SIFt-RLS: Subspace of Information Forgetting Recursive Least Squares

Brian Lai [a] and Dennis S. Bernstein [a]

[a] *Department of Aerospace Engineering, The University of Michigan, Ann Arbor, Michigan 48109, United States*

Abstract

This paper presents subspace of information forgetting recursive least squares (SIFt-RLS), a directional forgetting algorithm which, at each step, forgets only in row space of the regressor matrix, or the *information subspace*. As a result, SIFt-RLS tracks parameters that are in excited directions while not changing parameter estimation in unexcited directions. It is shown that SIFt-RLS guarantees an upper and lower bound of the covariance matrix, without assumptions of persistent excitation, and explicit bounds are given. Furthermore, sufficient conditions are given for the uniform Lyapunov stability and global uniform exponential stability of parameter estimation error in SIFt-RLS when estimating fixed parameters without noise. SIFt-RLS is compared to other RLS algorithms from the literature in a numerical example without persistently exciting data.

Key words: System Identification; Recursive Identification; Parameter Estimation; Adaptive Systems; Directional Forgetting

1 Introduction

Recursive least squares (RLS) is a fundamental algorithm in systems and control theory for the online identification of fixed parameters [1,21]. An important property of RLS is that the eigenvalues of the covariance matrix are monotonically decreasing and may become arbitrarily small [30, subsection 2.3.2], [10], resulting in eventual slow adaptation and inability to track time-varying parameters [19, 25, 29]. A classical extension of RLS used to track time-varying parameters is exponential forgetting [14, 15], which uses a forgetting factor to put exponentially lower weights on older terms in the least squares cost function. While this allows for continued adaptation of parameter estimation, a critical issue that arises is, without persistent excitation, at least one of the eigenvalues the covariance matrix may become arbitrarily large [10, 15], a phenomenon known as covariance blow up [22] or covariance windup [1, p. 473]. This results in significant sensitivity to measurement noise and poor parameter estimation [10].

As such, a well-established heuristic for RLS extensions is that the eigenvalues of the covariance matrix should have an upper bound and positive lower bound [29]. Later works have shown that conditions related to the boundedness of the covariance matrix are sufficient to guarantee stability of the parameter estimation error [18, 28, 31]. Common approaches to bound the covariance matrix and further improve tracking of time-varying parameters include variable-rate forgetting, in which a time-varying forgetting factor is used [4, 7, 8, 20, 23, 26], and resetting strategies, in which the covariance matrix is periodically reset to a desired value [11] or slowly converges back to a desired value when there is little excitation [17, 29, 32]. These methods address the fact that there may be periods of rich excitation, in which forgetting old data can more effectively adapt parameter estimates, and periods of poor excitation, in which it is preferable to rely on old data for parameter estimation.

While variable-rate forgetting and resetting strategies address periods of rich and poor excitation, these methods do not address when excitation is nonuniform. Nonuniform excitation may result in sufficient excitation to quickly adapt parameters in particular *directions* (that is, particular linear combinations of the parameters), while parameters in other directions can only be slowly adapted or not adapted at all [16, 29, 34]. If it is known a priori how quickly parameters will vary in different directions and which directions will be excited, techniques such as multiple forgetting may be suitable, where different forgetting factors are chosen for different directions [9, 34]. However, a more challenging problem is forgetting when such information is not

Email addresses: `brianlai@umich.edu` (Brian Lai), `dsbaero@umich.edu` (Dennis S. Bernstein).

known a priori. To tackle this problem, various directional forgetting techniques have been developed which analyze the regressor and/or the covariance matrix to gauge which directions are being excited and which are not [3, 5, 10, 16, 35].

This work presents a new directional forgetting algorithm called subspace of information forgetting (SIFt) RLS. We begin by presenting a subspace decomposition of a positive definite matrix into the sum of two positive semidefinite matrices, one *parallel* to the subspace and one *orthogonal* to the subspace. This decomposition reduces to the decomposition used in [5] when the subspace is of dimension 1. We give a thorough analysis of this decomposition, including the main properties of the decomposition, the uniqueness of this decomposition, and a duality between the parallel and orthogonal components.

Next, we develop the SIFt-RLS algorithm, where, at each step, we call the row space of the regressor the *information subspace*. At each step, the inverse covariance matrix (also called the information matrix) is decomposed into the components parallel and and orthogonal to the information subspace, and forgetting is applied only to the parallel component. This approach is developed for vector measurements and specializes to the method in [5] in the case of scalar measurements. Forgetting techniques for parameter estimation with vector measurements are useful, for example, in adaptive control of multiple-input, multiple-output (MIMO) systems [14, 24].

Finally, we give explicit upper and lower bounds for the eigenvalues of the covariance matrix in SIFt-RLS that are guaranteed without persistent excitation. This goes beyond the analysis of [5] which only shows the existence of bounds. Moreover, we show a striking similarity between the eigenvalue bounds of exponential forgetting that are guaranteed with persistent excitation and the eigenvalue bounds of SIFt. This shows that forgetting in the information subspace yields analogous bounds to uniform forgetting when all directions are excited. Furthermore, a counterexample shows that [35], a different vector measurement extension of [5], does not guarantee an upper bound on the eigenvalues of the covariance matrix as SIFt does. Finally, we provide sufficient conditions for the uniform Lyapunov stability and global uniform exponential stability of the estimation error dynamics of SIFt-RLS.

1.1 Notation and Terminology

I_n denotes the $n \times n$ identity matrix, and $0_{m\times n}$ denotes the $m \times n$ zero matrix. For $B \in \mathbb{R}^{m\times n}$, $\boldsymbol{\sigma}_{\mathbf{max}}(B)$ denotes the largest singular value of B, and $\boldsymbol{\sigma}_{\mathbf{min}}(B)$ denotes the smallest singular value of B. For symmetric $A \in \mathbb{R}^{n\times n}$, let the n real eigenvalues of A be denoted by $\boldsymbol{\lambda}_{\mathbf{min}}(A) \triangleq \boldsymbol{\lambda}_{\boldsymbol{n}}(A) \leq \cdots \leq \boldsymbol{\lambda}_{\mathbf{max}}(A) \triangleq \boldsymbol{\lambda}_{\mathbf{1}}(A)$. Additionally, the spectral radius of A, $\rho(A)$, is defined as

$$\rho(A) \triangleq \max\{|\boldsymbol{\lambda}_{\mathbf{1}}(A)|, \cdots, |\boldsymbol{\lambda}_{\boldsymbol{n}}(A)|\}. \tag{1}$$

Furthermore, if A is positive definite, the singular values of A are equivalent to the eigenvalues of A and the condition number of A, $\kappa(A)$, can be expressed as

$$\kappa(A) \triangleq \frac{\boldsymbol{\sigma}_{\mathbf{max}}(A)}{\boldsymbol{\sigma}_{\mathbf{min}}(A)} = \frac{\boldsymbol{\lambda}_{\mathbf{max}}(A)}{\boldsymbol{\lambda}_{\mathbf{min}}(A)}. \tag{2}$$

If A is positive semidefinite but not positive definite, then $\kappa(A) \triangleq \infty$.

