2	Data-dependent subspace constraint on θ_k
Proposition 3	Bounds on P_k for $\lambda = 1$
Proposition 4	Bounds on P_k for $\lambda \in (0,1)$
Proposition 5	Converse of Proposition 4
Proposition 6	Convergence of z_k with uniform-direction forgetting
Proposition 7	Persistent excitation and $\mathcal{A}_k$
Proposition 8	Recursive update of P_k^{-1} with variable-direction forgetting
Proposition 9	Convergence of z_k with variable-direction forgetting
Proposition 10	Bounds on P_k with variable-direction forgetting

of (2) by

$$\theta_{k+1} = \underset{\hat{\theta}\in\mathbb{R}^n}{\operatorname{argmin}}\ J_k(\hat{\theta}). \tag{3}$$

Then, for all $k \geq 0$, θ_{k+1} is given by

$$P_{k+1} = \frac{1}{\lambda}P_k - \frac{1}{\lambda}P_k\phi_k^{\mathrm{T}}\left(\lambda I_p + \phi_k P_k \phi_k^{\mathrm{T}}\right)^{-1}\phi_k P_k, \tag{4}$$

$$\theta_{k+1} = \theta_k + P_{k+1}\phi_k^{\mathrm{T}}(y_k - \phi_k\theta_k). \tag{5}$$

Proof: See [13]. □

2 The following result is a variation of Theorem 1, where the updates of P_k and θ_k are reversed.

Theorem 2: For all $k \geq 0$, let $\phi_k \in \mathbb{R}^{p\times n}$ and $y_k \in \mathbb{R}^p$, let $R \in \mathbb{R}^{n\times n}$ be positive definite,

TABLE 2: Summary of examples in this article.

Example 1	P_k converges to zero without persistent excitation
Example 2	Persistent excitation and bounds on P_k^{-1}
Example 3	Lack of persistent excitation and bounds on P_k^{-1}
Example 4	Convergence of z_k and θ_k
Example 5	Using $\kappa(P_k)$ to determine whether $(\phi_k)_{k=0}^{\infty}$ is persistently exciting
Example 6	Effect of λ on the rate of convergence of θ_k
Example 7	Lack of persistent excitation in scalar estimation
Example 8	Subspace constrained regressor
Example 9	Effect of lack of persistent excitation on θ_k
Example 10	Lack of persistent excitation and the information-rich subspace
Example 11	Variable-direction forgetting for a regressor lacking persistent excitation
Example 12	Effect of variable-direction forgetting on θ_k

and define $P_0 \triangleq R^{-1}$, $\theta_0 \in \mathbb{R}^n$, and $\lambda \in (0,1]$. Furthermore, for all $k \geq 0$, denote the minimizer of (2) by (3). Then, for all $k \geq 0$, θ_{k+1} is given by

$$\theta_{k+1} = \theta_k + P_k\phi_k^{\mathrm{T}}(\lambda I + \phi_k P_k \phi_k^{\mathrm{T}})^{-1}(y_k - \phi_k\theta_k), \tag{6}$$

$$P_{k+1} = \frac{1}{\lambda}P_k - \frac{1}{\lambda}P_k\phi_k^{\mathrm{T}}(\lambda I + \phi_k P_k \phi_k^{\mathrm{T}})^{-1}\phi_k P_k. \tag{7}$$

Proof: See [13]. □

Proposition 1: Let $\lambda \in (0,\infty)$, and let $(P_k)_{k=0}^{\infty}$ be a sequence of $n \times n$ positive-definite matrices. Then, for all $k \geq 0$, $(P_k)_{k=0}^{\infty}$ satisfies (4) if and only if, for all $k \geq 0$, $(P_k)_{k=0}^{\infty}$ satisfies

$$P_{k+1}^{-1} = \lambda P_k^{-1} + \phi_k^{\mathrm{T}}\phi_k. \tag{8}$$

Proof: To prove necessity, it follows from (8) and matrix-inversion lemma, that

$$\begin{aligned}
P_{k+1} &= (\lambda P_k^{-1} + \phi_k^{\mathrm{T}}\phi_k)^{-1} \\
&= (\lambda P_k^{-1})^{-1} - (\lambda P_k^{-1})^{-1}\phi_k^{\mathrm{T}}(I_p + \phi_k(\lambda P_k^{-1})^{-1}\phi_k^{\mathrm{T}})^{-1}\phi_k(\lambda P_k^{-1})^{-1} \\
&= \frac{1}{\lambda}P_k - \frac{1}{\lambda}P_k\phi_k^{\mathrm{T}}\left(\lambda I_p + \phi_k P_k \phi_k^{\mathrm{T}}\right)^{-1}\phi_k P_k.
\end{aligned}$$

2 Reversing these steps proves sufficiency. □

Let $k \geq 0$. By defining the *parameter error*

$$\tilde{\theta}_k \triangleq \theta_k - \theta, \tag{9}$$

it follows that

$$\phi_i\theta_k - y_i = \phi_i\tilde{\theta}_k. \tag{10}$$

Using (10) with k replaced by $k+1$, it follows that the minimum value of J_k is given by

$$J_k(\theta_{k+1}) = \sum_{i=0}^{k} \lambda^{k-i}\tilde{\theta}_{k+1}^{\mathrm{T}}\phi_i^{\mathrm{T}}\phi_i\tilde{\theta}_{k+1} + \lambda^{k+1}(\tilde{\theta}_{k+1}-\tilde{\theta}_0)^{\mathrm{T}}R(\tilde{\theta}_{k+1}-\tilde{\theta}_0). \tag{11}$$

Furthermore, (5) and (9) imply that $\tilde{\theta}_k$ satisfies

$$\tilde{\theta}_{k+1} = (I_n - P_{k+1}\phi_k^{\mathrm{T}}\phi_k)\tilde{\theta}_k \tag{12}$$

$$= \lambda P_{k+1}P_k^{-1}\tilde{\theta}_k. \tag{13}$$

Finally, it follows from (13) that, for all $k, l \geq 0$,

$$\tilde{\theta}_k = \lambda^{k-l}P_kP_l^{-1}\tilde{\theta}_l. \tag{14}$$

The following result shows that the estimate θ_k of θ is constrained to a data-dependent
2 subspace. Let $\mathcal{R}(A)$ denote the range of the matrix A.

Proposition 2: For all $k \geq 0$, let $\phi_k \in \mathbb{R}^{p\times n}$ and $y_k \in \mathbb{R}^p$, let $R \in \mathbb{R}^{n\times n}$ be positive definite, let $\theta_0 \in \mathbb{R}^n$, let $\lambda \in (0,1]$, and define θ_{k+1} by (3). Then, θ_{k+1} satisfies

$$\left(\sum_{i=0}^{k} \lambda^{k-i}\phi_i^{\mathrm{T}}\phi_i + \lambda^{k+1}R\right)\theta_{k+1} = \sum_{i=0}^{k} \lambda^{k-i}\phi_i^{\mathrm{T}}y_i + \lambda^{k+1}R\theta_0. \tag{15}$$

Furthermore,

$$\theta_{k+1} \in \mathcal{R}(\Phi_k^{\mathrm{T}}\Phi_k + R^{-1}\Phi_k^{\mathrm{T}}\Phi_kR^{-1} + \theta_0\theta_0^{\mathrm{T}}), \tag{16}$$

where

$$\Phi_k \triangleq [\phi_0^{\mathrm{T}} \quad \cdots \quad \phi_k^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{(k+1)p\times n}. \tag{17}$$

Proof: Note that

$$J_k(\hat{\theta}) = \hat{\theta}^{\mathrm{T}}A_k\hat{\theta} + \hat{\theta}^{\mathrm{T}}b_k + c_k,$$

where

$$A_k \triangleq \sum_{i=0}^{k} \lambda^{k-i}\phi_i^{\mathrm{T}}\phi_i + \lambda^{k+1}R,$$

$$b_k \triangleq \sum_{i=0}^{k} -\lambda^{k-i}\phi_i^{\mathrm{T}}y_i - \lambda^{k+1}R\theta_0,$$

$$c_k \triangleq \sum_{i=0}^{k} \lambda^{k-i}y_i^{\mathrm{T}}y_i + \lambda^{k+1}\theta_0^{\mathrm{T}}R\theta_0.$$

Since A_k is positive definite, it follows from Lemma 1 in [13] that the minimizer θ_{k+1} of J_k satisfies (15).

Next, define $W_k \overset{\triangle}{=} \operatorname{diag}(\lambda^{-1}I_p, \ldots, \lambda^{-1-k}I_p) \in \mathbb{R}^{(k+1)p\times(k+1)p}$. Using (15) and Lemma 1 from "Three Useful Lemmas," it follows that

$$\begin{aligned}
\theta_{k+1} &= \left(I_n + \Phi_k^{\mathrm{T}} W_k \Phi_k\right)^{-1} \left(\sum_{i=0}^{k} \lambda^{-i-1} R^{-1} \phi_i^{\mathrm{T}} y_i + \theta_0\right) \\
&= \sum_{i=0}^{k} \left(I_n + \Phi_k^{\mathrm{T}} W_k \Phi_k\right)^{-1} \lambda^{-i-1} R^{-1} \phi_i^{\mathrm{T}} y_i + \left(I_n + \Phi_k^{\mathrm{T}} W_k \Phi_k\right)^{-1} \theta_0 \\
&\in \sum_{i=0}^{k} \mathcal{R}([\Phi_k^{\mathrm{T}} \quad R^{-1}\phi_i^{\mathrm{T}}]) + \mathcal{R}([\Phi_k^{\mathrm{T}} \quad \theta_0]) \\
&= \mathcal{R}([\Phi_k^{\mathrm{T}} \quad R^{-1}\Phi_k^{\mathrm{T}} \quad \theta_0]) \\
&= \mathcal{R}(\Phi_k^{\mathrm{T}}\Phi_k + R^{-1}\Phi_k^{\mathrm{T}}\Phi_k R^{-1} + \theta_0\theta_0^{\mathrm{T}}).
\end{aligned}$$

$\square$

Table 3 summarizes various expressions for the RLS variables.

TABLE 3: Alternative expressions for the RLS variables.

Variable	Expression	Equation
P_k	• $P_{k+1} = \frac{1}{\lambda}P_k - \frac{1}{\lambda}P_k\phi_k^{\mathrm{T}}\left(\lambda I_p + \phi_k P_k \phi_k^{\mathrm{T}}\right)^{-1}\phi_k P_k$	(4)
	• $P_{k+1}^{-1} = \lambda P_k^{-1} + \phi_k^{\mathrm{T}}\phi_k$	(8)
	• $P_{k+1}^{-1} = \lambda^{k+1}P_0^{-1} + \sum_{i=0}^{k}\lambda^{k-i}\phi_i^{\mathrm{T}}\phi_i$	(8)
θ_k	• $\theta_{k+1} = \theta_k + P_{k+1}\phi_k^{\mathrm{T}}(y_k - \phi_k\theta_k)$	(5)
	• $\theta_{k+1} = \theta_k + P_k\phi_k^{\mathrm{T}}(\lambda I_p + \phi_k P_k \phi_k^{\mathrm{T}})^{-1}(y_k - \phi_k\theta_k)$	(6)
	• $\theta_{k+1} = P_{k+1}\left(\sum_{i=0}^{k}\lambda^{k-i}\phi_i^{\mathrm{T}}y_i + \lambda^{k+1}P_0^{-1}\theta_0\right)$	(15)
$\tilde{\theta}_k$	• $\tilde{\theta}_k = \theta_k - \theta$	(9)
	• $\tilde{\theta}_{k+1} = (I_n - P_{k+1}\phi_k^{\mathrm{T}}\phi_k)\tilde{\theta}_k$	(12)
	• $\tilde{\theta}_{k+1} = \lambda P_{k+1}P_k^{-1}\tilde{\theta}_k$	(13)
	• $\tilde{\theta}_k = \lambda^{k-l}P_k P_l^{-1}\tilde{\theta}_l$	(14)

Persistent Excitation and Forgetting

This section defines persistent excitation of the regressor sequence and investigates the effect of persistent excitation and forgetting on P_k. For all $j \geq 0$ and $k \geq j$, define

$$F_{j,k} \triangleq \sum_{i=j}^{k} \phi_i^{\mathrm{T}} \phi_i. \tag{18}$$

Definition 1: The sequence $(\phi_k)_{k=0}^{\infty} \subset \mathbb{R}^{p \times n}$ is *persistently exciting* if there exist $N \geq n/p$ and $\alpha, \beta \in (0, \infty)$ such that, for all $j \geq 0$,

$$\alpha I_n \leq F_{j,j+N} \leq \beta I_n. \tag{19}$$

Suppose that $(\phi_k)_{k=0}^{\infty}$ is persistently exciting and (19) is satisfied for given values of N, α, β. Then, with suitably modified values of α and β, (19) is satisfied for all larger values of N. For example, if N is replaced by $2N$, then (19) is satisfied with α replaced by 2α and β replaced by 2β. The following result expresses (8) in terms of $F_{0,k}$ in the case where $\lambda = 1$.

Lemma 1: Let $\lambda = 1$ and, for all $k \geq 1$, define P_k as in Theorem 1. Then,

$$P_k^{-1} = F_{0,k} + P_0^{-1}. \tag{20}$$

The following result shows that, if $(\phi_k)_{k=0}^{\infty}$ is persistently exciting and $\lambda = 1$, then P_k converges to zero.

Proposition 3: Assume that $(\phi_k)_{k=0}^{\infty} \in \mathbb{R}^{p \times n}$ is persistently exciting, let N, α, β be given by Definition 1, let $R \in \mathbb{R}^{n \times n}$ be positive definite, define $P_0 \triangleq R^{-1}$, let $\lambda = 1$, and, for all $k \geq 0$, let P_k be given by (4). Then, for all $k \geq N+1$,

$$\left\lfloor \tfrac{k}{N+1} \right\rfloor \alpha I_n + P_0^{-1} \leq P_k^{-1} \leq \left\lceil \tfrac{k}{N+1} \right\rceil \beta I_n + P_0^{-1}. \tag{21}$$

Furthermore,

$$\lim_{k \to \infty} P_k = 0. \tag{22}$$

Proof: First, note that, for all $k \geq 0$,

$$\begin{aligned} F_{0,k} &= \sum_{i=1}^{\left\lfloor \frac{k}{N+1} \right\rfloor} F_{(i-1)(N+1),i(N+1)-1} + F_{\left\lfloor \frac{k}{N+1} \right\rfloor (N+1),k} \\ &\leq \sum_{i=1}^{\left\lceil \frac{k}{N+1} \right\rceil} F_{(i-1)(N+1),i(N+1)-1}, \end{aligned}$$

and thus (19) implies that

$$
\begin{aligned}
\left\lfloor \tfrac{k}{N+1} \right\rfloor \alpha I_n &\leq \sum_{i=1}^{\left\lfloor \frac{k}{N+1} \right\rfloor} F_{(i-1)(N+1),i(N+1)-1} \\
&\leq \sum_{i=1}^{\left\lceil \frac{k}{N+1} \right\rceil} F_{(i-1)(N+1),i(N+1)-1} \\
&\leq \left\lceil \tfrac{k}{N+1} \right\rceil \beta I_n. \qquad (23)
\end{aligned}
$$

It follows from Lemma 1 and (23) that, for all $k \geq N+1$,

$$
\begin{aligned}
\left\lfloor \tfrac{k}{N+1} \right\rfloor \alpha I_n + P_0^{-1} &\leq F_{0,\left\lfloor \frac{k}{N+1} \right\rfloor (N+1)-1} + P_0^{-1} \\
&\leq F_{0,k} + P_0^{-1} \\
&= P_k^{-1} \\
&\leq F_{0,\left\lceil \frac{k}{N+1} \right\rceil (N+1)-1} + P_0^{-1} \\
&\leq \left\lceil \tfrac{k}{N+1} \right\rceil \beta I_n + P_0^{-1}.
\end{aligned}
$$

Finally, it follows from (21) that $\lim_{k\to\infty} P_k = 0$. □

The following example shows that $\lim_{k\to\infty} P_k = 0$ does not imply that $(\phi_k)_{k=0}^{\infty}$ is persistently exciting.

Example 1: P_k converges to zero without persistent excitation. For all $k \geq 0$, let $\phi_k = \frac{1}{\sqrt{k+1}}$. Let $\lambda = 1$. For all $N \geq 1$, note that $F_{j,j+N} \leq \frac{N+1}{j+1}$, and thus there does not exist α satisfying (19). Hence, $(\phi_k)_{k=0}^{\infty}$ is not persistently exciting. However, it follows from (8) that, for all $k \geq 0$,

$$
P_k^{-1} = \sum_{i=0}^{k} \frac{1}{i+1} + P_0^{-1}. \qquad (24)
$$

Thus, $\lim_{k\to\infty} P_k = 0$. ◇

The following result given in [34] shows that, if $(\phi_k)_{k=0}^{\infty}$ is persistently exciting and $\lambda \in (0,1)$, then P_k is bounded.

Proposition 4: Assume that $(\phi_k)_{k=0}^{\infty} \in \mathbb{R}^{p\times n}$ is persistently exciting, let N, α, β be given by Definition 1, let $R \in \mathbb{R}^{n\times n}$ be positive definite, define $P_0 \triangleq R^{-1}$, let $\lambda \in (0,1)$, and, for all $k \geq 0$, let P_k be given by (4). Then, for all $k \geq N+1$,

$$
\frac{\lambda^N(1-\lambda)\alpha}{1-\lambda^{N+1}} I_n \leq P_k^{-1} \leq \frac{\beta}{1-\lambda^{N+1}} I_n + P_N^{-1}. \qquad (25)
$$

Proof: It follows from (8) that, for all $i \geq 0$, $\lambda P_i^{-1} \leq P_{i+1}^{-1}$ and $\phi_i^{\mathrm{T}}\phi_i \leq P_{i+1}^{-1}$, and thus, for all $i, j \geq 0$, $\lambda^j P_i^{-1} \leq P_{i+j}^{-1}$. Hence, for all $k \geq N+1$,

$$\begin{aligned}\alpha I_n &\leq \sum_{i=k-N-1}^{k-1} \phi_i^{\mathrm{T}}\phi_i \\ &\leq \sum_{i=k-N}^{k} P_i^{-1} \\ &\leq (\lambda^{-N} + \cdots + 1)P_k^{-1} \\ &= \frac{1-\lambda^{N+1}}{\lambda^N(1-\lambda)}P_k^{-1},\end{aligned}$$

which proves the first inequality in (25). To prove the second inequality in (25), note that, for all $k \geq N+1$,

$$\begin{aligned}P_k^{-1} &\leq \frac{1-\lambda}{1-\lambda^{N+1}} \sum_{i=k-1}^{k+N-1} P_{i+1}^{-1} \\ &\leq \frac{1-\lambda}{1-\lambda^{N+1}} \left(\lambda \sum_{i=k-1}^{k+N-1} P_i^{-1} + \beta I_n\right) \\ &\leq \frac{1-\lambda}{1-\lambda^{N+1}} \left(\lambda^k \sum_{i=0}^{N} P_i^{-1} + \frac{1-\lambda^k}{1-\lambda}\beta I_n\right) \\ &\leq \lambda^{k-N} P_N^{-1} + \frac{(1-\lambda^k)\beta}{1-\lambda^{N+1}} I_n. \\ &\leq P_N^{-1} + \frac{\beta}{1-\lambda^{N+1}} I_n.\end{aligned}$$

□

The next result, which is an immediate consequence of (8), is a converse of Proposition 4.