For $x \in \mathbb{R}^n$, $\|x\|$ denotes the Euclidean norm of x. For $B \in \mathbb{R}^{m\times n}$, B^+ denotes the Moore-Penrose inverse of B. $\mathcal{R}(B)$ and $\mathcal{N}(B)$ denote the column space and nullspace of B, respectively. For positive-semidefinite $R \in \mathbb{R}^{n\times n}$ and $x \in \mathbb{R}^n$, $\|x\|_R \triangleq \sqrt{x^{\mathrm{T}}Rx}$. Let $(a_k)_{k=0}^{\infty} \subset \mathbb{R}^{m\times n}$ denote that, for all $k \geq 0$, $a_k \in \mathbb{R}^{m\times n}$. For $P, Q \in \mathbb{R}^{n\times n}$, let $P \prec Q$ (respectively $P \preceq Q$) denote that $Q - P$ is positive definite (respectively positive semidefinite).

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

$$\alpha I_n \preceq \sum_{i=k}^{k+N} \phi_i^{\mathrm{T}}\phi_i. \tag{3}$$

Furthermore, α and N are, respectively, the ***lower bound*** *and* ***persistency window*** *of $(\phi_k)_{k=0}^{\infty}$.*

Definition 2 *A sequence $(\phi_k)_{k=0}^{\infty} \subset \mathbb{R}^{p\times n}$ is* ***bounded*** *if there exist $\beta > 0$ such that, for all $k \geq 0$,*

$$\phi_k^{\mathrm{T}}\phi_k \preceq \beta I_n. \tag{4}$$

Furthermore, β is the ***upper bound*** *of $(\phi_k)_{k=0}^{\infty}$.*

2 RLS with Exponential Forgetting

To begin, we introduce RLS with exponential forgetting in Proposition 1, which gives a recursive way to compute the global minimizer of a least squares cost with exponentially smaller weighting on older terms. This background will be useful for later comparison with SIFt-RLS.

Proposition 1 *For all $k \geq 0$, let $\phi_k \in \mathbb{R}^{p\times n}$ and $y_k \in \mathbb{R}^p$. Furthermore, let $\theta_0 \in \mathbb{R}^n$, let $R_0 \in \mathbb{R}^{n\times n}$ be positive definite, and let $\lambda \in (0, 1)$. For all $k \geq 0$, define the function $J_k\colon \mathbb{R}^n \to \mathbb{R}$ by*

$$J_k(\hat\theta) \triangleq \lambda^{k+1}\|\hat\theta - \theta_0\|^2_{R_0} + \sum_{i=0}^{k} \lambda^{k-i}\|y_i - \phi_i\hat\theta\|_2. \tag{5}$$

Then, J_k has a unique global minimizer, denoted

$$\theta_{k+1} \triangleq \underset{\hat{\theta}\in\mathbb{R}^n}{\arg\min}\, J_k(\hat{\theta}), \tag{6}$$

which, for all $k \geq 0$, is given recursively by

$$R_{k+1} = \lambda R_k + \phi_k^{\mathrm{T}}\phi_k, \tag{7}$$
$$\theta_{k+1} = \theta_k + P_{k+1}\phi_k^{\mathrm{T}}(y_k - \phi_k\theta_k), \tag{8}$$

where, for all $k \geq 0$, $R_k \in \mathbb{R}^{n\times n}$ is positive definite and $P_k \triangleq R_k^{-1} \in \mathbb{R}^{n\times n}$. Moreover, for all $k \geq 0$, P_{k+1} is given recursively by

$$P_{k+1} = \frac{1}{\lambda}P_k - \frac{1}{\lambda}P_k\phi_k^{\mathrm{T}}(\lambda I_p + \phi_k P_k \phi_k^{\mathrm{T}})^{-1}\phi_k P_k. \tag{9}$$

Proof. See Theorem 1 of [14]. ■

For all $k \geq 0$, We call $y_k \in \mathbb{R}^p$ the *measurement*, $\phi_k \in \mathbb{R}^{p\times n}$ the *regressor*, and $\theta_k \in \mathbb{R}^n$ the *parameter estimate*. Moreover, we call $R_k \in \mathbb{R}^{n\times n}$ is the *information matrix*, and $P_k \in \mathbb{R}^{n\times n}$ the *covariance matrix*. Finally, we call $\lambda \in (0,1)$ the *forgetting factor*.

Note that RLS with exponential forgetting may experience covariance windup, where some of the eigenvalues of the covariance matrix P_k may become unbounded as $k \to \infty$ (e.g. see example 3 of [10]). However, if the sequence of regressors $(\phi_k)_{k=0}^\infty$ is persistently exciting and bounded, then P_k has guaranteed upper and lower bounds [10]. Proposition 2 gives an expression for these bounds in the special case where the persistency window is $N = 1$. For bounds in the case $N > 1$, see Proposition 4 of [10].

Proposition 2 *Let $\lambda \in (0,1)$, let $\theta_0 \in \mathbb{R}^n$, and let $R_0 \in \mathbb{R}^{n\times n}$ be positive definite. For all $k \geq 0$, let $y_k \in \mathbb{R}^p$, and $\phi_k \in \mathbb{R}^{p\times n}$. Finally, for all $k \geq 0$, let $R_k \in \mathbb{R}^{n\times n}$ and $\theta_k \in \mathbb{R}^n$ be recursively updated by* (7) *and* (8) *and $P_k \triangleq R_k^{-1}$. Then the following statements hold:*

(1) If $(\phi_k)_{k=0}^\infty$ is persistently exciting with lower bound $\alpha > 0$ and persistency window $N = 1$, then, for all $k \geq 0$,

$$\boldsymbol{\lambda}_{\mathbf{min}}(R_k) \geq \min\{\frac{\alpha}{1-\lambda}, \boldsymbol{\lambda}_{\mathbf{min}}(R_0)\}, \tag{10}$$
$$\boldsymbol{\lambda}_{\mathbf{max}}(P_k) \leq \max\{\frac{1-\lambda}{\alpha}, \boldsymbol{\lambda}_{\mathbf{max}}(P_0)\}. \tag{11}$$

(2) If $(\phi_k)_{k=0}^\infty$ is bounded with upper bound $\beta \in (0,\infty)$, then, for all $k \geq 0$,

$$\boldsymbol{\lambda}_{\mathbf{max}}(R_k) \leq \max\{\frac{\beta}{1-\lambda}, \boldsymbol{\lambda}_{\mathbf{max}}(R_0)\}, \tag{12}$$
$$\boldsymbol{\lambda}_{\mathbf{min}}(P_k) \geq \min\{\frac{1-\lambda}{\beta}, \boldsymbol{\lambda}_{\mathbf{min}}(P_0)\}. \tag{13}$$

Proof. We first prove *(1)* by induction. For brevity, let $a \triangleq \min\{\frac{\alpha}{1-\lambda}, \boldsymbol{\lambda}_{\mathbf{min}}(R_0)\}$. For the base case, note that $\boldsymbol{\lambda}_{\mathbf{min}}(R_0) \geq a$ immediately holds. Next, assume for inductive hypothesis that $\boldsymbol{\lambda}_{\mathbf{min}}(R_k) \geq a$. It follows from (7), Lemma A.3, and persistent excitation of $(\phi_k)_{k=0}^\infty$ that $\boldsymbol{\lambda}_{\mathbf{min}}(R_{k+1}) \geq \lambda a + \alpha$. Substituting $a = \min\{\frac{\alpha}{1-\lambda}, \boldsymbol{\lambda}_{\mathbf{min}}(R_0)\}$ yields $\boldsymbol{\lambda}_{\mathbf{min}}(R_{k+1}) \geq \min\{\frac{\alpha}{1-\lambda}, \lambda\boldsymbol{\lambda}_{\mathbf{min}}(R_0) + \alpha\}I_n$.

In the case where $\boldsymbol{\lambda}_{\mathbf{min}}(R_0) \geq \frac{\alpha}{1-\lambda}$, it follows that $\lambda\boldsymbol{\lambda}_{\mathbf{min}}(R_0) + \alpha \geq \lambda(\frac{\alpha}{1-\lambda}) + \alpha = \frac{\alpha}{1-\lambda} \geq a$. In the case where $\boldsymbol{\lambda}_{\mathbf{min}}(R_0) < \frac{\alpha}{1-\lambda}$, it follows that $\lambda\boldsymbol{\lambda}_{\mathbf{min}}(R_0) + \alpha \geq \boldsymbol{\lambda}_{\mathbf{min}}(R_0) \geq a$. Hence, $R_{k+1} \succeq aI_n$, and (10) holds by mathematical induction. Lastly (11) follows directly from (10) since, for all $k \geq 0$, $P_k = R_k^{-1}$. Statement *(2)* can be proven using similar reasoning. ■

3 Subspace Decomposition of a Positive-Definite Matrix

This section presents a novel decomposition of a positive-definite matrix $A \in \mathbb{R}^{n\times n}$ into the sum of two positive-semidefinite matrices based on the choice of a subspace $S \subset \mathbb{R}^n$. This decomposition will be the main tool used in SIFt-RLS. Note that this decomposition reduces to the decomposition proposed in [5] when subspace S is dimension 1. To begin, Definition 3 defines the proposed matrix decomposition.

Definition 3 *Let $S \subset \mathbb{R}^n$ be a subspace of dimension $p \leq n$ and let $A \in \mathbb{R}^{n\times n}$ be positive definite. Let $v_1, \ldots, v_p \in \mathbb{R}^n$ be a basis for S and $v \triangleq \begin{bmatrix} v_1 & \cdots & v_p \end{bmatrix} \in \mathbb{R}^{n\times p}$. Then, define the operators $\|$ and $\perp$ as*

$$A^{\|S} \triangleq Av(v^T Av)^{-1}v^T A \in \mathbb{R}^{n\times n}, \tag{14}$$
$$A^{\perp S} \triangleq A - A^{\|S} \in \mathbb{R}^{n\times n}, \tag{15}$$

where, by Lemma A.7, $v^{\mathrm{T}}Av \in \mathbb{R}^{p\times p}$ is nonsingular.

We call $A^{\|S}$ "A parallel to S" and $A^{\perp S}$ "A orthogonal to S". We call the equality $A = A^{\|S} + A^{\perp S}$ the "S subspace decomposition of A". Next, Proposition 3 shows that the matrices $A^{\|S}$ and $A^{\perp S}$ do not depend on the choice of basis for S.

Proposition 3 *Let $S \subset \mathbb{R}^n$ be a subspace of dimension $p \leq n$ and let $A \in \mathbb{R}^{n\times n}$ be positive definite. Let $v_1, \cdots, v_p \in \mathbb{R}^n$ and $w_1, \cdots, w_p \in \mathbb{R}^n$ be a bases for S and $v \triangleq \begin{bmatrix} v_1 & \cdots & v_p \end{bmatrix} \in \mathbb{R}^{n\times p}$, $w \triangleq \begin{bmatrix} w_1 & \cdots & w_p \end{bmatrix} \in \mathbb{R}^{n\times p}$. Then,*

$$Av(v^{\mathrm{T}}Av)^{-1}v^{\mathrm{T}}A = Aw(w^{\mathrm{T}}Aw)^{-1}w^{\mathrm{T}}A. \tag{16}$$

Proof. Since v and w are both bases for S, there exists invertible $P \in \mathbb{R}^{p\times p}$ such that $v = wP$. Then, (16) follows from substituting $v = wP$ in $Av(v^{\mathrm{T}}Av)^{-1}v^{\mathrm{T}}A$. ■

3.1 Properties

Theorem 1 now gives the main properties of the matrices $A^{\|S}$ and $A^{\perp S}$ in the S subspace decomposition of A.

Theorem 1 (Properties of $A^{\|S}$ and $A^{\perp S}$) *Let $S \subset \mathbb{R}^n$ be a subspace of dimension $p \le n$ and let $A \in \mathbb{R}^{n\times n}$ be positive definite. Then, A can be decomposed into*

$$A = A^{\|S} + A^{\perp S}, \tag{17}$$

where $A^{\|S} \in \mathbb{R}^{n\times n}$, given by (14)*, satisfies the properties:*

(1a) $A^{\|S}$ *is positive semidefinite,*
(2a) $\operatorname{rank}(A^{\|S}) = p$,
(3a) For all $x \in S$, $A^{\|S}x = Ax$.

and $A^{\perp S} \in \mathbb{R}^{n\times n}$, given by (15)*, satisfies the properties:*

(1b) $A^{\perp S}$ *is positive semidefinite,*
(2b) $\operatorname{rank}(A^{\perp S}) = n - p$,
(3b) For all $x \in S$, $A^{\perp S}x = 0$.