Proposition 5: Define ϕ_k, y_k, R, and P_0 as in Theorem 1, let $\lambda \in (0,1)$, and let P_k be given by (4). Furthermore, assume there exist $\overline{\alpha}, \overline{\beta} \in (0,\infty)$ such that, for all $k \geq 0$, $\overline{\alpha} I_n \leq P_k^{-1} \leq \overline{\beta} I_n$. Let $N \geq \frac{\lambda\overline{\beta}-\overline{\alpha}}{(1-\lambda)\overline{\alpha}}$. Then, for all $j \geq 0$,

$$[(1+(1-\lambda)N)\overline{\alpha} - \lambda\overline{\beta}]I_n \leq \sum_{i=j}^{j+N} \phi_i^{\mathrm{T}}\phi_i \leq \frac{1-\lambda^{N+1}}{\lambda^N(1-\lambda)}\overline{\beta} I_n. \tag{26}$$

Consequently, $(\phi_k)_{k=0}^{\infty}$ is persistently exciting.

Proof: Note that, for all $j \geq 0$,

$$\begin{aligned}
[(1+(1-\lambda)N)\overline{\alpha} - \lambda\overline{\beta}]I_n &= \overline{\alpha}I_n + (1-\lambda)N\overline{\alpha}I_n - \overline{\beta}I_n \\
&\leq P_{j+N+1}^{-1} + (1-\lambda)\sum_{i=j+1}^{j+N} P_i^{-1} - \lambda P_j^{-1} \\
&= \sum_{i=j}^{j+N}(P_{i+1}^{-1} - \lambda P_i^{-1}) \\
&= \sum_{i=j}^{j+N} \phi_i^{\mathrm{T}}\phi_i,
\end{aligned}$$

which proves the first inequality in (26). To prove the second inequality in (26), note that (8) implies that, for all $i \geq 0$, $\lambda P_i^{-1} \leq P_{i+1}^{-1}$ and $\phi_i^{\mathrm{T}}\phi_i \leq P_{i+1}^{-1}$, and thus, for all $i, j \geq 0$, $\lambda^j P_i^{-1} \leq P_{i+j}^{-1}$. Hence, for all $j \geq 0$,

$$\begin{aligned}
\sum_{i=j}^{j+N} \phi_i^{\mathrm{T}}\phi_i &\leq \sum_{i=j}^{j+N} P_{i+1}^{-1} \\
&\leq (\lambda^{-N} + \cdots + 1)P_{j+N+1}^{-1} \\
&\leq \frac{1-\lambda^{N+1}}{\lambda^N(1-\lambda)}\overline{\beta}I_n.
\end{aligned}$$

Finally, it follows from Definition 1 with $N \geq \frac{\lambda\overline{\beta}-\overline{\alpha}}{(1-\lambda)\overline{\alpha}}$, $\alpha = (1+(1-\lambda)N)\overline{\alpha} - \lambda\overline{\beta}$, and $\beta = \frac{1-\lambda^{N+1}}{\lambda^N(1-\lambda)}\overline{\beta}$, that $(\phi_k)_{k=0}^{\infty}$ is persistently exciting. □

The proof of Proposition 5 shows that the condition $N \geq \frac{\lambda\overline{\beta}-\overline{\alpha}}{(1-\lambda)\overline{\alpha}}$ is needed to satisfy the lower bound in Definition 1. However, the upper bound in Definition 1 is satisfied for all $N \geq 1$.

Example 2: Persistent excitation and bounds on P_k^{-1}. Let $\phi_k = [u_k \ \ u_{k-1}]$, where u_k is the periodic signal

$$u_k = \sin\frac{2\pi k}{17} + \sin\frac{2\pi k}{23} + \sin\frac{2\pi k}{53}. \tag{27}$$

Figure 1 shows the singular values of $F_{j,j+N}$ for $N = 2$ and $N = 10$, as well as the singular values of P_k^{-1} with the corresponding upper and lower bounds given by (25) for $N = 2$ and $N = 10$. ◇

Example 3: Lack of persistent excitation and bounds on P_k^{-1}. Let $\phi_k = [u_k \ \ u_{k-1}]$, where u_k is given by (27) for all $k < 2500$ and $u_k = 1$ for all $k \geq 2500$. Figure 2 shows the singular values of $F_{j,j+2}$ and the singular values of P_k^{-1} for $\lambda = 1$ and $\lambda = 0.9$, respectively. Note that, for $\lambda = 1$, one of the singular values of P_k^{-1} diverges, whereas, for $\lambda \in (0, 1)$, one of singular values of P_k^{-1} converges to zero. ◇

The following result shows that the *predicted error* $z_k \stackrel{\triangle}{=} \phi_k\theta_k - y_k$ converges to zero whether or not $(\phi_k)_{k=0}^\infty$ is persistent.

Proposition 6: For all $k \geq 0$, let $\phi_k \in \mathbb{R}^{p\times n}$ and $y_k \in \mathbb{R}^p$, let $R \in \mathbb{R}^{n\times n}$ be positive definite, and let $P_0 = R^{-1}$, $\theta_0 \in \mathbb{R}^n$, and $\lambda \in (0,1]$. Furthermore, for all $k \geq 0$, let P_k and θ_k be given by (4) and (5), respectively, and define the *predicted error* $z_k \stackrel{\triangle}{=} \phi_k\theta_k - y_k$. Then,

$$\lim_{k\to\infty} z_k = 0. \tag{28}$$

Proof: For all $k \geq 0$, note that $z_k = \phi_k\tilde{\theta}_k$, and define $V_k \stackrel{\triangle}{=} \tilde{\theta}_k^{\mathrm{T}} P_k^{-1}\tilde{\theta}_k$. Note that, for all $k \geq 0$ and $\tilde{\theta}_k \in \mathbb{R}^n$, $V_k \geq 0$. Furthermore, for all $k \geq 0$,

$$\begin{aligned}
V_{k+1} - V_k &= \tilde{\theta}_{k+1}^{\mathrm{T}} P_{k+1}^{-1}\tilde{\theta}_{k+1} - \tilde{\theta}_k^{\mathrm{T}} P_k^{-1}\tilde{\theta}_k \\
&= \lambda^2\tilde{\theta}_k^{\mathrm{T}} P_k^{-1} P_{k+1} P_k^{-1}\tilde{\theta}_k - \tilde{\theta}_k^{\mathrm{T}} P_k^{-1}\tilde{\theta}_k \\
&= (\lambda\tilde{\theta}_{k+1}^{\mathrm{T}} - \tilde{\theta}_k^{\mathrm{T}}) P_k^{-1}\tilde{\theta}_k \\
&= -[(1-\lambda)\tilde{\theta}_k^{\mathrm{T}} + \lambda\tilde{\theta}_k^{\mathrm{T}}\phi_k^{\mathrm{T}}\phi_k P_{k+1}] P_k^{-1}\tilde{\theta}_k \\
&= -[(1-\lambda)\tilde{\theta}_k^{\mathrm{T}} P_k^{-1}\tilde{\theta}_k + \lambda\tilde{\theta}_k^{\mathrm{T}}\phi_k^{\mathrm{T}}\phi_k P_{k+1} P_k^{-1}\tilde{\theta}_k] \\
&= -[(1-\lambda)\tilde{\theta}_k^{\mathrm{T}} P_k^{-1}\tilde{\theta}_k + \tilde{\theta}_k^{\mathrm{T}}\phi_k^{\mathrm{T}}[I_p - \phi_k P_k\phi_k^{\mathrm{T}}(\lambda I_p + \phi_k P_k\phi_k^{\mathrm{T}})^{-1}]\phi_k\tilde{\theta}_k] \\
&= -[(1-\lambda)V_k + z_k^{\mathrm{T}}[I_p - \phi_k P_k\phi_k^{\mathrm{T}}(\lambda I_p + \phi_k P_k\phi_k^{\mathrm{T}})^{-1}]z_k] \\
&\leq 0.
\end{aligned}$$

Note that, since $(V_k)_{k=1}^\infty$ is a nonnegative, nonincreasing sequence, it converges to a nonnegative number. Hence, $\lim_{k\to\infty}(V_{k+1} - V_k) = 0$, which implies that $\lim_{k\to\infty}[(1-\lambda)V_k + z_k^{\mathrm{T}} R_k z_k] = 0$, where $R_k \stackrel{\triangle}{=} I_p - \phi_k P_k\phi_k^{\mathrm{T}}(\lambda I_p + \phi_k P_k\phi_k^{\mathrm{T}})^{-1}$. Lemma 2 from "Three Useful Lemmas" implies that R_k is positive definite. Since $V_k \geq 0$, it follows that $\lim_{k\to\infty} z_k = 0$. □

The following example shows that θ_k may converge despite the fact that $(\phi_k)_{k=0}^\infty$ is not persistent.

Example 4: Convergence of z_k and θ_k. Consider the first-order system

$$y_k = \frac{0.8}{\mathbf{q} - 0.4} u_k, \tag{29}$$

where $\mathbf{q}$ is the forward-shift operator. Define $\phi_k \stackrel{\triangle}{=} [y_{k-1}\ u_{k-1}]$, so that $y_k = \phi_k\theta$, where θ consists of the coefficients in (29). To apply RLS, let $P_0 = I_2$, $\theta_0 = 0$, and $\lambda = 0.999$. Figure 3 shows the shows the singular values of $F_{j,j+10}$, the predicted error z_k, and the parameter estimate θ_k for two choices of the input u_k. In the first case, for all $k \geq 0$, $u_k = 1$, whereas in the second case, for all $k \geq 0$, $u_k = 1$. For both choices of u_k, the predicted error z_k converges to zero, which confirms Proposition 6, and θ_k converges. Note that, in these two cases, θ_k converges to different parameter values, neither of which is the true value. $\diamond$

Table 4 summarizes the results in this section.

TABLE 4: Behavior of P_k with and without persistent excitation.

Excitation \ λ	$\lambda = 1$	$\lambda \in (0, 1)$
Persistent	• P_k converges to zero • Proposition 3 • Example 2	• P_k is bounded • Propositions 4, 5 • Example 2
Not Persistent	• All singular values of P_k are bounded • Some of these converge to zero • Example 3	• Some singular values of P_k diverge • The remaining singular values are bounded • Example 3

Persistent Excitation and the Condition Number

For nonsingular $A \in \mathbb{R}^{n\times n}$, the condition number of A is defined by

$$\kappa(A) \triangleq \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}, \tag{30}$$

For $B \in \mathbb{R}^{n\times m}$, let $\|B\|$ denotes the maximum singular value of B. If A is positive definite, then

$$\|A^{-1}\|^{-1} I_n = \sigma_{\min}(A) I_n \le A \le \sigma_{\max}(A) I_n = \|A\| I_n. \tag{31}$$

Therefore, if $\alpha, \beta \in (0, \infty)$ satisfy $\alpha \le \sigma_{\min}(A)$ and $\sigma_{\max}(A) \le \beta$, then $\kappa(A) \le \beta/\alpha$. Thus, if $\lambda = 1$ and $(\phi_k)_{k=0}^{\infty}$ is persistently exciting with N, α, β given by Definition 1, then (21) implies that

$$\kappa(P_k) \le \frac{\beta}{\alpha}. \tag{32}$$

Similarly, if $\lambda \in (0, 1)$ and $(\phi_k)_{k=0}^{\infty}$ is persistently exciting with N, α, β given by Definition 1, then (25) implies that

$$\kappa(P_k) \le \frac{\beta + (1 - \lambda^{N+1})\|P_N^{-1}\|}{\lambda^N (1-\lambda)\alpha}. \tag{33}$$

However, as shown by Example 3, in the case where $(\phi_k)_{k=0}^{\infty}$ is not persistently exciting, there might not exist $\alpha > 0$ satisfying (19), and thus $\kappa(P_k)$ cannot be bounded. Hence $\kappa(P_k)$ can be used to determine whether or not $(\phi_k)_{k=0}^{\infty}$ is persistently exciting, where a bounded condition number implies that $(\phi_k)_{k=0}^{\infty}$ is persistently exciting, and a diverging condition number implies that ϕ_k is not persistently exciting, as illustrated by the following example. [35] provides a recursive algorithm for computing $\kappa(P_k)$.

Example 5: Using the condition number of P_k to determine whether $(\phi_k)_{k=0}^{\infty}$ is persistently exciting. Consider the 5th-order system

$$y_k = \frac{0.68\mathbf{q}^4 - 0.16\mathbf{q}^3 - 0.12\mathbf{q}^2 - 0.18\mathbf{q} + 0.09}{\mathbf{q}^5 - \mathbf{q}^4 + 0.41\mathbf{q}^3 - 0.17\mathbf{q}^2 - 0.03\mathbf{q} + 0.01} u_k, \tag{34}$$

where u_k is given by (27). To apply RLS, let θ consist of the coefficients in (34) and let

$$\phi_k = [u_{k-1} \ \cdots \ u_{k-5} \ y_{k-1} \ \cdots \ y_{k-5}], \tag{35}$$

so that $y_k = \phi_k \theta$. Letting $P_0 = I_{10}$, Figure 4 shows the singular values of $F_{j,j+20}$ and the singular values and condition number of P_k for $\lambda = 1$ and $\lambda = 0.99$. In particular, the smallest singular value of $F_{j,j+20}$ is essentially zero, which indicates that $(\phi_k)_{k=0}^{\infty}$ is not persistently exciting. Consequently, in the case where $\lambda = 0.99$, P_k becomes ill-conditioned. $\diamond$

In Example 5, the regressor $(\phi_k)_{k=0}^{\infty}$ is not persistently exciting. Consequently, in the case where $\lambda = 1$, it follows from (20) that P_k is bounded by P_0, and thus all of the singular values of P_k are bounded; this property is illustrated by Figure 4. However, Figure 4 also shows that not all of the singular values of P_k converge to zero. On the other hand, in the case where $\lambda = 0.99$, Figure 4 shows that some of the singular values of P_k are bounded, whereas the remaining singular values diverge. This example thus shows that singular values can diverge due to the lack of persistent excitation with $\lambda \in (0, 1)$.

Lyapunov Analysis of the Parameter Error

Let $k \geq 0$, and consider the system

$$x_{k+1} = f(k, x_k), \tag{36}$$

where $x_k \in \mathbb{R}^n$, $f\colon \{0, 1, 2, \ldots\} \times \mathbb{R}^n \to \mathbb{R}^n$ is continuous, and, for all $k \geq 0$, $f(k, 0) = 0$. Let $\mathcal{D} \subset \mathbb{R}^n$ be an open set such that $0 \in \mathcal{D}$.

Definition 2: The zero solution of (36) is *Lyapunov stable* if, for all $\varepsilon > 0$ and $k_0 \geq 0$, there exists $\delta(\varepsilon, k_0) > 0$ such that, for all $x_{k_0} \in \mathbb{R}^n$ satisfying $\|x_{k_0}\| < \delta(\varepsilon, k_0)$, it follows that, for all $k \geq k_0$, $\|x_k\| < \varepsilon$.

Definition 3: The zero solution of (36) is *uniformly Lyapunov stable* if, for all $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$ such that, for all $k_0 \geq 0$ and all $x_{k_0} \in \mathbb{R}^n$ satisfying $\|x_{k_0}\| < \delta(\varepsilon)$, it follows that, for all $k \geq k_0$, $\|x_k\| < \varepsilon$.

Definition 4: The zero solution of (36) is *globally asymptotically stable* if it is Lyapunov stable and, for all $k_0 \geq 0$ and all $x_{k_0} \in \mathbb{R}^n$, it follows that $\lim_{k\to\infty} x_k = 0$.

Definition 5: The zero solution of (36) is *uniformly globally geometrically stable* if there exist $\alpha > 0$ and $\beta > 1$ such that, for all $k_0 \geq 0$ and all $x_{k_0} \in \mathbb{R}^n$, it follows that, for all $k \geq k_0$, $\|x_k\| \leq \alpha \|x_{k_0}\| \beta^{-k}$.

Note that, if the zero solution of (36) is uniformly globally geometrically stable, then it is uniformly globally aymptotically stable as well as uniformly Lyapunov stable.

The following three results are specializations of Theorem 13.11 given in [36, pp. 784, 785].

Theorem 3: Consider (36), and assume there exist a continuous function $V\colon \{0,1,\ldots\} \times \mathcal{D} \to \mathbb{R}$ and $\alpha_1 > 0$ such that, for all $k \geq 0$ and $x \in \mathcal{D}$,

$$V(k,0) = 0, \tag{37}$$

$$\alpha_1 \|x\|^2 \leq V(k,x), \tag{38}$$

$$V(k+1, f(k,x)) - V(k,x) \leq 0. \tag{39}$$

Then, the zero solution of (36) is Lyapunov stable.

Theorem 4: Consider (36), and assume there exist a continuous function $V\colon \{0,1,\ldots\} \times \mathcal{D} \to \mathbb{R}$ and $\alpha_1, \beta_1 > 0$ such that, for all $k \geq 0$ and $x \in \mathcal{D}$,

$$V(k,0) = 0, \tag{40}$$

$$\alpha_1 \|x\|^2 \leq V(k,x) \leq \beta_1 \|x\|^2, \tag{41}$$

$$V(k+1, f(k,x)) - V(k,x) \leq 0. \tag{42}$$

Then, the zero solution of (36) is uniformly Lyapunov stable.

Theorem 5: Consider (36), and assume there exist a continuous function $V\colon \{0,1,\ldots\} \times \mathbb{R}^n \to \mathbb{R}$, and $\alpha_1, \beta_1, \gamma_1 > 0$, such that, for all $k \geq 0$ and $x \in \mathbb{R}^n$,

$$\alpha_1 \|x\|^2 \leq V(k,x) \leq \beta_1 \|x\|^2, \tag{43}$$

$$V(k+1, f(k,x)) - V(k,x) \leq -\gamma_1 \|x\|^2. \tag{44}$$

Then, the zero solution of (36) is uniformly globally geometrically stable.

The following result uses Theorems 3-5 to prove that, if $(\phi_k)_{k=0}^\infty$ is persistently exciting, then the RLS estimate θ_k with $\lambda \in (0,1)$ converges to θ in the sense of Definition 5. A related result is given in [34].

Theorem 6: Assume that $(\phi_k)_{k=0}^{\infty}$ is persistently exciting, let N, α, β be given by Definition 1, let $R \in \mathbb{R}^{n\times n}$ be positive definite, define $P_0 \triangleq R^{-1}$, let $\lambda \in (0,1]$, and, for all $k \geq 0$, let P_k be given by (4). Then the zero solution of (12) is Lyapunov stable. In addition, if $\lambda \in (0,1)$, then the zero solution of (12) is uniformly Lyapunov stable and uniformly globally geometrically stable.