Proof. See Appendix Section B. ■

Note that properties *2a)* and *3a)* imply that $\mathcal{R}(A^{\|S}) = AS$ and properties *2b)* and *3b)* imply that $\mathcal{N}(A^{\perp S}) = S$.

Next, Theorem 2 shows that $A^{\|S}$ is the unique matrix satisfying properties *(1a*)*, *(2a)*, and *(3a)* of Theorem 2. Note that since property *(1a)* of Theorem 1 implies property *(1a*)*, $A^{\|S}$ is also the unique matrix satisfying properties *(1a)*, *(2a)*, and *(3a)* of Theorem 1.

Theorem 2 (Uniqueness of $A^{\|S}$) *Let $S \subset \mathbb{R}^n$ be a subspace of dimension $p \le n$ and $A \in \mathbb{R}^{n\times n}$ be positive definite. If $\tilde{A} \in \mathbb{R}^{n\times n}$ satisfies the properties*

(1a)* $\tilde{A}$ *is symmetric,*
(2a) $\operatorname{rank}(\tilde{A}) = p$,
(3a) For all $x \in S$, $\tilde{A}x = Ax$,

then $\tilde{A} = A^{\|S}$.

Proof. See Appendix Section B. ■

Next, Definition 4 defines the $\langle\cdot,\cdot\rangle_A$ inner product for positive-definite matrix A and Definition 5 defines $S^{\perp_A}$, the orthogonal complement of subspace S under this inner product. Finally, Proposition 4 shows a duality between the $\|$ and $\perp$ operators and the subspaces S and $S^{\perp_A}$.

Definition 4 *Let $A \in \mathbb{R}^{n\times n}$ be positive definite. Define the inner product $\langle\cdot,\cdot\rangle_A \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ as*

$$\langle x, y\rangle_A \triangleq x^{\mathrm{T}}Ay. \tag{18}$$

Definition 5 *Let $S \subset \mathbb{R}^n$ be a subspace and let $A \in \mathbb{R}^{n\times n}$ be positive definite. Define the orthogonal complement of S under the inner product $\langle\cdot,\cdot\rangle_A$ as*

$$S^{\perp_A} \triangleq \{x \in \mathbb{R}^m : \text{for all } y \in S,\ \langle x, y\rangle_A = 0\}. \tag{19}$$

Proposition 4 (Duality) *Let $S \subset \mathbb{R}^n$ be a subspace of dimension $p \le n$ and let $A \in \mathbb{R}^{n\times n}$ be positive definite. Then,*

$$A^{\|S} = A^{\perp(S^{\perp_A})}, \tag{20}$$

$$A^{\perp S} = A^{\|(S^{\perp_A})}. \tag{21}$$

Proof. See Appendix Section B. ■

4 SIFt-RLS

To motivate this algorithm, suppose there exists true parameters $\theta \in \mathbb{R}^n$ such that, for all $k \ge 0$, $y_k = \phi_k\theta$. Then, at any $k \ge 0$, knowledge of the regressor $\phi_k \in \mathbb{R}^{p\times n}$ and measurement $y_k \in \mathbb{R}^p$ only provide information about the true parameters in the row space of ϕ_k. For example, a regressor $\phi_k = [1\ 1\ 0]$ only informs about the sum of the first two parameters of $\theta \in \mathbb{R}^3$. We call the row space of ϕ_k, i.e. $\mathcal{R}(\phi_k^{\mathrm{T}})$, the *information subspace*.

Note that in (7), R_k is multiplied by $\lambda \in (0, 1)$, which can be interpreted as forgetting uniformly over all directions. Covariance windup in exponential forgetting RLS occurs when the information subspace does not contain particular directions over a large number of time steps, while forgetting is uniform. As a result, persistent excitation conditions are often used to guarantee excitation in all directions.

We propose subspace of information forgetting recursive least squares, or SIFt-RLS, a directional forgetting algorithm which, at each step $k \ge 0$, only forgets in the information subspace of that step. To summarize the algorithm, at each step $k \ge 0$, SIFt-RLS involves three parts:

1. *Information Filtering*: First, we take a compact singular value decomposition (SVD) of the regressor $\phi_k \in \mathbb{R}^{p\times n}$ to truncate any singular values smaller than a desired threshold. This gives the filtered regressor $\bar{\phi}_k$ and filtered measurement $\bar{y}_k$. This step is critical to ensure the algorithm is numerically stable.

2. *Subspace of Information Forgetting (SIFTing)*: Next, we decompose the information matrix R_k using the subspace decomposition presented in section 3 with the information subspace $\mathcal{R}(\bar{\phi}_k^{\mathrm{T}})$. Forgetting is applied only to the component of R_k parallel to the information subspace.
3. *Update*: Lastly, we compute the updated information matrix R_{k+1} and parameter estimate θ_{k+1} using the filtered regressor and filtered measurement.

Furthermore, we also provide a way to directly update the covariance matrix $P_k \triangleq R_k^{-1}$ in steps 2 and 3 using the matrix inversion Lemma, which is more computationally efficient when $p \ll n$. The next three subsections provide details on the three parts of SIFt-RLS. The algorithm is initialized with forgetting factor $\lambda \in (0,1)$, tuning parameter $\varepsilon > 0$, initial estimate of parameters $\theta_0 \in \mathbb{R}^n$, initial positive-definite information matrix $R_0 \in \mathbb{R}^{n\times n}$, and initial covariance matrix P_0^{-1}.

4.1 Information Filtering

To begin, let $\varepsilon > 0$ be a tuning parameter which, for all $k \geq 0$, will be used to truncate small singular values of ϕ_k. Next, let $\phi_k = U_k \Sigma_k V_k^{\mathrm{T}}$ be the compact SVD of ϕ_k with the singular values on the diagonal of Σ_k in descending order. [1] [2] [3] Let q_k be the number of singular values of ϕ_k greater or equal to $\sqrt{\varepsilon}$ and define $\bar{U}_k \in \mathbb{R}^{p\times q_k}$ as the first q_k columns of U_k. Then, define the filtered regressor, $\bar{\phi}_k \in \mathbb{R}^{q_k\times n}$, and filtered measurement $\bar{y}_k \in \mathbb{R}^{q_k}$ by

$$\bar{\phi}_k \triangleq \bar{U}_k^{\mathrm{T}} \phi_k, \tag{22}$$
$$\bar{y}_k \triangleq \bar{U}_k^{\mathrm{T}} y_k. \tag{23}$$

Note that, by construction, all singular values of $\bar{\phi}_k$ are greater or equal to $\sqrt{\varepsilon}$.

4.2 Subspace of Information Forgetting (SIFting)

For all $k \geq 0$, define $R_k^{\parallel} \in \mathbb{R}^{n\times n}$ and $R_k^{\perp} \in \mathbb{R}^{n\times n}$ as

$$R_k^{\parallel} \triangleq R_k^{\parallel \mathcal{R}(\bar{\phi}_k^{\mathrm{T}})} = R_k \bar{\phi}_k^{\mathrm{T}} (\bar{\phi}_k R_k \bar{\phi}_k^{\mathrm{T}})^{-1} \bar{\phi}_k R_k, \tag{24}$$
$$R_k^{\perp} \triangleq R_k^{\perp \mathcal{R}(\bar{\phi}_k^{\mathrm{T}})} = R_k - R_k \bar{\phi}_k^{\mathrm{T}} (\bar{\phi}_k R_k \bar{\phi}_k^{\mathrm{T}})^{-1} \bar{\phi}_k R_k, \tag{25}$$

where the operators $\parallel$ and $\perp$ are defined in Definition 3 and $\mathcal{R}(\bar{\phi}_k^{\mathrm{T}})$ is the information subspace. It will later be shown in Corollary 2 that the matrix $\bar{\phi}_k R_k \bar{\phi}_k^{\mathrm{T}}$ is nonsingular and, under mild conditions, is well-conditioned. Hence, for all $k \geq 0$, the information subspace decomposition of R_k can be expressed as

$$R_k = R_k^{\parallel} + R_k^{\perp}. \tag{26}$$

Next, we forget in only the information subspace to get the *SIFted information matrix* $\bar{R}_k \in \mathbb{R}^{n\times n}$, defined as

$$\begin{aligned} \bar{R}_k &\triangleq \lambda R_k^{\parallel} + R_k^{\perp} \\ &= R_k - (1-\lambda) R_k \bar{\phi}_k^{\mathrm{T}} (\bar{\phi}_k R_k \bar{\phi}_k^{\mathrm{T}})^{-1} \bar{\phi}_k R_k. \end{aligned} \tag{27}$$

Furthermore, we define the *SIFted covariance matrix* as $\bar{P}_k \triangleq \bar{R}_k^{-1} \in \mathbb{R}^{n\times n}$. Note that it will later be shown in Corollary 1 that $\bar{R}_k$ is positive definite, and hence nonsingular. The matrix inversion lemma (Lemma A.1) can also be used to express $\bar{P}_k$ as

$$\bar{P}_k = P_k + \frac{1-\lambda}{\lambda} \bar{\phi}_k^{\mathrm{T}} (\bar{\phi}_k R_k \bar{\phi}_k^{\mathrm{T}})^{-1} \bar{\phi}_k. \tag{28}$$

4.3 Update

Lastly, for all $k \geq 0$, we compute the updated information matrix, $R_{k+1} \in \mathbb{R}^{n\times n}$, and updated parameter vector estimate, $\theta_{k+1} \in \mathbb{R}^n$ as

$$R_{k+1} = \bar{R}_k + \bar{\phi}_k^{\mathrm{T}} \bar{\phi}_k, \tag{29}$$
$$\theta_{k+1} = \theta_k + P_{k+1} \bar{\phi}_k^{\mathrm{T}} (\bar{y}_k - \bar{\phi}_k \theta_k), \tag{30}$$

where $P_{k+1} \triangleq R_{k+1}^{-1} \in \mathbb{R}^{n\times n}$. It will later be shown in Corollary 1 that R_{k+1} is positive definite, and hence nonsingular. Furthermore, note that the matrix inversion Lemma (Lemma A.1) can be used to express P_{k+1} as

$$P_{k+1} = \bar{P}_k - \bar{P}_k \bar{\phi}_k^{\mathrm{T}} (I_{q_k} + \bar{\phi}_k \bar{P}_k \bar{\phi}_k^{\mathrm{T}})^{-1} \bar{\phi}_k \bar{P}_k. \tag{31}$$

4.4 Implementation and Computational Cost

This subsection summarizes some subtleties in implementing SIFt-RLS which minimize computational cost and ensure numerical stability. We also analyze the computational complexity of SIFT-RLS and discuss when it is or is not efficient to use the matrix inversion Lemma to update the covariance matrix.

To begin our discussion on the computational cost of SIFt-RLS, note that, for all $k \geq 0$, the computational complexity of the compact SVD of ϕ_k is $\mathcal{O}(\min\{pn^2, p^2 n\})$ (See Lecture 31 of [33]). Next, note that $\bar{R}_k$, defined in (27), can be computed in $\mathcal{O}(q_k n^2)$ time complexity, where $q_k \leq \min\{p, n\}$ be the number of singular values of ϕ_k larger than $\sqrt{\varepsilon}$. This is

[1] If $p \leq n$, then $U_k \in \mathbb{R}^{p\times p}$, $\Sigma_k \in \mathbb{R}^{p\times p}$, and $V_k \in \mathbb{R}^{n\times p}$. If $p > n$, then $U_k \in \mathbb{R}^{p\times n}$, $\Sigma_k \in \mathbb{R}^{n\times n}$, and $V_k \in \mathbb{R}^{n\times n}$.

[2] This is more computationally efficient than computing the full SVD of ϕ_k. The compact SVD can, for example, be computed in MATLAB as `[U,S,V] = svd(phi,"econ")`.

[3] Note that, for computational efficiency, the right singular vectors V_k do not need to be computed.

done by first computing $L_k \triangleq \bar{\phi}_k R_k$, then computing $\bar{R}_k = R_k - (1-\lambda) L_k^{\mathrm{T}} (L_k \bar{\phi}_k^{\mathrm{T}})^{-1} L_k$.[4] A concern in practice is that the calculation of $\bar{R}_k$ over many steps may cause $\bar{R}_k$ to drift from symmetric due to round-off errors, resulting in numerical instability. A simple solution is to recompute $\bar{R}_k \leftarrow \frac{1}{2}(\bar{R}_k + \bar{R}_k^{\mathrm{T}})$ at each step.

Furthermore, note that, for all $k \geq 0$, the updated parameter estimate θ_{k+1}, given by (30) is most efficiently computed with the ordering $\theta_{k+1} = \theta_k + P_{k+1}[\bar{\phi}_k^{\mathrm{T}}(\bar{y}_k - \bar{\phi}_k \theta_k)]$. This computation requires the updated covariance matrix P_{k+1}, which can either be computed by first computing $\bar{P}_k$ using (28) then computing P_{k+1} using (31), or by directly inverting R_{k+1}.