Proof: Define the Lyapunov candidate

$$V(k,x) \triangleq x^{\mathrm{T}} P_k^{-1} x,$$

where $x \in \mathbb{R}^n$. Note that, for all $k \geq 0$, $V(k,0) = 0$, which confirms (37). Next, defining

$$f(k,x) \triangleq (I_n - P_{k+1}\phi_k^{\mathrm{T}}\phi_k)x,$$

it follows that

$$\begin{aligned}
V(k+1, f(k,x)) - V(k,x) &= f(k,x)^{\mathrm{T}} P_{k+1}^{-1} f(k,x) - x^{\mathrm{T}} P_k^{-1} x \\
&= x^{\mathrm{T}}[(I_n - \phi_k^{\mathrm{T}}\phi_k P_{k+1}) P_{k+1}^{-1} (I_n - P_{k+1}\phi_k^{\mathrm{T}}\phi_k) - P_k^{-1}]x \\
&= x^{\mathrm{T}}[(P_{k+1}^{-1} - \phi_k^{\mathrm{T}}\phi_k)(I_n - P_{k+1}\phi_k^{\mathrm{T}}\phi_k) - P_k^{-1}]x \\
&= x^{\mathrm{T}}[P_{k+1}^{-1} - 2\phi_k^{\mathrm{T}}\phi_k + \phi_k^{\mathrm{T}}\phi_k P_{k+1}\phi_k^{\mathrm{T}}\phi_k - P_k^{-1}]x \\
&= x^{\mathrm{T}}[(\lambda - 1)P_k^{-1} - \phi_k^{\mathrm{T}}(I_p - \phi_k P_{k+1}\phi_k^{\mathrm{T}})\phi_k]x. \qquad (45)
\end{aligned}$$

First, consider the case where $\lambda = 1$. It follows from (8) with $\lambda = 1$ that $P_0^{-1} \leq P_k^{-1}$, and thus, for all $k \geq 0$,

$$\sigma_{\min}(P_0^{-1})\|x\|^2 \leq V(k,x),$$

which confirms (38) with $\alpha_1(\|x\|) = \sigma_{\min}(P_0^{-1})\|x\|^2$. Next, note that

$$I_p - \phi_k P_{k+1}\phi_k^{\mathrm{T}} = I_p - [\phi_k P_k \phi_k^{\mathrm{T}} - \phi_k P_k \phi_k^{\mathrm{T}} \left(I_p + \phi_k P_k \phi_k^{\mathrm{T}}\right)^{-1} \phi_k P_k \phi_k^{\mathrm{T}}]. \qquad (46)$$

Using (45), (46), and Lemma 3 from “Three Useful Lemmas” yields (39). It thus follows from Theorem 3 that the zero solution of (12) is Lyapunov stable.

Next, consider the case where $\lambda \in (0,1)$. It follows from Proposition 4 that, for all $k \geq N+1$,

$$\begin{aligned}
\frac{\lambda^N(1-\lambda)\alpha}{1-\lambda^{N+1}}\|x\|^2 \leq V(k,x) &\leq \frac{\beta}{1-\lambda^{N+1}}\|x\|^2 + x^{\mathrm{T}} P_N^{-1} x \\
&\leq \left(\frac{\beta}{1-\lambda^{N+1}} + \|P_N^{-1}\|\right)\|x\|^2,
\end{aligned}$$

which confirms (41) for all $\lambda \in (0,1)$ with $\alpha_1 = \dfrac{\lambda^N(1-\lambda)\alpha}{1-\lambda^{N+1}}$, and $\beta_1 = \dfrac{\beta}{1-\lambda^{N+1}} + \|P_N^{-1}\|$. Using (45), (46), and Lemma 3 from "Three Useful Lemmas", (42) is confirmed. It thus follows from Theorem 4 that the zero solution of (12) is uniformly Lyapunov stable.

Furthermore, (43) is confirmed, $\alpha_1 = \frac{\lambda^N(1-\lambda)\alpha}{1-\lambda^{N+1}}$, and $\beta_1 = \frac{\beta}{1-\lambda^{N+1}} + \|P_N^{-1}\|$. Finally, if $\lambda \in (0,1)$, then

$$
\begin{aligned}
V(k+1, f(k,x)) - V(k,x) &\leq (\lambda-1)x^{\mathrm{T}}P_k^{-1}x \\
&\leq (\lambda-1)\left(\frac{\beta}{1-\lambda^{N+1}} + \|P_N^{-1}\|\right)\|x\|^2,
\end{aligned}
$$

which confirms (44) with , $\gamma_1 = (1-\lambda)(\frac{\beta}{1-\lambda^{N+1}} + \|P_N^{-1}\|)$. It thus follows from Theorem 5 that the zero solution of (12) is uniformly globally geometrically stable. □

The following result provides an alternative proof of Theorem 6 that does not depend on Theorems 3-5. In addition, this result considers the case $\lambda = 1$, where the RLS estimate θ_k converges to θ in the sense of Definition 4.

Theorem 7: Assume that $(\phi_k)_{k=0}^{\infty}$ is persistently exciting, let N, α, β be given by Definition 1, let $R \in \mathbb{R}^{n\times n}$ be positive definite, define $P_0 \triangleq R^{-1}$, let $\lambda \in (0,1]$, and, for all $k \geq 0$, let P_k be given by (4). Then the zero solution of (12) is globally asymptotically stable. Furthermore, if $\lambda \in (0,1)$, then the zero solution of (12) is uniformly globally geometrically stable.

Proof: Let $k_0 \geq 0$ and $\tilde{\theta}_{k_0} \in \mathbb{R}^n$. Then, it follows from (14) that, for all $k \geq k_0$,

$$
\begin{aligned}
\|\tilde{\theta}_k\| &= \lambda^{k-k_0}\|P_kP_{k_0}^{-1}\tilde{\theta}_{k_0}\| \\
&\leq \|P_kP_{k_0}^{-1}\tilde{\theta}_{k_0}\| \\
&\leq \|P_k\|\|P_{k_0}^{-1}\|\|\tilde{\theta}_{k_0}\|. \qquad (47)
\end{aligned}
$$

First, consider the case where $\lambda = 1$. Let $\delta > 0$, and suppose that $\tilde{\theta}_{k_0} \in \mathbb{R}^n$ satisfies $\|\tilde{\theta}_{k_0}\| < \delta$. It follows from (8) with $\lambda = 1$ that $\|P_k\| \leq \|P_0\|$ and (47), that, for all $k \geq k_0$, $\|\tilde{\theta}_k\| < \|P_0\|\|P_{k_0}^{-1}\|\delta$. It thus follows from Definition 2 with $\varepsilon = \|P_0\|\|P_{k_0}^{-1}\|\delta$ that the zero solution of (12) is Lyapunov stable.

Next, let $\tilde{\theta}_0 \in \mathbb{R}^n$. Then, Proposition 3 implies that

$$
\lim_{k\to\infty}\tilde{\theta}_k = \lim_{k\to\infty}P_kP_0^{-1}\tilde{\theta}_0 = 0.
$$

It thus follows from Definition 4 that the zero solution of (12) is globally asymptotically stable.

Next, consider the case where $\lambda \in (0,1)$. Let $k_0 \geq 0$ and $\delta > 0$, and let $\tilde{\theta}_{k_0} \in \mathbb{R}^n$ satisfy $\|\tilde{\theta}_{k_0}\| < \delta$. It follows from Proposition 4 and (47) that, for all $k \geq \max(N+1, k_0)$,

$$\|\tilde{\theta}_k\| < \varepsilon,$$

where

$$\varepsilon \triangleq \frac{\beta + (1-\lambda^{N+1})\|P_N^{-1}\|}{\lambda^N(1-\lambda)\alpha}\delta.$$

It thus follows from Definition 3 that the zero solution of (12) is uniformly Lyapunov stable.

Next, let $\tilde{\theta}_{k_0} \in \mathbb{R}^n$. Then, it follows from (14) and Proposition 4 that, for all $\tilde{\theta}_{k_0} \in \mathbb{R}^n$ and $k \geq N+1$,

$$\|\tilde{\theta}_k\| \leq \alpha_0 \|\tilde{\theta}_{k_0}\| \beta_0^{-k},$$

where $\beta_0 \triangleq 1/\lambda$ and

$$\alpha_0 \triangleq \frac{\beta + (1-\lambda^{N+1})\|P_N^{-1}\|}{\lambda^N(1-\lambda)\alpha}.$$

It thus follows from Definition 5 that the zero solution of (12) is uniformly globally geometrically stable, and thus globally asymptotically stable. □

The following result shows that persistent excitation produces an infinite sequence of matrices whose product converges to zero.

Proposition 7: Let $P_0 \in \mathbb{R}^{n\times n}$ be positive definite, let $\lambda \in (0,1]$, and, for all $k \geq 0$, let P_k be given by (4). Then, for all $k \geq 0$, all of the eigenvalues of $P_{k+1}\phi_k^{\mathrm{T}}\phi_k$ are contained in $[0,1]$. If, in addition, $(\phi_k)_{k=0}^{\infty}$ is persistently exciting, then

$$\lim_{k\to\infty} \mathcal{A}_k = 0, \tag{48}$$

where

$$\mathcal{A}_k \triangleq (I_n - P_{k+1}\phi_k^{\mathrm{T}}\phi_k)\cdots(I_n - P_1\phi_0^{\mathrm{T}}\phi_0). \tag{49}$$

Proof: It follows from (8) that, for all $k \geq 0$, $\phi_k^{\mathrm{T}}\phi_k \leq P_{k+1}^{-1}$, and thus, for all $k \geq 0$, $P_{k+1}^{1/2}\phi_k^{\mathrm{T}}\phi_k P_{k+1}^{1/2} \leq I_n$. Hence, for all $k \geq 0$,

$$0 \leq \lambda_{\max}(P_{k+1}\phi_k^{\mathrm{T}}\phi_k) = \lambda_{\max}(P_{k+1}^{1/2}\phi_k^{\mathrm{T}}\phi_k P_{k+1}^{1/2}) \leq 1.$$

To prove (48), suppose that $(\phi_k)_{k=0}^{\infty}$ is persistently exciting, let $i \in \{1,\ldots,n\}$, and define $\theta_0 \triangleq e_i + \theta$, where e_i is the ith column of I_n. Note that $\tilde{\theta}_0 \triangleq \theta_0 - \theta = e_i$. Then, (14) implies that, for all $k \geq 0$,

$$\tilde{\theta}_{k+1} = \mathcal{A}_k e_i = \lambda^{k+1} P_{k+1} P_0^{-1} e_i. \tag{50}$$

It follows from Theorem 7 that $\tilde{\theta}_k$ converges to zero. Hence, (50) implies that the ith column of $\mathcal{A}_k$ converges to zero as $k \to \infty$. It thus follows that every column of $\mathcal{A}_k$ converges to zero as $k \to \infty$, which implies (48). $\square$

It follows from Theorem 7 that, if $(\phi_k)_{k=0}^{\infty}$ is persistently exciting, then, for all $\lambda \in (0, 1]$, $\tilde{\theta}_k$ converges to zero. In addition, if $\lambda \in (0, 1)$, then $\tilde{\theta}_k$ converges to zero geometrically, and thus the rate of convergence of $\|\tilde{\theta}_k\|$ is $O(\lambda^k)$. However, in the case $\lambda = 1$, as shown in [34] and the next example, $\tilde{\theta}_k$ converges to zero as $O(1/k)$, and thus the convergence is not geometric.

Example 6: Effect of λ on the rate of convergence of θ_k. Consider the 3rd-order FIR system

$$y_k = \frac{\mathbf{q}^2 + 0.8\mathbf{q} + 0.5}{\mathbf{q}^3} u_k. \tag{51}$$

To apply RLS, let $\theta = [1\ 0.8\ 0.5]$, $\theta_0 = 0$, and $\phi_k = [u_{k-1}\ u_{k-2}\ u_{k-3}]$, where the input u_k is zero-mean Gaussian white noise with standard deviation 1. Note that $(\phi_k)_{k=0}^{\infty}$ is persistently exciting. It thus follows from Theorem 7 that $\tilde{\theta}_k$ converges to zero. Figure 5 shows the parameter-error norm $\|\tilde{\theta}_k\|$ for several values of P_0 and λ as well as the condition number of the corresponding P_k. Note that the convergence rate of $\|\tilde{\theta}_k\|$ is $O(1/k)$ for $\lambda = 1$ and geometric for all $\lambda \in (0, 1)$. Furthermore, as λ is decreased, the convergence rate of θ_k increases; however, the condition number of P_k degrades, and the effect of P_0 is reduced. $\diamond$

Lack of Persistent Excitation

This section presents numerical examples to investigate the effect of lack of persistent excitation. As shown in Example 3 and Example 5, if $(\phi_k)_{k=0}^{\infty}$ is not persistently exciting and $\lambda = 1$, then some of the singular values of P_k converge to zero, whereas the remaining singular values remain bounded. On the other hand, if $(\phi_k)_{k=0}^{\infty}$ is not persistently exciting and $\lambda \in (0, 1)$, then some of the singular values of P_k remain bounded, whereas the remaining singular values diverge. Furthermore, Proposition 6 implies that the predicted error z_k converges to zero whether or not $(\phi_k)_{k=0}^{\infty}$ is persistent.

Example 7: Lack of persistent excitation in scalar estimation. Let $n = 1$, so that (4), (5) are given by

$$P_{k+1} = \frac{P_k}{\lambda + P_k \phi_k^2}, \tag{52}$$

$$\tilde{\theta}_{k+1} = \frac{\lambda \tilde{\theta}_k}{\lambda + P_k \phi_k^2}. \tag{53}$$

Now, let $k_0 \geq 0$ and assume that, for all $k \geq k_0$, $\phi_k = 0$. Therefore, for all $j \geq 0$ and $N \geq 1$, $F_{j,j+N}$ cannot be lower bounded as in (19), and thus $(\phi_k)_{k=0}^{\infty}$ is not persistently exciting.

Furthermore, in the case where $\lambda = 1$, it follows from the fact that $\phi_k = 0$ for all $k \geq k_0$ that P_k and $\tilde{\theta}_k$ converge in k_0 steps to $\overline{P} \neq 0$ and $\overline{\tilde{\theta}}$, respectively. Furthermore, if $\theta_0 \neq \theta$, then $\overline{\tilde{\theta}} \neq 0$. However, in the case where $\lambda \in (0, 1)$, it follows that P_k diverges geometrically, whereas, as in the case where $\lambda = 1$, $\tilde{\theta}_k$ converges in k_0 steps. Therefore, for all $\lambda \in (0, 1]$, since $\phi_k = 0$ for all $k \geq k_0$, it follows from (52) and (53) that, for all $k \geq k_0$, the minimum value of (2) is achieved in a finite number of steps. Consequently, RLS provides no further refinement of the estimate θ_k of θ, and thus $\overline{\tilde{\theta}} \neq 0$ implies that θ_k does not converge to θ.

Alternatively, assume that, for all $k \geq 0$, $\phi_k = \overline{\phi}$, where $\overline{\phi} \neq 0$. Then it follows from Definition 1 with $N = 1$, $\alpha = \overline{\phi}^2$, and $\beta = 3\overline{\phi}^2$ that $(\phi_k)_{k=0}^{\infty}$ is persistently exciting. If $\lambda = 1$, then both P_k and $\tilde{\theta}_k$ converge to zero. However, if $\lambda \in (0, 1)$, then P_k converges to $\frac{1-\lambda}{\overline{\phi}^2}$ and $\tilde{\theta}_k$ converges geometrically to zero. Table 5 shows the asymptotic behavior of $\tilde{\theta}_k$ and P_k for both of these cases. ◇

Excitation \ λ	$\lambda = 1$	$\lambda \in (0, 1)$
Not persistently exciting	$\tilde{\theta}_k \to \overline{\tilde{\theta}}, P_k \to \overline{P}$	$\tilde{\theta}_k \to \overline{\tilde{\theta}}$, P_k diverges
Persistently exciting	$\tilde{\theta}_k \to 0, P_k \to 0$	$\tilde{\theta}_k \to 0, P_k \to \frac{1-\lambda}{\overline{\phi}^2}$

TABLE 5: Asymptotic behavior of RLS in Example 7. In the case of persistent excitation with $\lambda < 1$, the convergence of $\tilde{\theta}_k$ is geometric.

Example 8: Subspace-constrained regressor. Consider (1), where $\phi_k = (\sin \frac{2\pi k}{100})[1\ 1]$ and $\theta = [0.4\ 1.4]^{\mathrm{T}}$. To estimate θ using RLS, let $P_0 = I_2$ and $\theta_0 = 0$. Figure 6 shows the estimate θ_k of θ with $\lambda = 1$ and $\lambda = 0.99$. Note that all regressors ϕ_k lie along the same one-dimensional subspace, and thus, $(\phi_k)_{k=0}^{\infty}$ is not persistently exciting. It follows from (16) that the estimate θ_k of θ lies in this subspace.

For $\lambda = 1$, note that one singular value decreases to zero, whereas the other singular value is bounded. Note that $\tilde{\theta}_k$ converges along the singular vector corresponding to the bounded singular value. For $\lambda = 0.99$, one singular value is bounded, whereas the other singular value diverges. Note that $\tilde{\theta}_k$ converges along the singular vector corresponding to the diverging singular value. ◇

Example 9: Lack of persistent excitation and finite-precision arithmetic. Consider the problem of fitting a 5th-order model to measured input-output data from the system (34), where the input u_k is given by (27). Note that ϕ_k is given by (35), and is not persistently exciting as shown in Example 5. Let $P_0 = I_{10}$, $\theta_0 = 0$, and $\lambda = 0.999$. Figure 7 shows the predicted error

z_k, the norm of the parameter error $\tilde{\theta}_k$, and the singular values and the condition number of P_k. Note that the $\tilde{\theta}_k$ does not converge to zero and that six singular values of P_k remain bounded due to the presence of three harmonics in the regresssor. Due to finite-precision arithmetic, the computation becomes erroneous as P_k becomes numerically ill-conditioned, and thus the estimate θ_k diverges. $\diamond$

The numerical examples in this section show that, if $\lambda \in (0,1]$ and $(\phi_k)_{k=0}^{\infty}$ is not persistently exciting, then $\tilde{\theta}_k$ does not necessarily converge to zero. Furthermore, if $\lambda \in (0,1)$ and $(\phi_k)_{k=0}^{\infty}$ is not persistently exciting, then some of the singular values of P_k diverge, and θ_k diverges due to finite-precision arithmetic when P_k becomes numerically ill-conditioned.

Information Subspace

Using the singular value decomposition, (8) can be written as

$$P_{k+1}^{-1} = \lambda U_k \Sigma_k U_k^{\mathrm{T}} + U_k \psi_k^{\mathrm{T}} \psi_k U_k^{\mathrm{T}}, \tag{54}$$

where $U_k \in \mathbb{R}^{n \times n}$ is an orthonormal matrix whose columns are the singular vectors of P_k^{-1}, $\Sigma_k \in \mathbb{R}^{n \times n}$ is a diagonal matrix whose diagonal entries are the corresponding singular values, and

$$\psi_k \triangleq \phi_k U_k. \tag{55}$$

The columns of U_k are the *information directions* at step k, and each row of ψ_k is the projection of the corresponding row of ϕ_k onto the information directions. The norm of each column of ψ_k thus indicates the *information content* present in ϕ_k along the corresponding information direction. The smallest subspace that is spanned by a subset of the information directions and that contains all rows of ϕ_k is the *information-rich subspace* $\mathcal{I}_k$ at step k. Figure 8 illustrates the information-rich subspace.