Using the matrix inversion lemma, the matrix $\bar{P}_k$ can be computed in $\mathcal{O}(q_k n^2)$ as $\bar{P}_k = P_k + \frac{1-\lambda}{\lambda} \bar{\phi}_k^{\mathrm{T}} (L_k \bar{\phi}_k^{\mathrm{T}})^{-1} \bar{\phi}_k$. Then, P_{k+1} can be computed in $\mathcal{O}(q_k n^2)$ by first computing $M_k \triangleq \bar{\phi}_k \bar{P}_k$, then computing $P_{k+1} = \bar{P}_k - M_k^{\mathrm{T}}(I_{q_k} + M_k \bar{\phi}_k^{\mathrm{T}})^{-1} M_k$.[5] Similarly, P_{k+1} may drift from symmetric when computed using the matrix inversion Lemma due to round-off errors. A solution is to recompute $P_{k+1} \leftarrow \frac{1}{2}(P_{k+1} + P_{k+1}^{\mathrm{T}})$ at each step. In contrast, the computational complexity to compute P_{k+1} by directly inverting R_{k+1} is $\mathcal{O}(n^3)$.

In practice, however, it is only faster to compute P_{k+1} using the matrix inversion lemma via (28) and (31) if $q_k \ll n$. Note that $q_k \in \{0, 1, \ldots, \min\{p, n\}\}$ may change at each step but n remains constant. Hence, we suggest to select $q_{\max} \in \{1, 2, \ldots, \min\{p, n\}\}$ such that if $q_k \leq q_{\max}$, P_{k+1} is computed using (28) and (31), otherwise P_{k+1} is computed by directly inverting R_{k+1}. Figure 1 gives data on which method is faster for various values of q_k and n as well as suggest values of $q_{\max}$ as a function of n. It is important to note that the choice of $q_{\max}$ only affects computational cost, not the algorithm itself. Alternatively, for simpler implementation, one can select $q_{\max} = 0$ to always perform direct inversion or $q_{\max} = p$ to always use the matrix inversion lemma.

To summarize, the algorithm SIFt-RLS can be performed in $\max\{\min\{pn^2, p^2 n\}, q_k n^2\}$ computational complexity per step. While no assumptions are made whether $p < n$ or $p \geq n$, SIFt-RLS is best used when $p < n$. In the case where $p < n$, the computational complexity of SIFt-RLS is $\max\{p^2 n, q_k n^2\}$ per step. Note that if $p \geq n$ and, at step $k \geq 0$, all singular values of ϕ_k are greater or equal to $\sqrt{\varepsilon}$, then the information subspace is all of $\mathbb{R}^n$ and SIFt-RLS simplifies to exponential forgetting RLS, as shown in Proposition 5. However, SIFt-RLS can still be still be useful in the case $p \geq n$ if it is expected that ϕ_k will have one or more singular values smaller than $\sqrt{\varepsilon}$.

[4] Computing L_k is important because standard left to right multiplication of (27) results in $\mathcal{O}(n^3)$ time complexity.

[5] First computing M_k is important because standard left to right matrix multiplication of (31) results in $\mathcal{O}(n^3)$ time complexity.

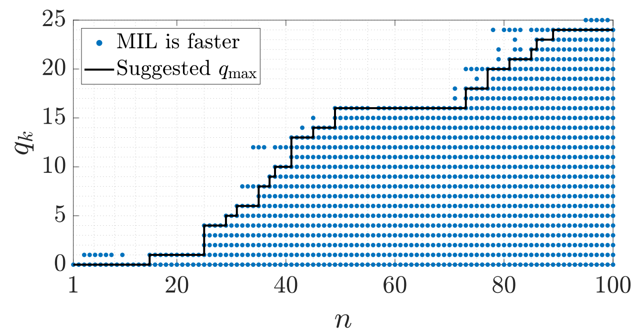

Fig. 1. Comparison of computation time to compute P_{k+1} in SIFt-RLS, tested in MATLAB on an i7-6600U processor with 16 GB of RAM. Blue x indicates that, for the given values of q_k and n, it is faster, on average, to compute $P_{k+1} \in \mathbb{R}^{n \times n}$ using the matrix inversion lemma via (28) and (31) than by direct inversion of R_{k+1}. However, further testing has shown that these results may differ greatly between different machines.

Proposition 5 *Let $k \geq 0$, let $\phi_k \in \mathbb{R}^{p \times n}$, $y_k \in \mathbb{R}^p$, $\theta_k \in \mathbb{R}^n$, and let $R_k \in \mathbb{R}^{n \times n}$ be positive definite. Also let $\varepsilon > 0$. Let $\theta_{k+1} \in \mathbb{R}^n$ and $R_{k+1} \in \mathbb{R}^{n \times n}$ be computed using (22), (23), (27), (29), and (30). If $p \geq n$ and all singular values of ϕ_k are greater or equal to $\sqrt{\varepsilon}$, then*

$$R_{k+1} = \lambda R_k + \phi_k^{\mathrm{T}} \phi_k, \tag{32}$$

$$\theta_{k+1} = \theta_k + P_{k+1} \phi_k^{\mathrm{T}} (y_k - \phi_k \theta_k). \tag{33}$$

Proof. Let $\phi_k = U_k \Sigma_k V_k^{\mathrm{T}}$ be the compact SVD of ϕ_k, where $U_k \in \mathbb{R}^{p \times n}$, $\Sigma_k \in \mathbb{R}^{n \times n}$, and $V_k \in \mathbb{R}^{n \times n}$. Since $p \geq n$ and all singular values of ϕ_k are greater or equal to $\sqrt{\varepsilon}$, it follows that $q_k = n$ and $\bar{U}_k = U_k$. Hence, $\bar{\phi}_k = U_k^{\mathrm{T}} \phi_k \in \mathbb{R}^{n \times n}$ and $\bar{y}_k = U_k^{\mathrm{T}} y_k \in \mathbb{R}^n$. Moreover, since U_k is semi-orthogonal, $\operatorname{rank}(\bar{\phi}_k) = n$, hence $\bar{\phi}_k$ is nonsingular. Therefore, it follows that $(\bar{\phi}_k R_k \bar{\phi}_k^{\mathrm{T}})^{-1} = \bar{\phi}_k^{-\mathrm{T}} P_k \bar{\phi}_k^{-1}$. Substituting this equality into (27), it follows that $\bar{R}_k = \lambda R_k$. Furthermore, since U_k is semi-orthogonal, it follows that $\bar{\phi}_k^{\mathrm{T}} \bar{\phi}_k = \phi_k^{\mathrm{T}} \phi_k$ and $\bar{\phi}_k^{\mathrm{T}} \bar{y}_k = \phi_k^{\mathrm{T}} y_k$. Substituting into (29) and (30), (32) and (33) follow. ■

The implementation details discussed in this subsection are summarized in Algorithm 1. Some recommendations for tuning parameters $\lambda \in (0,1)$, $\varepsilon > 0$, $\theta_0 \in \mathbb{R}^n$, and positive-definite $P_0 \in \mathbb{R}^{n \times n}$ are as follows. θ_0 and P_0 are a prior belief of the mean and covariance, respectively, of the parameters being estimated and can be tuned similarly as they would in other extensions of RLS. ε can be determined based on the noise level of the data but should be small to avoid overly truncating the data. Finally, while a forgetting factor close to 1 is traditionally recommended to avoid covariance windup [4, 26, 27], we have found $\lambda \in (0,1)$ can be tuned more aggressively (more small) in SIFt-RLS to track quickly changing parameters since forgetting only occurs in the information subspace. This is particularly useful if there is ample excitation in certain directions, allowing for identification of a subspace of parameters, while there is little excitation in other directions.