Now, consider the case where

$$\psi_k = \begin{bmatrix} \psi_{k,1} & 0_{p \times (n-n_1)} \end{bmatrix}, \tag{56}$$

where $\psi_{k,1} \in \mathbb{R}^{p \times n_1}$. It follows from (56) that ϕ_k provides new information along the first n_1 columns of U_k; these directions constitute the information-rich subspace. It thus follows from (54) and (56) that P_{k+1}^{-1} is given by

$$P_{k+1}^{-1} = U_k \begin{bmatrix} \lambda \Sigma_{k,1} + \psi_{k,1}^{\mathrm{T}} \psi_{k,1} & 0 \\ 0 & \lambda \Sigma_{k,2} \end{bmatrix} U_k^{\mathrm{T}}, \tag{57}$$

where $\Sigma_{k,1} \in \mathbb{R}^{n_1 \times n_1}$ is the diagonal matrix whose diagonal entries are the first n_1 singular values of P_k^{-1}, and $\Sigma_{k,2}$ is the diagonal matrix whose diagonal entries are the remaining $n - n_1$ singular values of P_k^{-1}. In particular, writing

$$U_k = \begin{bmatrix} U_{k,1} & U_{k,2} \end{bmatrix}, \tag{58}$$

where $U_{k,1} \in \mathbb{R}^{n \times n_1}$ contains the first n_1 columns of U_k, and $U_{k,2} \in \mathbb{R}^{n \times n - n_1}$ contains the remaining $n - n_1$ columns of U_k, it follows that

$$P_{k+1}^{-1} = \begin{bmatrix} U_{k+1,1} & U_{k+1,2} \end{bmatrix} \begin{bmatrix} \Sigma_{k+1,1} & 0 \\ 0 & \Sigma_{k+1,2} \end{bmatrix} \begin{bmatrix} U_{k+1,1}^{\mathrm{T}} \\ U_{k+1,2}^{\mathrm{T}} \end{bmatrix}, \tag{59}$$

where

$$U_{k+1,1} = U_{k,1} V_k, \tag{60}$$

$$\Sigma_{k+1,1} = D_k, \tag{61}$$

$$U_{k+1,2} = U_{k,2}, \tag{62}$$

$$\Sigma_{k+1,2} = \lambda \Sigma_{k,2}, \tag{63}$$

where $V_k \in \mathbb{R}^{n_1 \times n_1}$ contains the singular vectors of $\lambda \Sigma_{k,1} + \psi_{k,1}^{\mathrm{T}} \psi_{k,1}$ and $D_k \in \mathbb{R}^{n_1 \times n_1}$ is the diagonal matrix containing the corresponding singular values. It follows from (62), (63) that if, for all $k \geq 0$, ψ_k is given by (56) and $\lambda \in (0, 1)$, then the last $n - n_1$ singular vectors of P_k^{-1} do not change and the corresponding singular values of P_k^{-1} decrease to zero geometrically. It thus follows from Proposition 4 that $(\phi_k)_{k=0}^{\infty}$ is not persistently exciting. Furthermore, since P_k and P_k^{-1} have the same singular vectors and the singular values of P_k are the reciprocals of the singular values of P_k^{-1}, it follows that the last $n - n_1$ singular values of P_k diverge.

The next example considers the case where there exists a proper subspace $\mathcal{S} \subset \mathbb{R}^n$ such that, for all $k \geq 0$, $\mathcal{R}(\phi_k^{\mathrm{T}}) \subseteq \mathcal{S}$. Hence, $(\phi_k)_{k=0}^{\infty}$ is not persistently exciting. In this case, for all $k \geq 0$, the information-rich subspace $\mathcal{I}_k$ is a proper subspace of $\mathbb{R}^n$, and the singular values of P_k^{-1} corresponding to the singular vectors in the orthogonal complement of $\mathcal{I}_k$ converge to zero.

Example 10: Lack of persistent excitation and the information-rich subspace. Consider the regressor ϕ_k given by (35) used in Example 5. Recall that $(\phi_k)_{k=0}^{\infty}$ is not persistently exciting. Let $P_0 = I_{10}$. Figure 9 shows the information content $|\psi_{k,(i)}|$ for several values of λ along with the singular values of the corresponding P_k^{-1}. Note that the information-rich subspace is six dimensional due to the presence of three harmonics in u_k as shown by six relatively large components of ψ_k and, in the case where $\lambda < 1$, the singular values that correspond to the singular vectors not in the information-rich subspace converge to zero in machine precision. ◇

Variable-Direction forgetting

Examples 3, 5, 7, 8, and 9 show that some of the singular values of P_k^{-1} converge to zero in the case where ϕ_k is not persistently exciting. To address this situation, (8) is modified by replacing the scalar forgetting factor λ by a data-dependent forgetting matrix Λ_k. Similar modifications are discussed in "Toward Matrix Forgetting". In particular, P_{k+1}^{-1} is redefined as

$$P_{k+1}^{-1} = \Lambda_k P_k^{-1} \Lambda_k + \phi_k^{\mathrm{T}} \phi_k, \tag{64}$$

where Λ_k is a positive-definite (and thus symmetric) matrix constructed below. Note that, for all $k \geq 0$, P_{k+1}^{-1} given by (64) is positive definite. Using the singular value decomposition, (64) can be written as

$$P_{k+1}^{-1} = \Lambda_k U_k \Sigma_k U_k^{\mathrm{T}} \Lambda_k + U_k \psi_k^{\mathrm{T}} \psi_k U_k^{\mathrm{T}}, \tag{65}$$

2 where U_k, Σ_k, and ψ_k are as defined in the previous section.

The objective is to apply forgetting to only those singular values of P_k^{-1} that correspond to the singular vectors in the information-rich subspace, that is, forgetting is restricted to the subspace of P_k^{-1} where sufficient new information is provided by ϕ_k. Specifically, forgetting is applied to those information directions where the information content is greater than $\varepsilon > 0$, where ε should be selected to be larger than the noise to signal ratio or larger than the machine zero, if no noise is present. To do so, (65) is written as

$$P_{k+1}^{-1} = U_k \overline{\Lambda}_k \Sigma_k \overline{\Lambda}_k U_k^{\mathrm{T}} + U_k \psi_k^{\mathrm{T}} \psi_k U_k^{\mathrm{T}}, \tag{66}$$

where $\overline{\Lambda}_k$ is a diagonal matrix whose diagonal entries are either $\sqrt{\lambda}$ or 1. In particular,

$$\overline{\Lambda}_k(i,i) \triangleq \begin{cases} \sqrt{\lambda}, & \|\mathrm{col}_i(\psi_k)\| > \varepsilon, \\ 1, & \text{otherwise}, \end{cases} \tag{67}$$

where $\mathrm{col}_i(\psi_k)$ is the ith column of ψ_k and $\lambda \in (0,1]$. Note that, it follows from (66) and (67) that P_{k+1}^{-1} is positive definite. Next, it follows from (65) and (66) that

$$\Lambda_k = U_k \overline{\Lambda}_k U_k^{\mathrm{T}}, \tag{68}$$

which is positive definite. Note that

$$\Lambda_k^{-1} = U_k \overline{\Lambda}_k^{-1} U_k^{\mathrm{T}}. \tag{69}$$

The next result provides a recursive formula to update P_{k+1} given by (64).

Proposition 8: Let $\lambda \in (0,1]$, $\varepsilon > 0$, let $(P_k)_{k=0}^{\infty}$ be a sequence of $n \times n$ positive-definite matrices, and let $U_k \in \mathbb{R}^{n \times n}$ be an orthonormal matrix whose columns are the singular vectors of P_k. Furthermore, let $\psi_k \in \mathbb{R}^{p\times n}$ be given by (55), let $\overline{\Lambda}_k$ be given by (67), and let Λ_k be given by (68). Then, for all $k \geq 0$, $(P_k)_{k=0}^{\infty}$ satisfies (64) if and only if, for all $k \geq 0$, $(P_k)_{k=0}^{\infty}$ satisfies

$$P_{k+1} = \overline{P}_k - \overline{P}_k \phi_k (I_p + \phi_k^{\mathrm{T}} \overline{P}_k \phi_k)^{-1} \phi_k^{\mathrm{T}} \overline{P}_k, \tag{70}$$

where

$$\overline{P}_k = \Lambda_k^{-1} P_k \Lambda_k^{-1}. \tag{71}$$

Proof: To prove necessity, it follows from (64) and matrix-inversion lemma, that

$$\begin{aligned}
P_{k+1} &= (\Lambda_k P_k^{-1} \Lambda_k + \phi_k^{\mathrm{T}} \phi_k)^{-1} \\
&= (\Lambda_k P_k^{-1} \Lambda_k)^{-1} - (\Lambda_k P_k^{-1} \Lambda_k)^{-1} \phi_k^{\mathrm{T}} [I_p + \phi_k (\Lambda_k P_k^{-1} \Lambda_k)^{-1} \phi_k^{\mathrm{T}}]^{-1} \phi_k (\Lambda_k P_k^{-1} \Lambda_k)^{-1} \\
&= \overline{P}_k - \overline{P}_k \phi_k (I_p + \phi_k^{\mathrm{T}} \overline{P}_k \phi_k)^{-1} \phi_k^{\mathrm{T}} \overline{P}_k,
\end{aligned}$$

where $\overline{P}_k$ is given by (71). Reversing these steps proves sufficiency. □

The modified update (64) is shown to be optimal for a specific cost function in "A Modified Quadratic Cost Function Supporting Variable-Direction RLS".

Next, the matrix-forgetting scheme (64) is shown to prevent the singular values of P_k from diverging. Consider the case where, for all $k \geq 0$,

$$\psi_k = \begin{bmatrix} \psi_{k,1} & 0 \end{bmatrix}, \tag{72}$$

where $\psi_{k,1} \in \mathbb{R}^{p\times n_1}$, that is, the information-rich subspace is spanned by the first n_1 columns of U_k. It thus follows from (66) and (72) that P_{k+1}^{-1} is given by

$$P_{k+1}^{-1} = U_k \begin{bmatrix} \lambda \Sigma_{k,1} + \psi_{k,1}^{\mathrm{T}} \psi_{k,1} & 0 \\ 0 & \Sigma_{k,2} \end{bmatrix} U_k^{\mathrm{T}}. \tag{73}$$

It follows from the $(2,2)$ block of (73) that the last $n - n_1$ information directions and the corresponding singular values are not affected by ϕ_k. Furthermore, if $n_1 = n$, that is, new information is present in ϕ_k along every information direction, then forgetting is applied to all of the singular values of P_k^{-1}, and thus variable-direction forgetting specializes to uniform-direction forgetting, that is, RLS with the update for P_k given by (8).

The next result shows that, as in the case of uniform-direction forgetting, z_k converges to zero with variable-direction forgetting for every choice of $\varepsilon > 0$, whether or not $(\phi_k)_{k=0}^{\infty}$ is persistently exciting.

Proposition 9: For all $k \geq 0$, let $\phi_k \in \mathbb{R}^{p\times n}$ and $y_k \in \mathbb{R}^p$, let $R \in \mathbb{R}^{n\times n}$ be positive definite, and let $P_0 = R^{-1}$, $\theta_0 \in \mathbb{R}^n$, and $\lambda \in (0,1]$. Furthermore, for all $k \geq 0$, let P_k and θ_k be given by (64) and (5), respectively. Then,

$$\lim_{k\to\infty} z_k = 0. \tag{74}$$

Proof: Using (67), (68), and $P_k^{-1} = U_k \Sigma_k U_k^{\mathrm{T}}$, it follows that, for all $k \geq 0$,

$$\Lambda_k P_k^{-1} \Lambda_k = U_k \overline{\Lambda}_k \Sigma_k \overline{\Lambda}_k U_k^{\mathrm{T}} \leq U_k \Sigma_k U_k^{\mathrm{T}} = P_k^{-1}. \tag{75}$$

For all $k \geq 0$, note that $z_k = \phi_k \tilde{\theta}_k$, and define $V_k \triangleq \tilde{\theta}_k^{\mathrm{T}} P_k^{-1} \tilde{\theta}_k$. Note that, for all $k \geq 0$ and $\tilde{\theta}_k \in \mathbb{R}^n$, $V_k \geq 0$. Furthermore, for all $k \geq 0$,

$$\begin{aligned}
V_{k+1} - V_k &= \tilde{\theta}_{k+1}^{\mathrm{T}} P_{k+1}^{-1} \tilde{\theta}_{k+1} - \tilde{\theta}_k^{\mathrm{T}} P_k^{-1} \tilde{\theta}_k \\
&= \tilde{\theta}_k^{\mathrm{T}} \Lambda_k P_k^{-1} \Lambda_k P_{k+1} \Lambda_k P_k^{-1} \Lambda_k \tilde{\theta}_k - \tilde{\theta}_k^{\mathrm{T}} P_k^{-1} \tilde{\theta}_k \\
&= \tilde{\theta}_k^{\mathrm{T}} [\Lambda_k P_k^{-1} \Lambda_k P_{k+1} \Lambda_k P_k^{-1} \Lambda_k - P_k^{-1}] \tilde{\theta}_k \\
&= \tilde{\theta}_k^{\mathrm{T}} [\Lambda_k P_k^{-1} (P_k - P_k \Lambda_k^{-1} \phi_k (I_p + \phi_k^{\mathrm{T}} \overline{P}_k \phi_k)^{-1} \phi_k^{\mathrm{T}} \Lambda_k^{-1} P_k) P_k^{-1} \Lambda_k - P_k^{-1}] \tilde{\theta}_k \\
&= \tilde{\theta}_k^{\mathrm{T}} [\Lambda_k P_k^{-1} \Lambda_k - \phi_k (I_p + \phi_k^{\mathrm{T}} \overline{P}_k \phi_k)^{-1} \phi_k^{\mathrm{T}} - P_k^{-1}] \tilde{\theta}_k \\
&= -[\tilde{\theta}_k^{\mathrm{T}} (P_k^{-1} - \Lambda_k P_k^{-1} \Lambda_k) \tilde{\theta}_k + z_k (I_p + \phi_k^{\mathrm{T}} \overline{P}_k \phi_k)^{-1} z_k] \\
&\leq 0.
\end{aligned}$$

Note that, since $(V_k)_{k=1}^{\infty}$ is a nonnegative, nonincreasing sequence, it converges to a nonnegative number. Hence, $\lim_{k\to\infty}(V_{k+1} - V_k) = 0$, which implies that

$$\lim_{k\to\infty} [\tilde{\theta}_k^{\mathrm{T}} (P_k^{-1} - \Lambda_k P_k^{-1} \Lambda_k) \tilde{\theta}_k + z_k (I_p + \phi_k^{\mathrm{T}} \overline{P}_k \phi_k)^{-1} z_k] = 0.$$

Since, for all $k \geq 0$, $P_k^{-1} - \Lambda_k P_k^{-1} \Lambda_k \geq 0$ and $(I_p + \phi_k^{\mathrm{T}} \overline{P}_k \phi_k)^{-1} > 0$, it follows that $\lim_{k\to\infty} z_k = 0$. □

The next result shows that P_k is bounded from above with variable-direction forgetting for every choice of $\varepsilon > 0$ in the case where $(\phi_k)_{k=0}^{\infty}$ is persistently exciting.

Proposition 10: Assume that $(\phi_k)_{k=0}^{\infty}$ is persistently exciting, let N, α, β be given by Definition 1, let $R \in \mathbb{R}^{n\times n}$ be positive definite, define $P_0 \triangleq R^{-1}$, let $\lambda \in (0,1)$, and, for all $k \geq 0$, let P_k be given by (64). Then, for all $k \geq N+1$,

$$\frac{\lambda^N (1-\lambda)\alpha}{1-\lambda^{N+1}} I_n \leq P_k^{-1}. \tag{76}$$

Proof: It follows from (64), that, for all $k \geq 0$, $\Lambda_k P_k^{-1} \Lambda_k \leq P_{k+1}^{-1}$ and $\phi_k^{\mathrm{T}} \phi_k \leq P_{k+1}^{-1}$. Next, using (68) and $P_k^{-1} = U_k \Sigma_k U_k^{\mathrm{T}}$, it follows that, for all $k \geq 0$,

$$\lambda P_k^{-1} = \lambda U_k \Sigma_k U_k^{\mathrm{T}} \leq U_k \overline{\Lambda}_k \Sigma_k \overline{\Lambda}_k U_k^{\mathrm{T}} = \Lambda_k P_k^{-1} \Lambda_k \leq P_{k+1}^{-1}.$$

Finally, for all $k \geq N+1$,

$$\begin{aligned}\alpha I_n &\leq \sum_{i=k-N-1}^{k-1} \phi_i^{\mathrm{T}} \phi_i \\ &\leq \sum_{i=k-N}^{k} P_i^{-1} \\ &\leq (\lambda^{-N} + \cdots + 1) P_k^{-1} \\ &= \frac{1-\lambda^{N+1}}{\lambda^N(1-\lambda)} P_k^{-1},\end{aligned}$$

which proves (76). □

The next two examples consider variable-direction forgetting in the case where $(\phi_k)_{k=0}^{\infty}$ is not persistently exciting. In these examples, P_k is bounded, z_k converges to zero, and θ_k converges, although not to the true value θ.

Example 11: Variable-direction forgetting for a regressor lacking persistent excitation. Reconsider Example 10. Let $P_0 = I_{10}$, and P_k^{-1} be given by (70), where $\varepsilon = 10^{-8}$. Figure 10 shows the information content $|\mathrm{col}_i(\psi_k)|$ and the singular values of the P_k^{-1} for several values of λ. Note that the information-rich subspace is six dimensional due to the presence of three harmonics in u_k as shown by six relatively large components of ψ_k and the singular values that correspond to the singular vectors not in the information-rich subspace do not converge to zero. ◇

Example 12: Effect of variable-direction forgetting on θ_k. Reconsider Example 9. Let $P_0 = I_{10}$, and P_k^{-1} be given by (70), where $\varepsilon = 10^{-8}$. Figure 11 shows the predicted error z_k, the norm of the parameter error $\tilde{\theta}_k$, and the singular values and the condition number of P_k. Note that the $\tilde{\theta}_k$ does not converge to zero and, unlike uniform-direction forgetting, all of the singular values of P_k remain bounded and θ_k is bounded. ◇

Concluding Remarks

This tutorial article presented a self-contained exposition of uniform-direction and variable-direction forgetting within the context of RLS. It was shown that, in the case of persistent excitation without forgetting, the parameter estimates converge asymptotically, whereas, with forgetting, the parameter estimates converge geometrically. Numerical examples were presented to illustrate this behavior.

In the case where forgetting is used but the excitation is not persistent, it was shown that forgetting is enforced in all information directions, whether or not new information is present along these directions. Consequently, the parameter estimates converge, but not necessarily to their true values; furthermore, the matrix P_k diverges, leading to numerical instability. This phenomenon was traced to the divergence of the singular values of P_k corresponding to singular vectors that are orthogonal to the information-rich subspace.

In order to address this problem, a data-dependent forgetting matrix was constructed to restrict forgetting to the information-rich subspace. The RLS cost function that corresponds to this extension of RLS was presented. Numerical examples showed that this variable-direction forgetting technique prevents P_k from diverging under lack of persistent excitation.

Since RLS is fundamentally least squares optimization, its estimates are not consistent in the case of sensor noise [37]. An open problem is thus to develop extensions of RLS that provide consistent parameter estimates in the presence of errors-in-variable noise arising in system identification problems [38].

Acknowledgments

This research was partially supported by AFOSR under DDDAS grant FA9550-16-1-0071 (Dynamic Data-Driven Applications Systems http://www.1dddas.org/).

References

[1] M. Grewal and K. Glover, “Identifiability of linear and nonlinear dynamical systems,” *IEEE Trans. Autom. Contr.*, vol. 21, no. 6, pp. 833–837, 1976.

[2] I. Y. Mareels and M. Gevers, “Persistence of excitation criteria,” in *Proc. Conf. Dec. Contr*, 1986, pp. 1933–1935.

[3] I. M. Mareels, R. R. Bitmead, M. Gevers, C. R. Johnson, and R. L. Kosut, “How exciting can a signal really be?” *Sys. Contr. Lett.*, vol. 8, no. 3, pp. 197–204, 1987.

[4] I. M. Mareels and M. Gevers, “Persistency of excitation criteria for linear, multivariable, time-varying systems,” *Mathematics of Control, Signals, and Systems*, vol. 1, no. 3, pp. 203–226, 1988.

[5] B. D. O. Anderson, “Adaptive systems, lack of persistency of excitation and bursting phenomena,” *Automatica*, vol. 21, no. 3, pp. 247–258, 1985.

[6] G. Chowdhary and E. Johnson, “Concurrent learning for convergence in adaptive control without persistency of excitation,” in *Proc. Conf. Dec. Contr.*, 2010, pp. 3674–3679.

[7] G. Chowdhary, M. Mühlegg, and E. Johnson, “Exponential parameter and tracking error convergence guarantees for adaptive controllers without persistency of excitation,” *Int. J.*

Contr., vol. 87, no. 8, pp. 1583–1603, 2014.

[8] S. Aranovskiy, A. Bobtsov, R. Ortega, and A. Pyrkin, "Performance enhancement of parameter estimators via dynamic regressor extension and mixing," *IEEE Trans. Autom. Contr.*, vol. 62, no. 7, pp. 3546–3550, 2017.

[9] P. Panda, J. M. Allred, S. Ramanathan, and K. Roy, "Learning to forget with adaptive synaptic plasticity in spiking neural networks," *J. Emerg. Selec. Top. Circ. Syst.*, vol. 8, no. 1, pp. 51–64, 2018.

[10] J. M. Allred and K. Roy, "Unsupervised incremental stdp learning using forced firing of dormant or idle neurons," in *2016 International Joint Conference on Neural Networks*, July 2016, pp. 2492–2499.

[11] N. Frémaux and W. Gerstner, "Neuromodulated spike-timing-dependent plasticity, and theory of three-factor learning rules," *Frontiers in Neural Circuits*, vol. 9, pp. 85–103, 2016.

[12] B. Han, A. Ankit, A. Sengupta, and K. Roy, "Cross-layer design exploration for energy-quality tradeoffs in spiking and non-spiking deep artificial neural networks," *IEEE Transactions on Multi-Scale Computing Systems*, 2017.

[13] S. A. U. Islam and D. S. Bernstein, "Recursive Least Squares for Real-Time Implementation," *IEEE Contr. Sys. Mag.*, vol. 39, pp. 82–85, June 2019.

[14] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Prentice Hall, 1999.

[15] A. H. Sayed, *Fundamentals of Adaptive Filtering*. Wiley, 2003.

[16] K. J. Astrom and B. Wittenmark, *Computer-Controlled Systems: Theory and Design*, 3rd ed. Prentice-Hall, 1996.

[17] T. R. Fortescue, L. S. Kershenbaum, and B. E. Ydstie, "Implementation of Self-tuning Regulators with Variable Forgetting Factors," *Automatica*, vol. 17, no. 6, pp. 831–835, 1981.

[18] C. Paleologu, J. Benesty, and C. Silviu, "A Robust Variable Forgetting Factor Recursive Least-Squares Algorithm for System Identification," *IEEE Sig. Proc. Lett.*, vol. 15, pp. 597–600, 2008.

[19] S.-H. Leung and C. F. So, "Gradient-Based Variable Forgetting Factor RLS Algorithm in Time-Varying Environments," *IEEE Trans. Sig. Proc.*, vol. 53, no. 8, pp. 3141–3150, 2005.

[20] S. Song, J.-S. Lim, S. J. Baek, and K.-M. Sung, "Gauss Newton variable forgetting factor recursive least squares for time varying parameter tracking," *Electron. Lett.*, vol. 36, no. 11, pp. 988–990, 2000.

[21] D. J. Park, B. E. Jun, and K. J. H., "Fast tracking RLS algorithm using novel variable forgetting factor with unity zone," *Electron. Lett.*, vol. 27, no. 23, pp. 2150–2151, 1991.

[22] A. A. Ali, J. B. Hoagg, M. Mossberg, and D. S. Bernstein, "On the stability and convergence

of a sliding-window variable-regularization recursive-least-squares algorithm," *Int. J. Adapt. Contr. Sig. Proc.*, vol. 30, pp. 715–735, 2016.

[23] R. M. Canetti and M. D. España, "Convergence analysis of the least-squares identification algorithm with a variable forgetting factor for time-varying linear systems," *Automatica*, vol. 25, no. 4, pp. 609–612, 1989.

[24] M. E. Salgado, G. C. Goodwin and R. H. Middleton, "Modified least squares algorithm incorporating exponential resetting and forgetting," *Int. J. Contr.*, vol. 47, no. 2, pp. 477–491, 1988.

[25] G. Goodwin, E. Teoh, and H. Elliott, "Deterministic convergence of a self-tuning regulator with covariance resetting," in *IEE Proceedings D-Control Theory and Applications*, vol. 130, no. 1, 1983, pp. 6–8.

[26] R. Kulhavỳ, "Restricted exponential forgetting in real-time identification," *Automatica*, vol. 23, no. 5, pp. 589–600, 1987.

[27] R. Kulhav\'y and M. Kárn\'y, "Tracking of slowly varying parameters by directional forgetting," *IFAC Proc. Vol.*, vol. 17, no. 2, pp. 687–692, 1984.

[28] G. Kreisselmeier, "Stabilized least-squares type adaptive identifiers," *IEEE Trans. Autom. Contr.*, vol. 35, no. 3, pp. 306–310, 1990.

[29] L. Cao and H. Schwartz, "Directional forgetting algorithm based on the decomposition of the information matrix," *Automatica*, vol. 36, no. 11, pp. 1725–1731, 2000.

[30] G. Kubin, "Stabilization of the RLS algorithm in the absence of persistent excitation," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, 1988, pp. 1369–1372.

[31] S. Bittanti, P. Bolzern, and M. Campi, "Convergence and exponential convergence of identification algorithms with directional forgetting factor," *Automatica*, vol. 26, no. 5, pp. 929–932, 1990.

[32] ——, "Exponential convergence of a modified directional forgetting identification algorithm," *Sys. Contr. Lett.*, vol. 14, no. 2, pp. 131–137, 1990.

[33] A. Goel and D. S. Bernstein, "A targeted forgetting factor for recursive least squares," in *Proc. Conf. Dec. Contr.*, 2018, pp. 3899–3903.

[34] R. M. Johnstone, C. R. Johnson, R. R. Bitmead, and B. D. O. Anderson, "Exponential convergence of recursive least squares with exponential forgetting factor," in *Proc. Conf. Dec. Contr.*, 1982, pp. 994–997.

[35] J. Benesty and T. Gänsler, "New Insights into the RLS Algorithm," *Eurasip Jour. App. Sig. Proc.*, no. 3, pp. 331–339, 2004.

[36] W. M. Haddad and V. Chellaboina, *Nonlinear Dynamical Systems and Control: A Lyapunov-based Approach*. Princeton University Press, 2008.

[37] P. Eykhoff, *System Identification: Parameter and State Estimation*. Wiley-Interscience,

1974.

[38] T. Söderström, *Errors-in-variables methods in system identification.* Springer, 2018.

Sidebar: Summary

Learning depends on the ability to acquire and assimilate new information. This ability depends—somewhat counterintuitively—on the ability to forget. In particular, effective forgetting requires the ability to recognize and utilize new information to order to update a system model. This article is a tutorial on forgetting within the context of recursive least squares (RLS). To do this, RLS is first presented in its classical form, which employs uniform-direction forgetting. Next, examples are given to motivate the need for variable-direction forgetting, especially in cases where the excitation is not persistent. Some of these results are well known, whereas others complement the prior literature. The goal is to provide a self-contained tutorial of the main ideas and techniques for students and researchers whose research may benefit from variable-direction forgetting.

Sidebar: Three Useful Lemmas

Lemma 1: Let $X \in \mathbb{R}^{n\times p}$ and $y \in \mathbb{R}^n$, and let $W \in \mathbb{R}^{p\times p}$ be positive definite. Then,

$$(I_n + XWX^{\mathrm{T}})^{-1} y \in \mathcal{R}([X \quad y]). \tag{S1}$$

Proof: Note that

$$\begin{aligned}
& y \in \mathcal{R}([X \quad y]) \\
& \quad = \mathcal{R}[X \quad y + XWX^{\mathrm{T}} y] \\
& \quad = \mathcal{R}\left([X \quad (I_n + XWX^{\mathrm{T}})y] \begin{bmatrix} I_p + WX^{\mathrm{T}}X & 0 \\ 0 & 1 \end{bmatrix} \right) \\
& \quad = \mathcal{R}([X(I_p + WX^{\mathrm{T}}X) \quad (I_n + XWX^{\mathrm{T}})y]) \\
& \quad = \mathcal{R}([(I_n + XWX^{\mathrm{T}})X \quad (I_n + XWX^{\mathrm{T}})y]) \\
& \quad = (I_n + XWX^{\mathrm{T}})\mathcal{R}([X \quad y]),
\end{aligned}$$

which implies (S1). □

Lemma 2: Let $A \in \mathbb{R}^{n\times n}$ be positive semidefinite, and let $\lambda > 0$. Then,

$$I_n - A(\lambda I_n + A)^{-1} > 0. \tag{S2}$$

Proof: Write $A = SDS^{\mathrm{T}}$, where $D = \operatorname{diag}(d_1, \ldots, d_n)$ is diagonal and S is unitary. For all $i \in \{1, \ldots, n\}$, $d_i \geq 0$, and thus $\frac{d_i}{\lambda + d_i} < 1$. Hence,

$$D(\lambda I_n + D)^{-1} = \operatorname{diag}\left(\tfrac{d_1}{\lambda+d_1}, \ldots, \tfrac{d_n}{\lambda+d_n}\right) < I_n. \tag{S3}$$

Pre-multiplying and post-multiplying (S3) by S and S^{T}, respectively, yields (S2). □

Lemma 3: Let $A \in \mathbb{R}^{n\times n}$ be positive semidefinite, and let $\lambda > 0$. Then,

$$I_n - \frac{1}{\lambda}\left(A - A(\lambda I_n + A)^{-1} A\right) > 0. \tag{S4}$$

Proof: Write $A = SDS^{\mathrm{T}}$, where $D = \operatorname{diag}(d_1, \ldots, d_n)$ is diagonal and S is unitary. For all $i \in \{1, \ldots, n\}$, $d_i \geq 0$, and thus $\frac{d_i}{\lambda + d_i} < 1$. Hence,

$$\frac{1}{\lambda}\left(D - D(\lambda I_n + D)^{-1} D\right) = \operatorname{diag}\left(\tfrac{d_1}{\lambda+d_1}, \ldots, \tfrac{d_n}{\lambda+d_n}\right) < I_n. \tag{S5}$$

Pre-multiplying and post-multiplying (S5) by S and S^{T}, respectively, yields (S4). □

Sidebar: RLS as a One-Step Optimal Predictor

Consider the linear system

$$x_{k+1} = A_k x_k + B_k u_k + w_{1,k}, \tag{S1}$$

$$y_k = C_k x_k + w_{2,k}, \tag{S2}$$

where, for all $k \geq 0$, $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^m$, $y_k \in \mathbb{R}^p$, and A_k, B_k, C_k are real matrices of appropriate sizes. The input u_k and output y_k are assumed to be measured. The process noise $w_{1,k} \in \mathbb{R}^n$ and the sensor noise $w_{2,k} \in \mathbb{R}^p$ are zero-mean white noise processes with variances $\mathbb{E}[w_{1,k} w_{1,k}^{\mathrm{T}}] = Q_k$ and $\mathbb{E}[w_{2,k} w_{2,k}^{\mathrm{T}}] = R_k$, respectively. The expected value of the initial state is assumed to be $\overline{x}_0$, and the variance of the initial state is P_0, that is, $\mathbb{E}[x_0] = \overline{x}_0$ and $\mathbb{E}[(x_0 - \overline{x}_0)(x_0 - \overline{x}_0)^{\mathrm{T}}] = P_0$. The objective is to estimate the state x_k given the measurements of u_k and y_k.

To estimate x_k, consider the estimator

$$\hat{x}_{k+1} = A_k \hat{x}_k + B_k u_k + K_k (y_k - C_k \hat{x}_k), \tag{S3}$$

where $\hat{x}_k$ is the estimate of x_k at step k and $\hat{x}_0 = \overline{x}_0$. The matrix K_k is constructed as follows. Define the *state-estimate error* $e_k \triangleq x_k - \hat{x}_k$ and the *state error covariance* $P_k \triangleq \mathbb{E}[e_k e_k^{\mathrm{T}}] \in \mathbb{R}^{n \times n}$. Then, e_k and P_k satisfy

$$e_{k+1} = (A_k - K_k C_k) e_k + w_{1,k} - K_k w_{2,k}, \tag{S4}$$

$$P_{k+1} = A_k P_k A_k^{\mathrm{T}} + Q_k + K_k \left(R_k + C_k P_k C_k^{\mathrm{T}}\right) K_k^{\mathrm{T}} - A_k P_k C_k^{\mathrm{T}} K_k^{\mathrm{T}} - C_k P_k A_k^{\mathrm{T}}. \tag{S5}$$

Proposition S1: Let P_{k+1} be given by (S5). The matrix K_k that minimizes $\operatorname{tr} P_{k+1}$ is given by

$$K_k = A_k P_k C_k^{\mathrm{T}} \left(R_k + C_k P_k C_k^{\mathrm{T}}\right)^{-1}, \tag{S6}$$

and the minimized state-error covariance P_k is updated as

$$P_{k+1} = A_k P_k A_k^{\mathrm{T}} + Q_k - A_k P_k C_k^{\mathrm{T}} \left(R_k + C_k P_k C_k^{\mathrm{T}}\right)^{-1} C_k P_k A_k^{\mathrm{T}}. \tag{S7}$$

Proof: See [S1]. □

Let $A_k = I_n$, $B_k = 0$, $C_k = \phi_k$, $Q_k = 0$, and $R_k = I_p$. Then,

$$\hat{x}_{k+1} = \hat{x}_k + P_k \phi_k^{\mathrm{T}} \left(I_p + \phi_k P_k \phi_k^{\mathrm{T}}\right)^{-1} (y_k - \phi_k \hat{x}_k), \tag{S8}$$

$$P_{k+1} = P_k - P_k \phi_k^{\mathrm{T}} \left(I_p + \phi_k P_k \phi_k^{\mathrm{T}}\right)^{-1} \phi_k P_k. \tag{S9}$$

Note that (6), (7) with $\lambda = 1$ have the same form as (S8), (S9). In particular, RLS without forgetting is the state estimator for the linear time-varying system with $A_k = I_n$, $B_k = 0$, $C_k = \phi_k$, $Q_k = 0$, and $R_k = I_p$.

References

[S1] S. A. U. Islam, A. Goel, and D. S. Bernstein, "Real-Time Implementation of the Optimal Predictor and Optimal Filter: Accuracy versus Latency," *IEEE Contr. Sys. Mag.*, to appear.

Sidebar: RLS as a Maximum Likelihood Estimator

Let $k \geq 0$ and, for all $i \in \{0, 1, \ldots, k\}$, consider the process

$$y_i = \phi_i \theta_{\rm true} + v_i, \tag{S1}$$

where $\theta_{\rm true} \in \mathbb{R}^n$ is the unknown parameter, $\phi_i \in \mathbb{R}^{p \times n}$ is the regressor matrix, $v_i \in \mathbb{R}^p$ is the measurement noise, and $y_i \in \mathbb{R}^p$ is the measurement. The goal is to estimate $\theta_{\rm true}$ using the data $(\phi_i)_{i=0}^k$ and $(y_i)_{i=0}^k$.

Let $\theta_{\rm true}$ be modeled by the n-dimensional, real-valued normal random variable Θ with mean $\theta_0 \in \mathbb{R}^n$ and covariance $(\lambda^{k+1}R)^{-1}$, where $\lambda \in (0, 1]$ and $R \in \mathbb{R}^{n \times n}$ is positive definite. For $\theta \in \mathbb{R}^n$, the density of Θ is thus given by

$$f_\Theta(\theta) = \frac{1}{\sqrt{(2\pi)^n \det{(\lambda^{k+1}R)^{-1}}}} \exp[-\tfrac{1}{2}(\theta - \theta_0)^{\rm T} \lambda^{k+1} R (\theta - \theta_0)]. \tag{S2}$$

For all $i \in \{0, 1, \ldots, k\}$, assume that v_i is a sample of the zero-mean, p-dimensional, real-valued normal random variable V_i with covariance $\lambda^{i-k} I_p$. For $v_i \in \mathbb{R}^p$, the density of V_i is thus given by

$$f_{V_i}(v_i) = \frac{1}{\sqrt{(2\pi)^p \lambda^{i-k}}} \exp(-\tfrac{1}{2} v_i^{\rm T} \lambda^{k-i} I_p v_i). \tag{S3}$$

Assume that $V_0, V_1, \ldots, V_k$ are independent.

Since $\theta_{\rm true}$ and v_i are modeled as normal random variables, it follows from (S1) that y_i is a sample of the p-dimensional, real-valued normal random variable $Y_i = \phi_i \theta_{\rm true} + V_i$. Note that, since $V_0, V_1, \ldots, V_k$ are independent, it follows that $Y_0, Y_1, \ldots, Y_k$ are independent. Using (S1) and (S3), it thus follows that

$$f_{Y_i|\theta}(y_i) = \frac{1}{\sqrt{(2\pi)^p \lambda^{i-k}}} \exp[-\tfrac{1}{2}(y_i - \phi_k \theta)^{\rm T} \lambda^{k-i} I_p (y_i - \phi_k \theta)], \tag{S4}$$

where $f_{Y_i|\theta}(y_i)$ is the density of the random variable Y_i conditions on Θ taking the value θ.