Algorithm 1 Subspace of Information Forgetting Recursive Least Squares (SIFt-RLS)

Tuning Parameters: $\lambda \in (0, 1)$, $\varepsilon > 0$, $\theta_0 \in \mathbb{R}^n$, positive-definite $R_0 \in \mathbb{R}^{n\times n}$ and $P_0 \triangleq R_0^{-1}$.
Computational Parameters: $q_{\max} \in [1, \min\{p, n\}]$.

for all $k \geq 0$ **do**
  <u>**1. Information Filtering:**</u>
  Compute the compact SVD $\phi_k = U_k \Sigma_k V_k^{\mathrm{T}}$ with the singular values on the diagonal of Σ_k in descending order. Let q_k be the number of singular values of ϕ_k greater or equal to $\sqrt{\varepsilon}$.
  if $q_k = 0$ **then**
    $R_{k+1} \leftarrow R_k$, $\theta_{k+1} \leftarrow \theta_k$, skip to step $k+1$.
  Let $\bar{U}_k$ be the first q_k columns of U_k.
  $\bar{\phi}_k \in \mathbb{R}^{q_k \times n} \leftarrow \bar{U}_k^{\mathrm{T}} \phi_k$
  $\bar{y}_k \in \mathbb{R}^{q_k} \leftarrow \bar{U}_k^{\mathrm{T}} y_k$
  <u>**2. SIFting:**</u>
  $L_k \leftarrow \bar{\phi}_k R_k$
  $\bar{R}_k \leftarrow R_k - (1-\lambda) L_k^{\mathrm{T}} (L_k \bar{\phi}_k^{\mathrm{T}})^{-1} L_k$
  $\bar{R}_k \leftarrow \frac{1}{2}(\bar{R}_k + \bar{R}_k^{\mathrm{T}})$
  if $q_k \leq q_{\max}$ **then**
    $\bar{P}_k \leftarrow P_k + \frac{1-\lambda}{\lambda} \bar{\phi}_k^{\mathrm{T}} (L_k \bar{\phi}_k^{\mathrm{T}})^{-1} \bar{\phi}_k$.
  <u>**3. Update:**</u>
  $R_{k+1} \leftarrow \bar{R}_k + \bar{\phi}_k^{\mathrm{T}} \bar{\phi}_k$
  if $q_k \leq q_{\max}$ **then**
    $M_k \leftarrow \bar{\phi}_k \bar{P}_k$
    $P_{k+1} \leftarrow \bar{P}_k - M_k^{\mathrm{T}} (I_{q_k} + M_k \bar{\phi}_k^{\mathrm{T}})^{-1} M_k$
    $P_{k+1} \leftarrow \frac{1}{2}(P_{k+1} + P_{k+1}^{\mathrm{T}})$
  else
    $P_{k+1} \leftarrow R_{k+1}^{-1}$
  $\theta_{k+1} \leftarrow \theta_k + P_{k+1}[\bar{\phi}_k^{\mathrm{T}}(\bar{y}_k - \bar{\phi}_k \theta_k)]$

5 Covariance Bounds and Stability of SIFt-RLS

This section discusses the theoretical guarantees of SIFt-RLS including bounds on the covariance matrix and estimation error stability assuming estimation of fixed parameters without noise.

5.1 Covariance Bounds

To begin, we show that, without any assumptions of persistent excitation, the eigenvalues of the information matrix R_k in SIFt-RLS are lower bounded. Furthermore, we show that if the sequence of regressors $(\phi_k)_{k=0}^{\infty}$ is upper bounded, then the eigenvalues of the information matrix R_k in SIFt-RLS are also upper bounded. This result also immediately implies that the reciprocal expressions are bounds for the covariance matrix $P_k = R_k^{-1}$. These results are given in Theorem 3. Note that explicit bounds are given which improves upon the works of [5] and [35] which only state the existence of bounds. Moreover, the bounds of Theorem 3 are stronger than those of [35], see Appendix D for details.

Theorem 3 *Let $\lambda \in (0, 1)$, let $\varepsilon > 0$, and let $R_0 \in \mathbb{R}^{n\times n}$ be positive definite. For all $k \geq 0$, let $\phi_k \in \mathbb{R}^{p\times n}$, let $\bar{\phi}_k \in \mathbb{R}^{q_k \times n}$ be defined by (22), let $\bar{R}_k \in \mathbb{R}^{n\times n}$ be defined by (27), and let $R_{k+1} \in \mathbb{R}^{n\times n}$ be defined by (29). Then the following statements hold:*

(1) For all $k \geq 0$,

$$\boldsymbol{\lambda}_{\mathbf{min}}(R_k) \geq \min\{\frac{\varepsilon}{1-\lambda}, \boldsymbol{\lambda}_{\mathbf{min}}(R_0)\}, \tag{34}$$

$$\boldsymbol{\lambda}_{\mathbf{max}}(P_k) \leq \max\{\frac{1-\lambda}{\varepsilon}, \boldsymbol{\lambda}_{\mathbf{max}}(P_0)\}. \tag{35}$$

(2) If $(\phi_k)_{k=0}^{\infty}$ is bounded with upper bound $\beta \in (0, \infty)$, then, for all $k \geq 0$,

$$\boldsymbol{\lambda}_{\mathbf{max}}(R_k) \leq \max\{\frac{\beta}{1-\lambda}, \boldsymbol{\lambda}_{\mathbf{max}}(R_0)\}, \tag{36}$$

$$\boldsymbol{\lambda}_{\mathbf{min}}(P_k) \geq \min\{\frac{1-\lambda}{\beta}, \boldsymbol{\lambda}_{\mathbf{min}}(P_0)\}. \tag{37}$$

Proof. See Appendix section C for a proof of (34) and (36). Bounds (35) and (37) follow directly from (34) and (36) respectively since, for all $k \geq 0$, $P_k = R_k^{-1}$. ■

Note that the bounds of Theorem 3 are strikingly analogous to the bounds of RLS with exponential forgetting in Proposition 2, where $(\phi_k)_{k=0}^{\infty}$ is persistently exciting with persistency window $N = 1$. This is a result of the information filtering step of SIFt-RLS truncating the singular values of ϕ_k smaller than $\sqrt{\varepsilon}$, and the SIFting step of SIFt-RLS forgetting only in the information subspace, $\mathcal{R}(\phi_k^{\mathrm{T}})$.