It follows from Bayes' rule [S1, p. 413] that

$$f_{\Theta|\{y_0, \ldots, y_k\}}(\theta) = \alpha^{-1} f_\Theta(\theta) \prod_{i=0}^k f_{Y_i|\theta}(y_i), \tag{S5}$$

where

$$\alpha \triangleq \int_{\mathbb{R}^n} f_\Theta(\theta) \prod_{i=0}^k f_{Y_i|\theta}(y_i) \, {\rm d}\theta. \tag{S6}$$

Substituting (S2) and (S4) into (S5), it follows that

$$f_{\Theta|\{y_0,\ldots,y_k\}}(\theta) = \beta \exp\left[\sum_{i=0}^{k} -\tfrac{1}{2}\lambda^{k-i}(y_i - \phi_k\theta)^{\mathrm{T}}(y_i - \phi_k\theta) - \tfrac{1}{2}\lambda^{k+1}(\theta-\theta_0)^{\mathrm{T}}R(\theta-\theta_0)\right], \tag{S7}$$

where

$$\beta \triangleq \frac{1}{\alpha\sqrt{(2\pi)^p\lambda^{i-k}}}\frac{1}{\sqrt{(2\pi)^n \det{(\lambda^{k+1}R)^{-1}}}}. \tag{S8}$$

Finally, the *maximum likelihood estimate* of θ_{true} is given by the maximizer of (S7), that is,

$$\theta_{\mathrm{ML}} = \underset{\theta\in\mathbb{R}^n}{\operatorname{argmax}}\ f_{\Theta|\{y_0,\ldots,y_k\}}(\theta). \tag{S9}$$

In fact, $\theta_{\mathrm{ML}} = \underset{\theta\in\mathbb{R}^n}{\operatorname{argmin}}\ J_k(\theta)$, where $J_k(\theta)$ is given by (2). Therefore, RLS with forgetting can be interpreted as the maximum likelihood estimator of the random variable Θ.

References

[S1] D. P. Bertsekas, and J. N. Tsitsiklis. *Introduction to Probability*, Second Edition, Athena Scientific, 2008.

Sidebar: Toward Matrix Forgetting

In [S1], P_k^{-1} is updated by

$$P_{k+1}^{-1} = (I_n + M_k P_k) P_k^{-1} + \phi_k^{\mathrm{T}} \phi_k, \tag{S1}$$

where $M_k \in \mathbb{R}^{n \times n}$ is chosen to guarantee asymptotic stability and boundedness. Two choices of matrix M_k are considered. In the first case,

$$M_k \triangleq -(1-\lambda)(I - \alpha P_k)^N P_k^{-1}, \tag{S2}$$

where $\lambda \in (0,1)$, $\alpha > 0$, and N is an odd, positive integer. In the second case,

$$M_k = -(1-\lambda)(P_k^{-1} - \alpha I_n)^N (P_k^{-1} + \beta I_n)^{-N} P_k^{-1}, \tag{S3}$$

where $\lambda \in (0,1)$, $\alpha > 0$, $\beta \geq 0$, and N is an odd, positive integer. Note that RLS with constant forgetting is obtained by setting $M_k = (\lambda - 1) P_k^{-1}$ in (S1).

Proposition S1: Consider (S1) with (S2) or (S3). Let P_0 be symmetric and nonsingular. Then, the following statements hold:

$i)$ For all $k \geq 0$, P_k is symmetric and nonsingular.
$ii)$ If $P_0^{-1} \geq \frac{\alpha}{2} I_n$, then, $P_k^{-1} = \alpha I$ is an asymptotically stable equilibrium of (S1).
$iii)$ If $P_0^{-1} \geq \alpha I_n$, then, for all $k \geq 0$, $P_k^{-1} \geq \alpha I_n$.
$iv)$ If $P_0^{-1} \geq \alpha I_n$ and, for all $k \geq 0$, ϕ_k is bounded, then P_k^{-1} is bounded.
$v)$ If $P_0^{-1} \geq \alpha I_n$ and ϕ_k is persistently exciting, then there exists $k_0 > 0$ such that, for all $k \geq k_0$, $P_k^{-1} > \alpha I_n$.

Proof: See [28]. □

The main goal of (S1) is stabilization of P_k in the case where $(\phi_k)_{k=0}^{\infty}$ is not persistently exciting. Proposition S1 implies that P_k remains bounded whether or not $(\phi_k)_{k=0}^{\infty}$ is persistent. However, (S1) is not designed to implement forgetting. Furthermore, note that (S1) requires the computation of the inverse of an $n \times n$ matrix at each step.

An alternative directional forgetting scheme given in [S2] considers the update

$$P_{k+1}^{-1} = M_k P_k^{-1} + \phi_k^{\mathrm{T}} \phi_k, \tag{S4}$$

where $M_k \in \mathbb{R}^{n \times n}$ is designed to apply forgetting to a specific subspace. In the case of a scalar measurement, that is, $p = 1$, P_k^{-1} is decomposed as

$$P_k^{-1} = P_{1,k}^{-1} + P_{2,k}^{-1}, \tag{S5}$$

where $P_{1,k}^{-1}$ is chosen such that $P_{1,k}^{-1}\phi_k^{\mathrm{T}} = 0$, that is, ϕ_k^{T} is in the null space of $P_{1,k}^{-1}$. Next, forgetting is restricted to $P_{2,k}^{-1}$, that is,

$$P_{k+1}^{-1} = P_{1,k}^{-1} + \lambda P_{2,k}^{-1} + \phi_k^{\mathrm{T}}\phi_k. \tag{S6}$$

The matrix $P_{2,k}^{-1}$ is chosen to be positive semidefinite with rank 1 by using

$$P_{2,k}^{-1} \triangleq P_k^{-1}\phi_k^{\mathrm{T}}\left(\phi_k P_k^{-1}\phi_k^{\mathrm{T}}\right)^{-1}\phi_k P_k^{-1}, \tag{S7}$$

and thus $P_{1,k}^{-1} = P_k^{-1} - P_{2,k}^{-1}$. Finally, it follows from (S4), (S6), and (S7) that

$$M_k = I_n - (1-\lambda)\left(\phi_k P_k^{-1}\phi_k^{\mathrm{T}}\right)^{-1} P_k^{-1}\phi_k^{\mathrm{T}}\phi_k \tag{S8}$$

and P_{k+1} is computed as

$$\overline{P}_k = \begin{cases} P_k + \dfrac{1-\lambda}{\lambda}\left(\phi_k P_k^{-1}\phi_k^{\mathrm{T}}\right)^{-1}\phi_k^{\mathrm{T}}\phi_k, & \phi_k \neq 0, \\ P_k, & \phi_k = 0, \end{cases} \tag{S9}$$

$$P_{k+1} = \overline{P}_k - \overline{P}_k\phi_k(1 + \phi_k^{\mathrm{T}}\overline{P}_k\phi_k)^{-1}\phi_k^{\mathrm{T}}\overline{P}_k. \tag{S10}$$

It is shown in [S2] that, if P_k^{-1} is positive definite, then, for all $\lambda \in (0,1]$, $M_k P_k^{-1}$ is positive definite. Furthermore, if, for all $k \geq 0$, ϕ_k is bounded, then there exists $\beta > 0$ such that, for all $k \geq 0$, $P_k < \beta I_n$.

References

[S1] G. Kreisselmeier, "Stabilized least-squares type adaptive identifiers," *IEEE Trans. Autom. Contr.*, vol. 35, no. 3, pp. 306–310, 1990.

[S2] L. Cao and H. Schwartz, "Directional forgetting algorithm based on the decomposition ofthe information matrix," *Automatica*, vol. 36, no. 11, pp. 1725–1731, 2000.

Sidebar: A Cost Function for Variable-Direction RLS

Theorem S1: For all $k \geq 0$, let $\phi_k \in \mathbb{R}^{p \times n}$ and $y_k \in \mathbb{R}^p$. Furthermore, let $R \in \mathbb{R}^{n \times n}$ be positive definite, let $\lambda \in (0,1]$, and, for all $k \geq 0$, let P_k be given by

$$P_{k+1}^{-1} = \Lambda_k P_k^{-1} \Lambda_k + \phi_k^{\mathrm{T}} \phi_k, \tag{S1}$$

where $P_0 \triangleq R^{-1}$ and let Λ_k be given by (68). In addition, let $\theta_0 \in \mathbb{R}^n$, and define

$$J_k(\hat{\theta}) \triangleq \sum_{i=0}^{k} (y_i - \phi_i \hat{\theta})^{\mathrm{T}} (y_i - \phi_i \hat{\theta}) + (\hat{\theta} - \theta_0)^{\mathrm{T}} R_k (\hat{\theta} - \theta_0), \tag{S2}$$

where, for all $k \geq 0$,

$$R_k = R_{k-1} + \Lambda_k P_k^{-1} \Lambda_k - P_k^{-1}, \tag{S3}$$

where $R_{-1} \triangleq R$. Then, for all $k \geq 0$, (S2) has a unique global minimizer

$$\theta_{k+1} = \underset{\hat{\theta} \in \mathbb{R}^n}{\operatorname{argmin}} \; J_k(\hat{\theta}), \tag{S4}$$

which is given by

$$\theta_{k+1} = \theta_k + P_{k+1} \phi_k^{\mathrm{T}} (y_k - \phi_k \theta_k) + P_{k+1} (R_k - R_{k-1})(\theta_0 - \theta_k). \tag{S5}$$

Proof: Note that, for all $k \geq 0$,

$$J_k(\hat{\theta}) = \hat{\theta}^{\mathrm{T}} A_k \hat{\theta} + \hat{\theta}^{\mathrm{T}} b_k + c_k,$$

where

$$A_k \triangleq \sum_{i=0}^{k} \phi_i^{\mathrm{T}} \phi_i + R_k, \tag{S6}$$

$$b_k \triangleq \sum_{i=0}^{k} -\phi_i^{\mathrm{T}} y_i - R_k \theta_0, \tag{S7}$$

$$c_k \triangleq \sum_{i=0}^{k} y_i^{\mathrm{T}} y_i + \theta_0^{\mathrm{T}} R_k \theta_0.$$

Using (S3), (S6), and (S7), it follows that, for all $k \geq 0$,

$$A_k = A_{k-1} + \Lambda_k P_k^{-1} \Lambda_k - P_k^{-1} + \phi_k^{\mathrm{T}} \phi_k, \tag{S8}$$

$$b_k = b_{k-1} - \phi_k^{\mathrm{T}} y_k - (R_k - R_{k-1}) \theta_0, \tag{S9}$$

where $A_{-1} \triangleq R$ and $b_{-1} \triangleq -R\theta_0$. Using (S1) and (S8), it follows that, for all $k \geq 0$,

$$\begin{aligned} A_k - P_{k+1}^{-1} &= A_{k-1} - P_k^{-1} \\ &= A_{-1} - P_0^{-1} \\ &= 0. \end{aligned}$$

It follows from (65) that, for all $k \geq 0$, P_{k+1}^{-1} is positive definite, and thus A_k is positive definite. Furthermore, for all $k \geq 0$, A_k is given by

$$A_k = \Lambda_k A_{k-1} \Lambda_k + \phi_k^{\mathrm{T}} \phi_k.$$

Finally, since A_k is positive definite, it follows from Lemma 1 in [S1] that

$$\begin{aligned} \theta_{k+1} &= -A_k^{-1} b_k \\ &= -A_k^{-1}(b_{k-1} - \phi_k^{\mathrm{T}} y_k - (R_k - R_{k-1})\theta_0) \\ &= -A_k^{-1}(-A_{k-1}\theta_k - \phi_k^{\mathrm{T}} y_k - (R_k - R_{k-1})\theta_0) \\ &= A_k^{-1}((A_k - R_k + R_{k-1} - \phi_k^{\mathrm{T}}\phi_k)\theta_k + \phi_k^{\mathrm{T}} y_k + (R_k - R_{k-1})\theta_0) \\ &= A_k^{-1}(A_k\theta_k + \phi_k^{\mathrm{T}}(y_k - \phi_k\theta_k) + (R_k - R_{k-1})(\theta_0 - \theta_k)) \\ &= \theta_k + A_k^{-1}\phi_k^{\mathrm{T}}(y_k - \phi_k\theta_k) + A_k^{-1}(R_k - R_{k-1})(\theta_0 - \theta_k) \\ &= \theta_k + P_{k+1}\phi_k^{\mathrm{T}}(y_k - \phi_k\theta_k) + P_{k+1}(R_k - R_{k-1})(\theta_0 - \theta_k). \end{aligned}$$

Hence, (S5) is satisfied. □

Using $R_k - R_{k-1} = \Lambda_k A_{k-1} \Lambda_k - A_{k-1}$, it follows that (S5) can be implemented without computing P_k^{-1}.

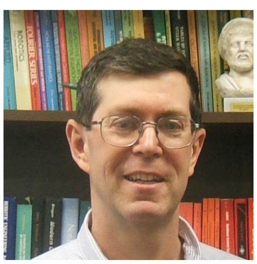
References

[S1] S. A. U. Islam, and D. S. Bernstein, "Recursive Least Squares for Real-Time Implementation," *Contr. Sys. Mag.*, vol. 39, pp. 82-85, June 2019.

Author Biographies

Ankit Goel received the B.E. degree in mechanical engineering from the Delhi College of Engineering, Delhi, the M.S. and the Ph.D. degree in aerospace engineering from the University of Michigan in Ann Arbor. His interests are in data-driven estimation and control of high-dimensional complex systems.

Adam L. Bruce received the B.A. in Physics and Applied Mathematics from the University of California, Berkeley, the M.S. in Aeronautics and Astronautics from Purdue University, and is currently a PhD Candidate at the University of Michigan, Ann Arbor. His interests are in identification, estimation, and adaptive control for nonlinear, uncertain, and complex systems.



Dennis S. Bernstein received the Sc.B. degree from Brown University and the Ph.D. degree from the University of Michigan in Ann Arbor, Michigan, where he is currently professor in the Aerospace Engineering Department. His interests are in identification, estimation, and control for aerospace applications. He is the author of *Scalar, Vector, and Matrix Mathematics*, published by Princeton University Press.

About This Issue: *In "Recursive Least Squares with Variable-Direction Forgetting: Compensating for the loss of persistency ," Ankit Goel, Adam Bruce and Dennis Bernstein revisit the classical recursive least squares (RLS) algorithm with uniform forgetting and propose a variable-direction forgetting scheme. They carefully define the notion of persistent excitation and present convergence properties of the classical RLS algorithm in the case of persistent excitation. Motivated by numerical studies, they propose a variable-direction forgetting scheme to prevent the covariance matrix from diverging in the case where persistent excitation is lacking. Their article provides a tutorial introduction on an important topic in systems theory.*

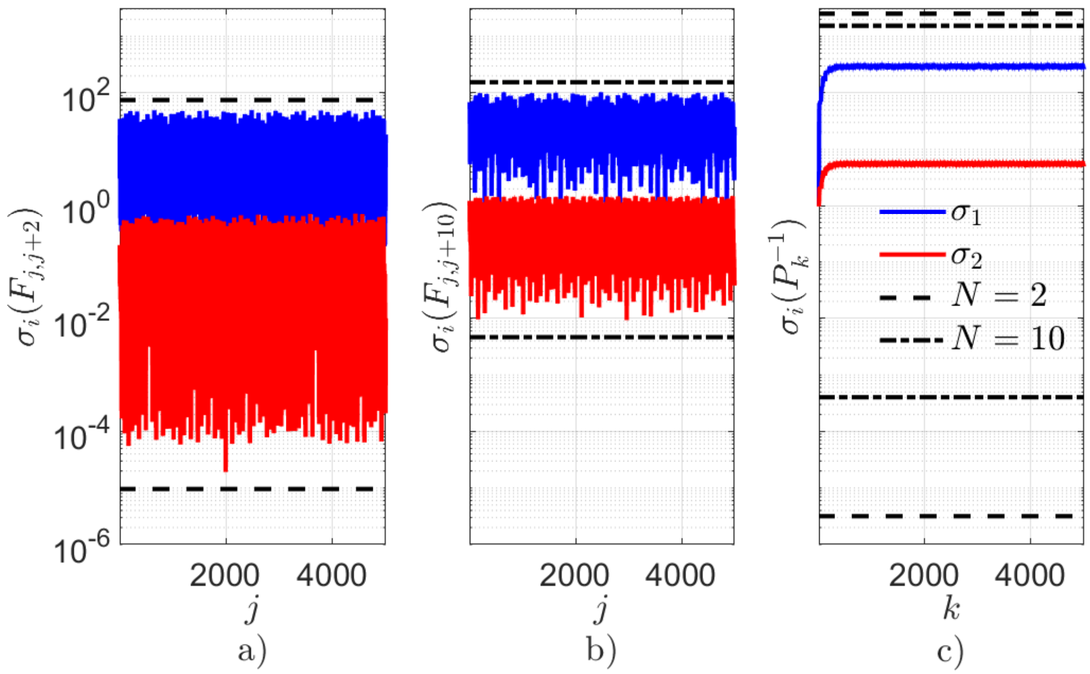


Figure 1: Example 2. Persistent excitation and bounds on P_k^{-1}. a) and b) show the singular values of $F_{j,j+N}$ for $N = 2$ and $N = 10$, where α and β are chosen to satisfy (19). Since u_k is periodic, it follows that, for all $j \geq 0$, the lower and upper bounds (19) for $F_{j,j+N}$ are satisfied. Hence, $(\phi_k)_{k=0}^{\infty}$ is persistently exciting. c) shows the singular values of P_k^{-1}, with corresponding bounds given by (25) for $\lambda = 0.99$. Note that α and β are larger for $N = 10$ than for $N = 2$, as expected.

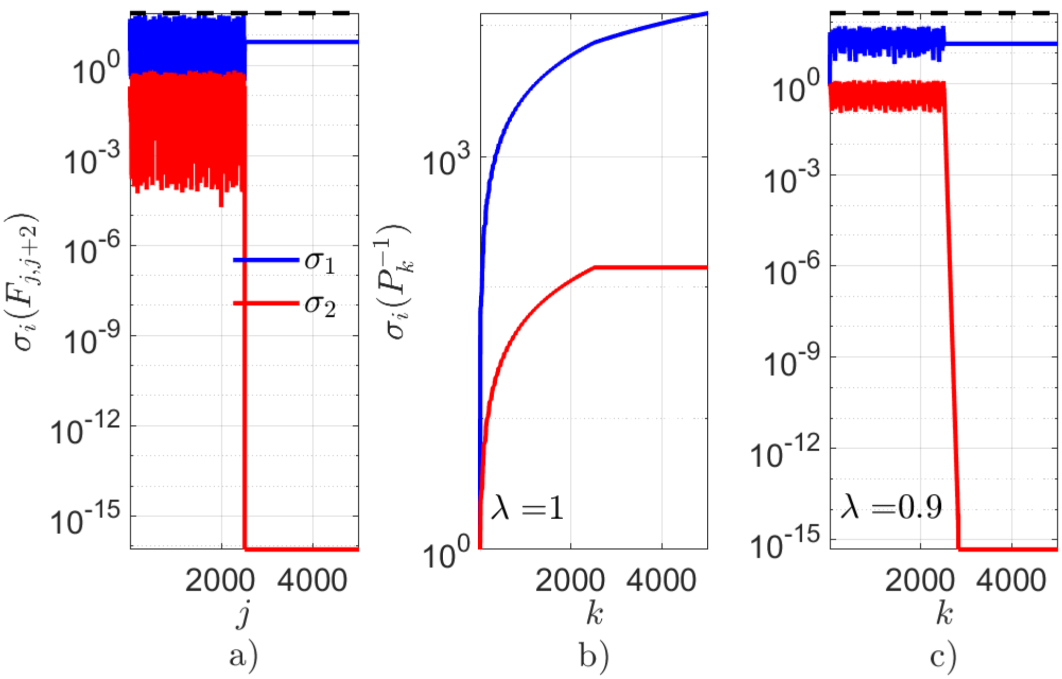


Figure 2: Example 3. Lack of persistent excitation and bounds on P_k^{-1}. a) shows the singular values of $F_{j,j+2}$. Note that the smaller singular value of $F_{j,j+2}$ reaches zero in machine precision, and thus that $\alpha > 0$ satisfying (19) does not exist. Hence, ϕ_k is not persistently exciting. The upper bound β shown by the dashed line is chosen to satisfy (19). b) and c) show the singular values of P_k^{-1} for $\lambda = 1$ and $\lambda = 0.9$, respectively. Note that, if $\lambda = 1$, then one of the singular values of P_k^{-1} diverges, whereas, if $\lambda \in (0, 1)$, then one of singular values of P_k^{-1} converges to zero.

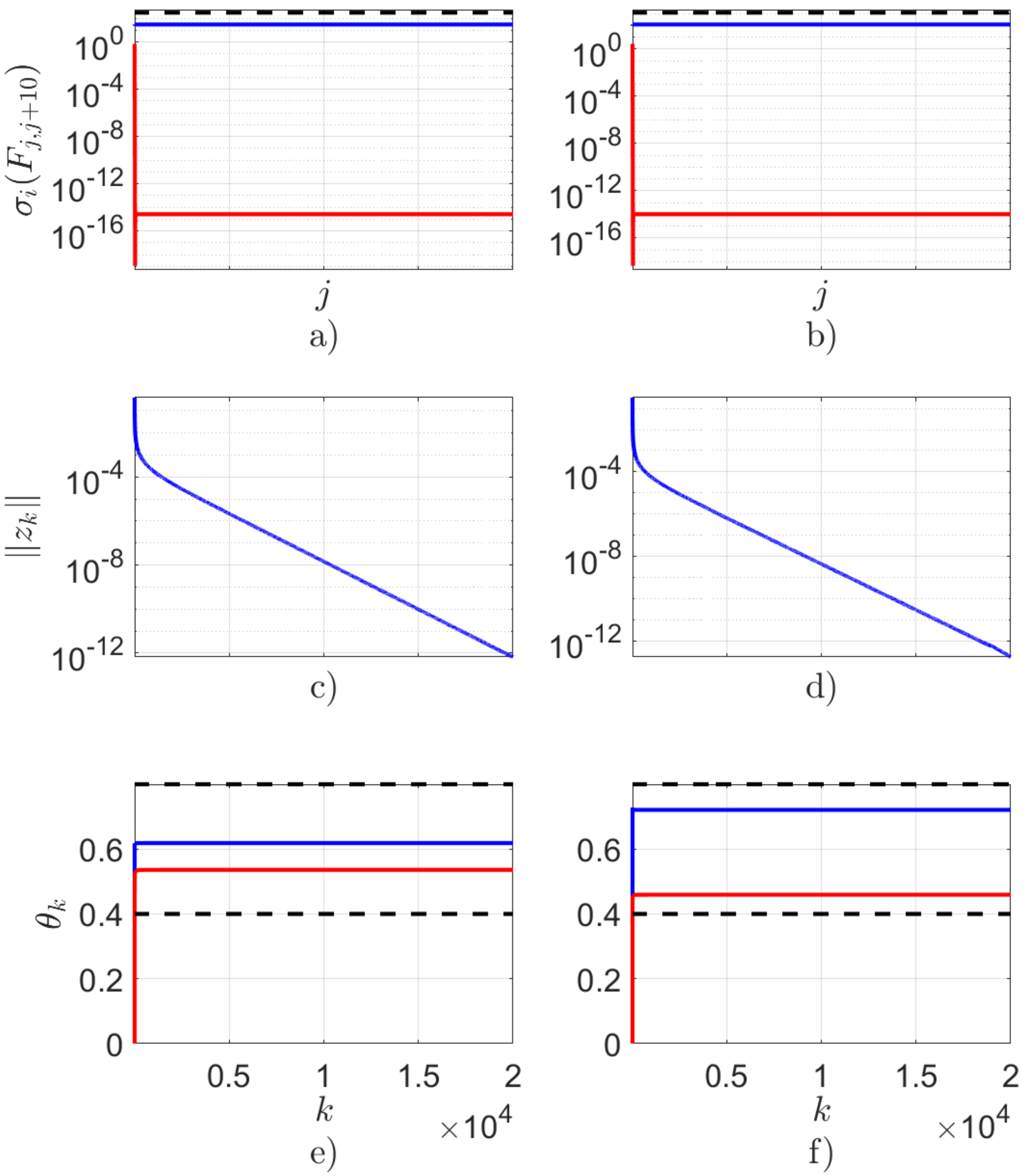


Figure 3: Example 4. Convergence of z_k and θ_k. a) and b) show the singular values of $F_{j,j+10}$ for two choices of u_k. Note that the singular value of $F_{j,j+10}$ that is close to machine precision ($\approx 10^{-15}$) is essentially zero. Definition 1 thus implies that $(\phi_k)_{k=0}^{\infty}$ is not persistently exciting. c) and d) show the predicted error z_k for both cases. Note that z_k converges to zero in both cases. Finally, e) and f) show the parameter estimate θ_k for both cases. Note that, for both choices of input u_k, θ_k converge, but to different parameter values.

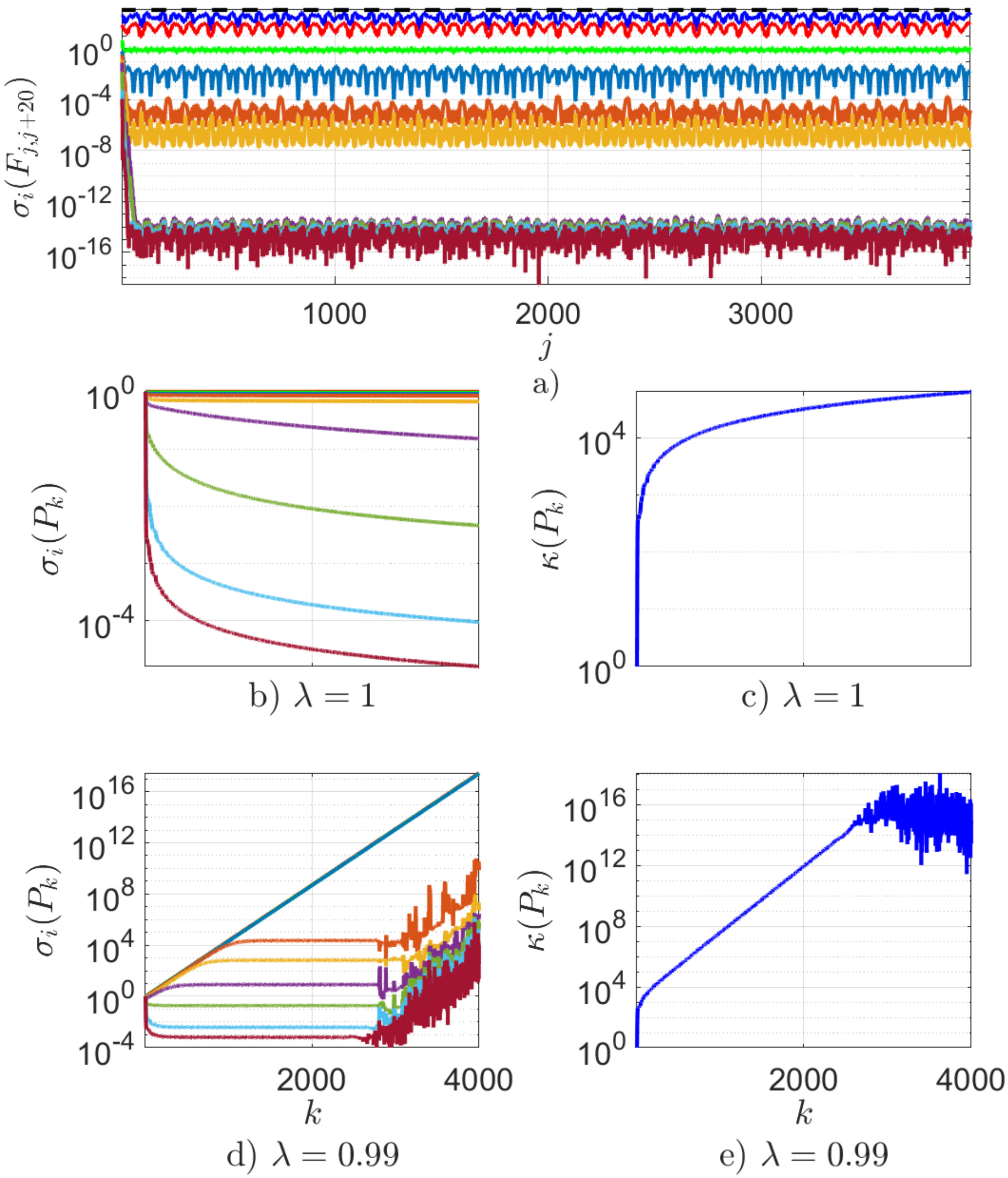


Figure 4: Example 5. Using the condition number of P_k to evaluate persistency. a) shows the singular values of $F_{j,j+20}$, where the singular values of $F_{j,j+20}$ close to machine precision ($\approx 10^{-15}$) are essentially zero, thus implying that $(\phi_k)_{k=0}^{\infty}$ is not persistently exciting. b) and c) shows the singular values and the condition number of P_k for $\lambda = 1$. Note that the six singular values of P_k decrease due to the presence of three harmonics in u_k. d) and e) shows the singular values and the condition number of P_k for $\lambda = 0.99$. Note that the six singular values of P_k remain bounded due to the presence of three harmonics in u_k. However, P_k becomes ill-conditioned due to the lack of persistent excitation.

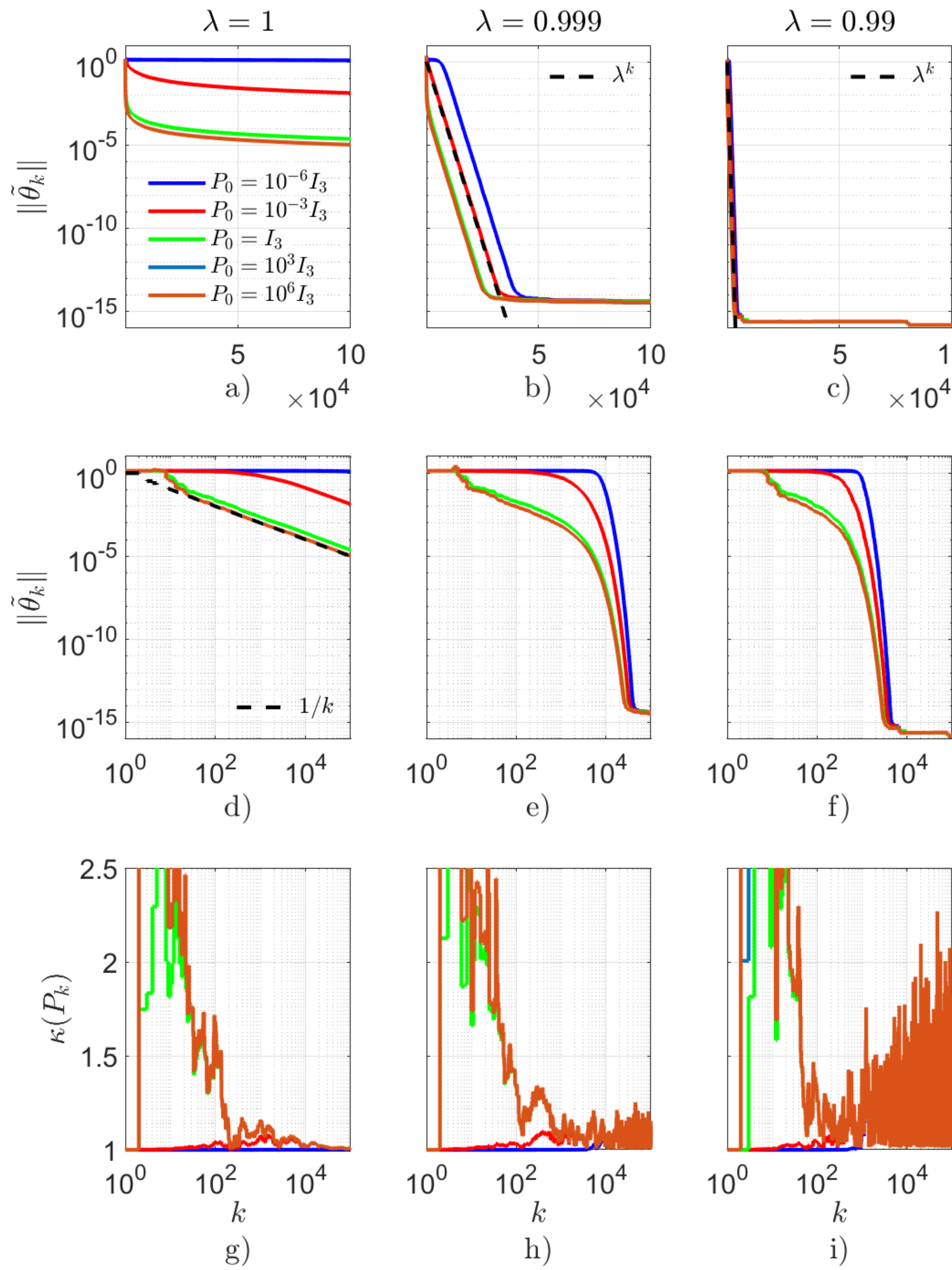
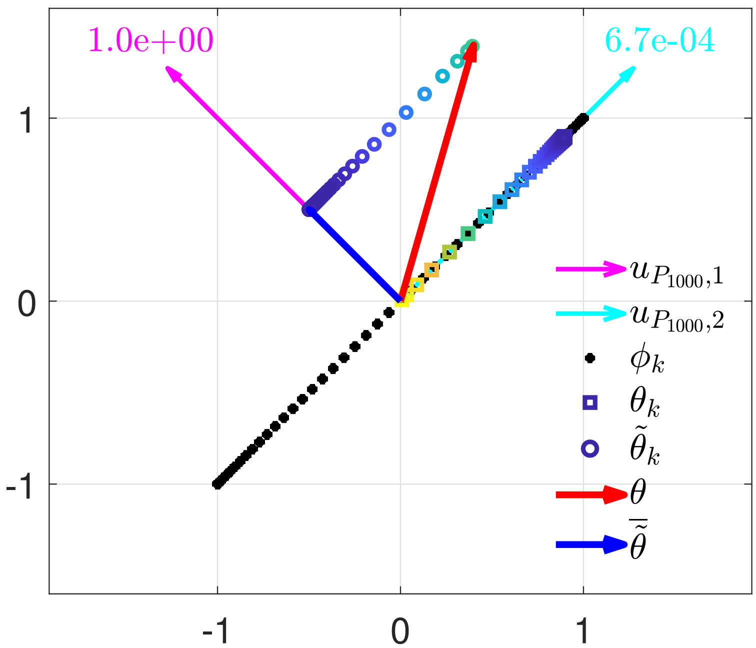
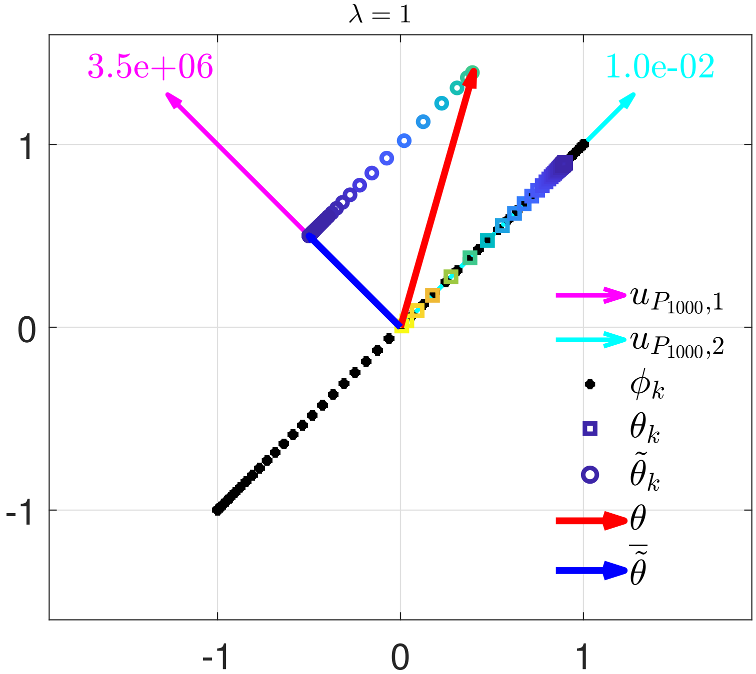


Figure 5: Example 6. Effect of λ on the rate of convergence of θ_k. a)-f) show the parameter error norm $\|\tilde{\theta}_k\|$ for several values of P_0 and λ. Note that the slope of -1 between $\log \|\tilde{\theta}_k\|$ and $\log k$ in d) is consistent with the fact that the rate of convergence of $\|\tilde{\theta}_k\|$ is $O(1/k)$ for $\lambda = 1$. Similarly, the slope of $\log \lambda$ between $\log \|\tilde{\theta}_k\|$ and k in b) and c) is consistent with the fact that the rate of convergence of $\|\tilde{\theta}_k\|$ is $O(\lambda^k)$ for $\lambda \in (0, 1)$. g), h), and i) show the condition number of the corresponding P_k for several values of P_0 and λ. Note that, as λ is decreased, the convergence rate of θ_k increases; however, the condition number of P_k degrades, and the effect of P_0 is reduced.



$\lambda = 1$



$\lambda = 0.99$

Figure 6: Example 8. Subspace constrained regressor. The first component of each vector is plotted along the horizontal axis, and the second component is plotted along the vertical axis. The singular values $\sigma_i(P_{1000})$ are shown with the corresponding singular vector $u_{P_{1000},i}$. All regressors ϕ_k lie along the same one-dimensional subspace, and thus, $(\phi_k)_{k=0}^{\infty}$ is not persistently exciting. Consequently, each estimate θ_k of θ lies in this subspace. The color gradient from yellow to blue of θ_k and $\tilde{\theta}_k$ shows the evolution from $k = 1$ to $k = 1000$. In a), the singular value corresponding to the cyan singular vector decreases to zero, whereas the singular value corresponding to the magenta singular vector is bounded. Note that $\tilde{\theta}_k$ converges along the singular vector corresponding to the bounded singular value. In b), the singular value corresponding to the cyan singular vector is bounded, whereas the singular value corresponding to the magenta singular vector diverges. Note that $\tilde{\theta}_k$ converges along the singular vector corresponding to the diverging singular value.

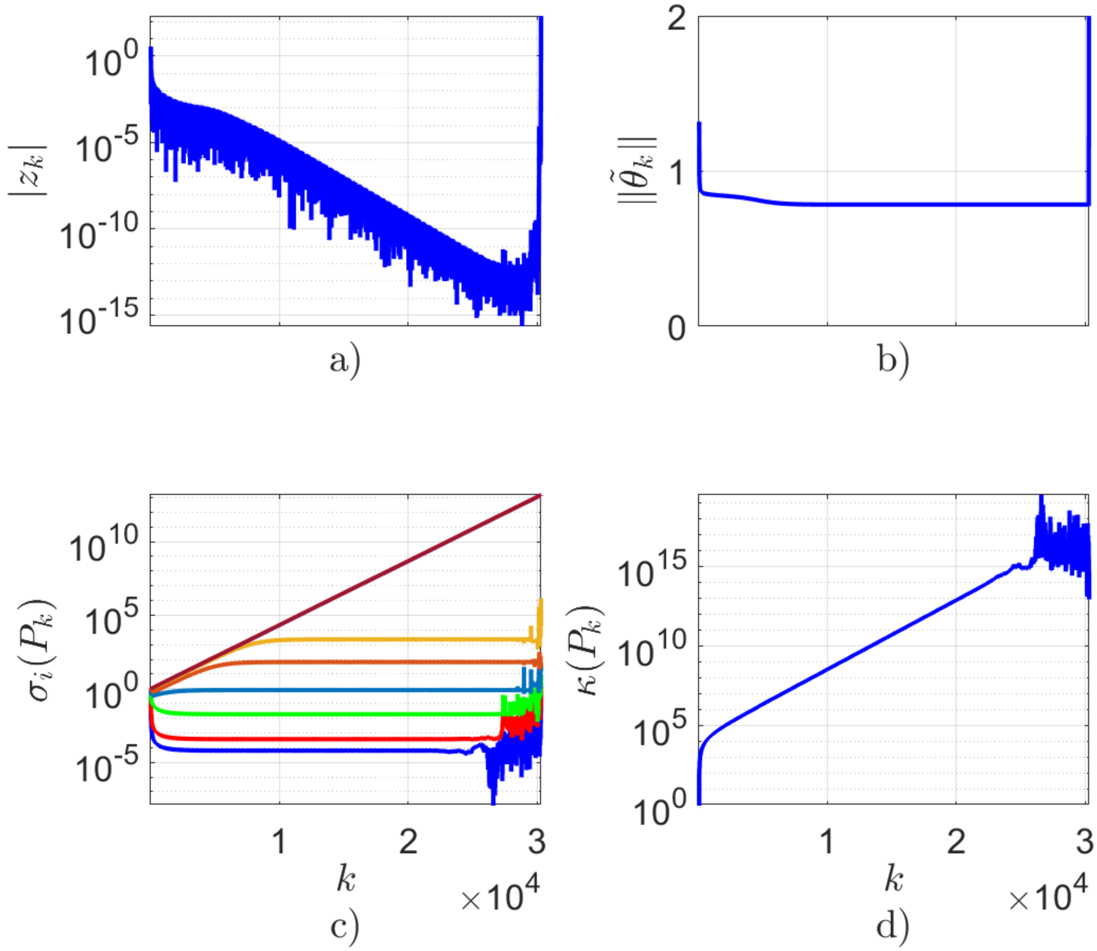


Figure 7: Example 9. Effect of lack of persistent excitation on θ_k. a) shows the predicted error z_k, b) shows the norm of the parameter error $\tilde{\theta}_k$, c) shows the singular values of P_k, and d) shows the condition number of P_k. Note that six singular values of P_k remain bounded due to the presence of three harmonics in the regresssor. Due to finite-precision arithmetic, the computation becomes erroneous as P_k becomes numerically ill-conditioned, and thus, the estimate θ_k diverges.

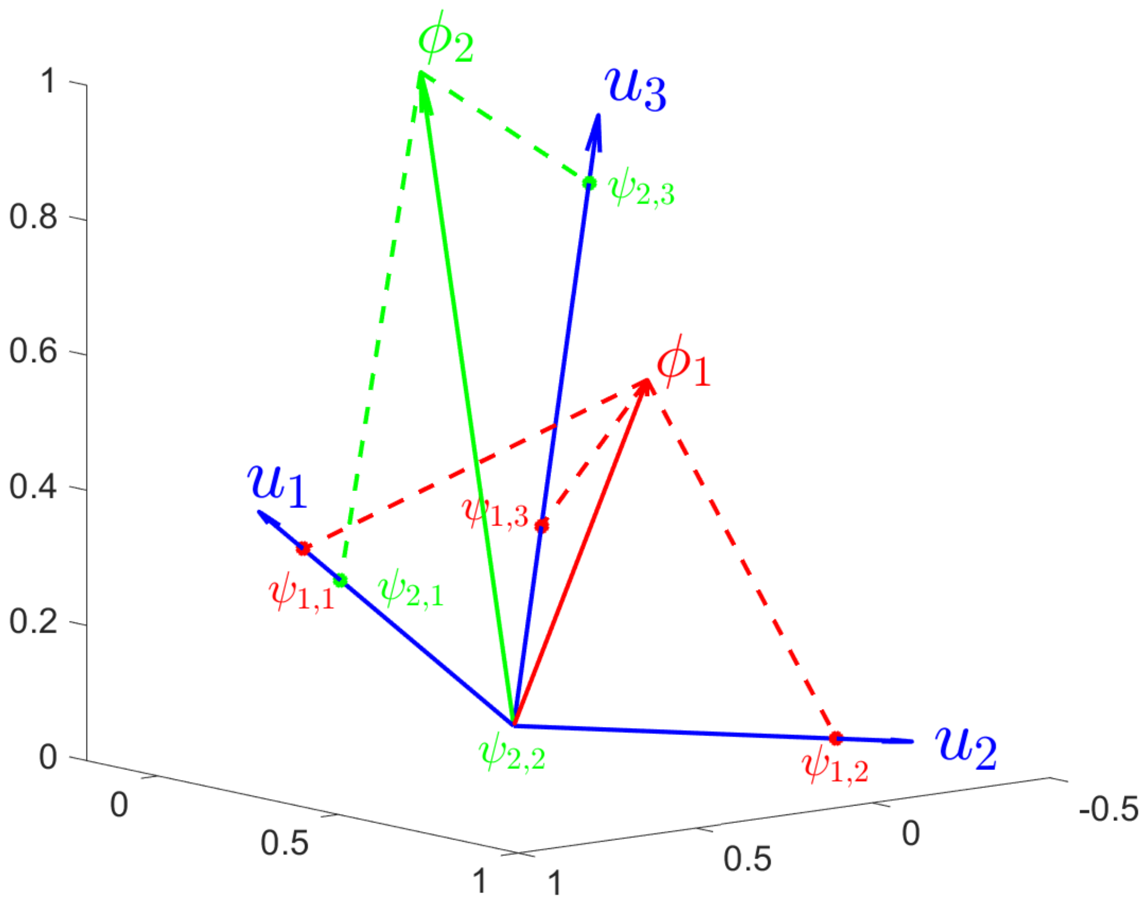


Figure 8: Illustrative example of the information-rich subspace. Let u_1, u_2, and u_3 be the information directions (shown in blue). The regressor ϕ_1 (shown in red) has new information along all three information directions, as shown by the nonzero values $\psi_{1,1}$, $\psi_{1,2}$, and $\psi_{1,3}$; the information-rich subspace is thus $\mathcal{R}([u_1\ u_2\ u_3])$. On the other hand, the regressor ϕ_2 (shown in green) has new information only along u_1 and u_3, as shown by the nonzero values $\psi_{2,1}$ and $\psi_{2,3}$; the information-rich subspace is thus $\mathcal{R}([u_1\ u_3])$.

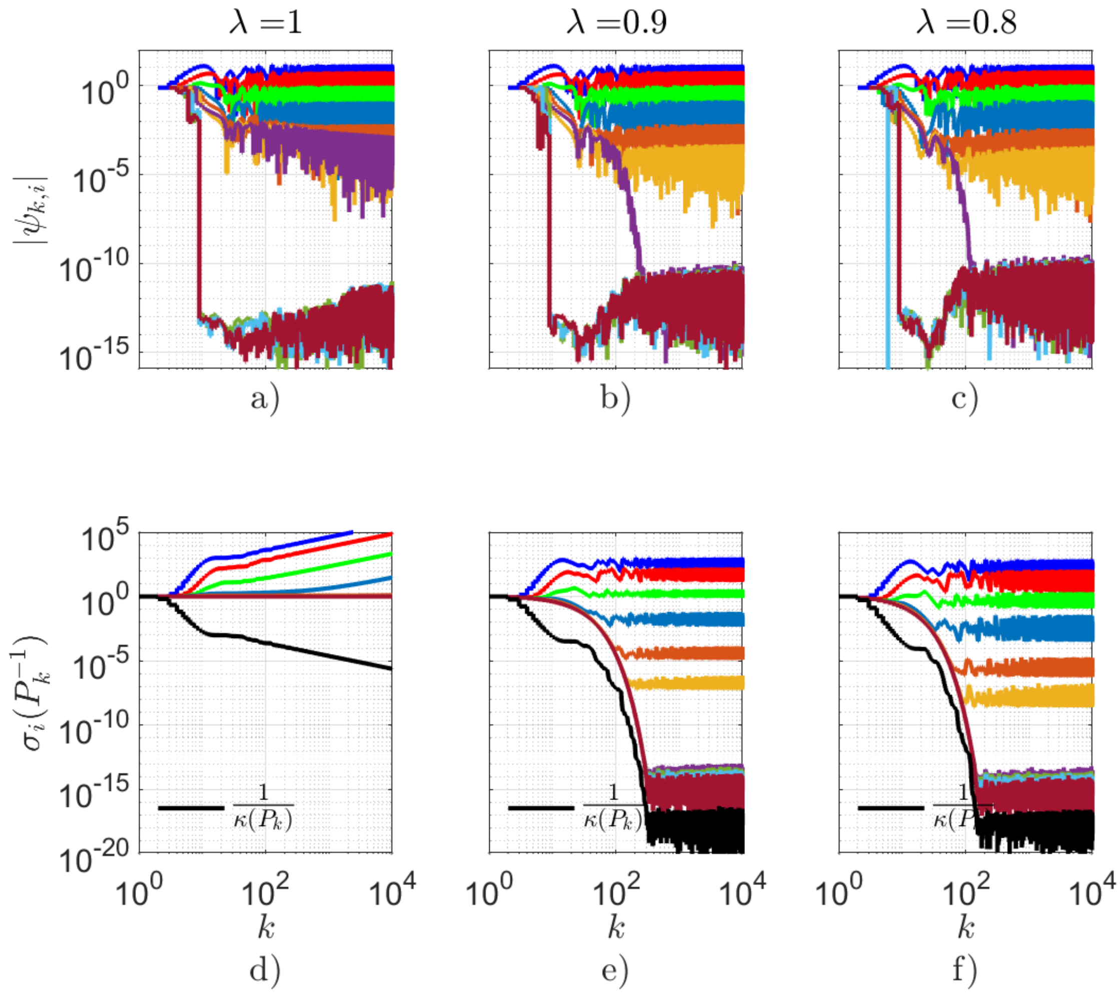


Figure 9: Example 10. Relation between P_k and the information content ψ_k. a), b), and c) show the information content $\mathrm{col}_i(\psi_k)$ for several values of λ. Note that, in each case, the information-rich subspace is six dimensional due to the presence of three harmonics in u_k. d), e), and (f) show the singular values of P_k^{-1} for several values of λ. The inverse of the condition number of P_k is shown in black. Note that, for $\lambda < 1$, the singular values of P_k^{-1} corresponding to the singular vectors in the orthogonal complement of the information-rich subspace converge to zero.

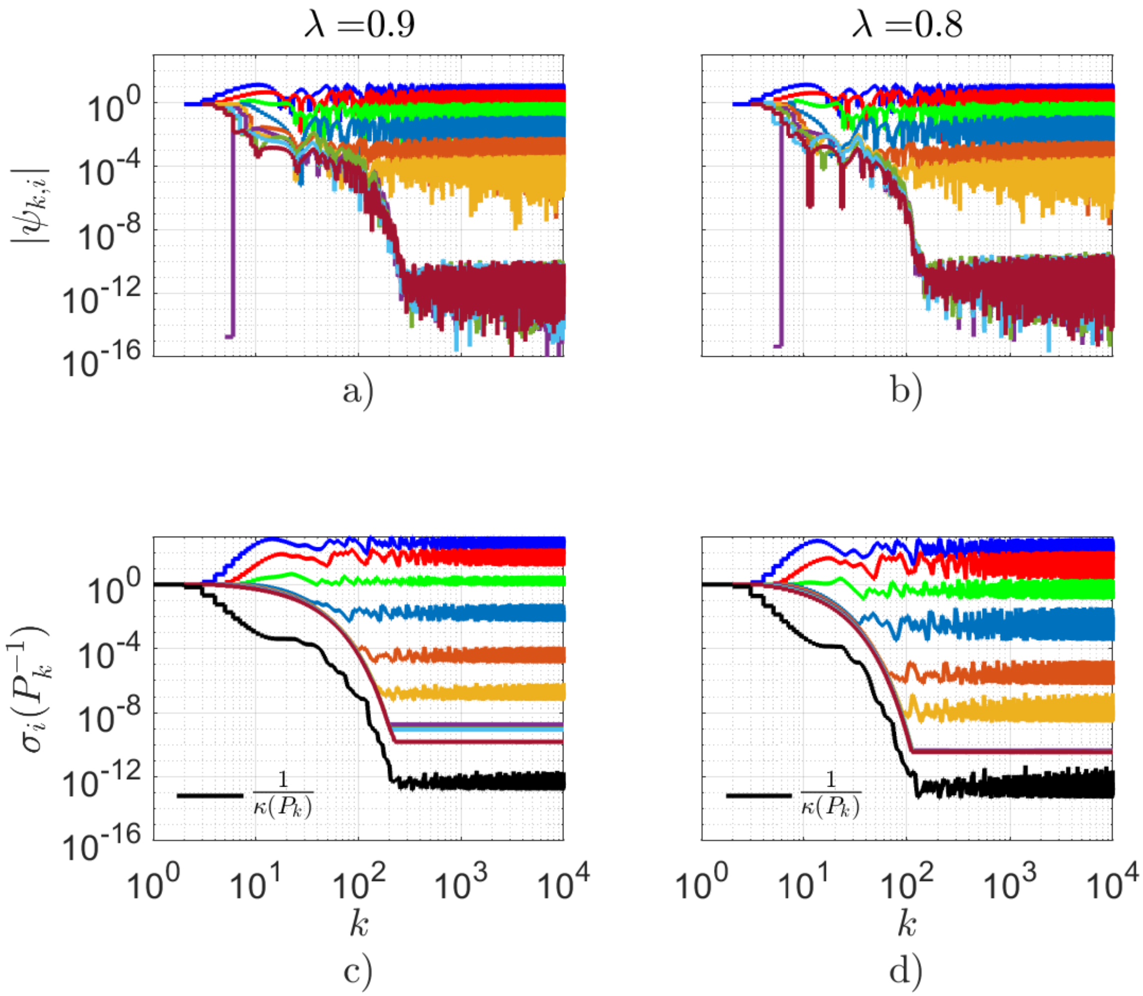


Figure 10: Example 11. Variable-direction forgetting for a regressor lacking persistent excitation. a) and b) show the information content $\|\psi_k\|$ for $\lambda = 0.9$ and $\lambda = 0.8$. c) and d) show the singular values of P_k^{-1} for $\lambda = 0.9$ and $\lambda = 0.8$. The inverse of the condition number of P_k is shown in black. Note that, for $\lambda < 1$, the singular values that correspond to the singular vectors not in the information-rich subspace do not converge to zero.

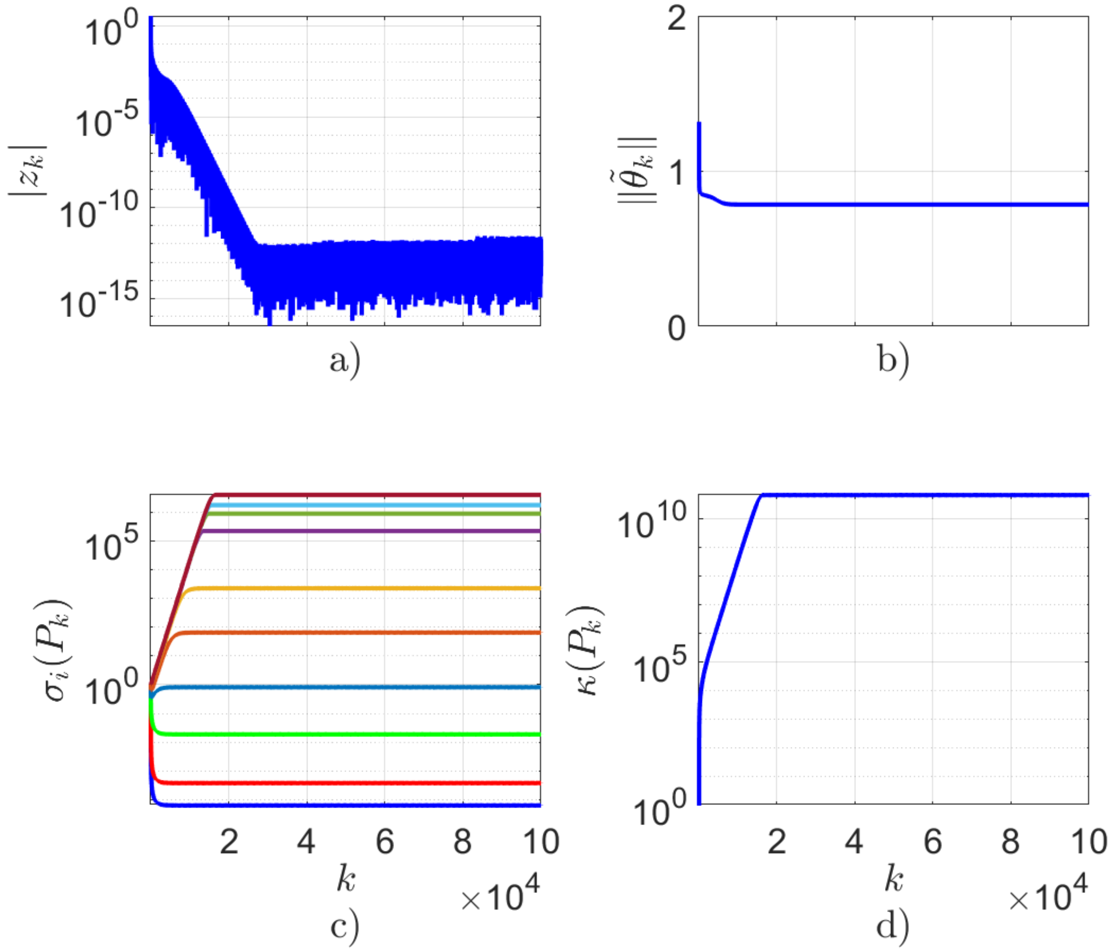


Figure 11: Example 12. Effect of variable-direction forgetting on θ_k. a) shows the predicted error z_k, b) shows the norm of the parameter error $\tilde{\theta}_k$, c) shows the singular values of P_k, and d) shows the condition number of P_k. Note that all of the singular values of P_k remain bounded.