Next, Corollary 1 shows that, for all $k \geq 0$, R_k and $\bar{R}_k$ are positive definite. Furthermore, Corollary 2 shows that, for all $k \geq 0$, the matrix $\bar{\phi}_k R_k \bar{\phi}_k^{\mathrm{T}}$ is nonsingular and, assuming bounded regressors, has a bounded condition number. This is ensures that computing $(\bar{\phi}_k R_k \bar{\phi}_k^{\mathrm{T}})^{-1}$ in the SIFting step is well-conditioned.

Corollary 1 *Consider the notation of Theorem 3. For all $k \geq 0$, R_k and $\bar{R}_k$ are positive definite.*

Proof. Since $\min\{\frac{\varepsilon}{1-\lambda}, \boldsymbol{\lambda}_{\mathbf{min}}(R_0)\} > 0$, it follows from (34) that, for all $k \geq 0$, $R_k \succ 0$. Next, note that, for all $k \geq 0$, $\bar{R}_k$, defined in (27), can be written as $\bar{R}_k = \lambda R_k + (1-\lambda) R_k^{\perp}$. It follows from Theorem 1 that $R_k^{\perp}$ is positive semidefinite, and hence $\bar{R}_k \succeq \lambda R_k \succ 0$. ■

Corollary 2 *Consider the notation of Theorem 3. For all $k \geq 0$, $\bar{\phi}_k R_k \bar{\phi}_k^{\mathrm{T}}$ is nonsingular. Moreover, if $(\phi_k)_{k=0}^{\infty}$ is bounded with upper bound $\beta \in (0,\infty)$, then, for all $k \geq 0$,*

$$\kappa(\bar{\phi}_k R_k \bar{\phi}_k^{\mathrm{T}}) \leq \frac{\beta \max\{\frac{\beta}{1-\lambda}, \boldsymbol{\lambda}_{\mathbf{max}}(R_0)\}}{\varepsilon \min\{\frac{\varepsilon}{1-\lambda}, \boldsymbol{\lambda}_{\mathbf{min}}(R_0)\}}. \tag{38}$$

Proof. It follows from Corollary 1 that R_k is positive definite. Next, since $\operatorname{rank}(\bar{\phi}_k) = q_k$ by construction, it follows from Lemma A.7 that $\bar{\phi}_k R_k \bar{\phi}_k^{\mathrm{T}}$ is positive definite, hence nonsingular.

Next, suppose $(\phi_k)_{k=0}^{\infty}$ is bounded with upper bound $\beta \in (0,\infty)$. Since $\boldsymbol{\sigma}_{\mathbf{max}}(\cdot)$ is submultiplicative, it follows that $\boldsymbol{\sigma}_{\mathbf{max}}(\bar{\phi}_k R_k \bar{\phi}_k^{\mathrm{T}}) \leq \boldsymbol{\sigma}_{\mathbf{max}}(\bar{\phi}_k)\boldsymbol{\sigma}_{\mathbf{max}}(R_k)\boldsymbol{\sigma}_{\mathbf{max}}(\bar{\phi}_k^{\mathrm{T}})$. Furthermore, since $(\phi_k)_{k=0}^{\infty}$ is bounded with upper bound $\beta \in (0,\infty)$, $(\bar{\phi}_k)_{k=0}^{\infty}$ is also bounded with upper bound $\beta \in (0,\infty)$. Therefore, $\boldsymbol{\sigma}_{\mathbf{max}}(\bar{\phi}_k) \leq \sqrt{\beta}$. This combined with (36) imply that $\boldsymbol{\sigma}_{\mathbf{max}}(\bar{\phi}_k R_k \bar{\phi}_k^{\mathrm{T}}) \leq \beta \max\{\frac{\beta}{1-\lambda}, \boldsymbol{\lambda}_{\mathbf{max}}(R_0)\}$.

Next, as a consequence of the min-max theorem given in Lemma A.2, note that

$$\boldsymbol{\sigma}_{\mathbf{min}}(\bar{\phi}_k R_k \bar{\phi}_k^{\mathrm{T}}) = \min_{x \in \mathbb{R}^{q_k}, x \neq 0} \frac{\|\bar{\phi}_k R_k \bar{\phi}_k^{\mathrm{T}} x\|}{\|x\|}.$$

Since $\bar{\phi}_k^{\mathrm{T}}$ has full column rank, $x \neq 0$ implies that $\bar{\phi}_k^{\mathrm{T}} x \neq 0$. Therefore, we can write

$$\begin{aligned}
\min_{x \in \mathbb{R}^{q_k}, x \neq 0} \frac{\|\bar{\phi}_k R_k \bar{\phi}_k^{\mathrm{T}} x\|}{\|x\|} &= \min_{x \in \mathbb{R}^{q_k}, x \neq 0} \frac{\|\bar{\phi}_k R_k \bar{\phi}_k^{\mathrm{T}} x\|}{\|\bar{\phi}_k^{\mathrm{T}} x\|} \frac{\|\bar{\phi}_k^{\mathrm{T}} x\|}{\|x\|} \\
&\geq \min_{y \in \mathbb{R}^{n}, y \neq 0} \frac{\|\bar{\phi}_k R_k y\|}{\|y\|} \min_{x \in \mathbb{R}^{q_k}, x \neq 0} \frac{\|\bar{\phi}_k^{\mathrm{T}} x\|}{\|x\|} \\
&= \boldsymbol{\sigma}_{\mathbf{min}}(\bar{\phi}_k R_k)\boldsymbol{\sigma}_{\mathbf{min}}(\bar{\phi}_k^{\mathrm{T}}).
\end{aligned}$$

Furthermore, since $\bar{\phi}_k$ has full row rank, it can be shown by similar reasoning that $\boldsymbol{\sigma}_{\mathbf{min}}(\bar{\phi}_k R_k) \geq \boldsymbol{\sigma}_{\mathbf{min}}(\bar{\phi}_k)\boldsymbol{\sigma}_{\mathbf{min}}(R_k)$. By construction of $\bar{\phi}_k$, $\boldsymbol{\sigma}_{\mathbf{min}}(\bar{\phi}_k) = \boldsymbol{\sigma}_{\mathbf{min}}(\bar{\phi}_k^{\mathrm{T}}) \geq \sqrt{\varepsilon}$. Combined with (34), it follows that $\boldsymbol{\sigma}_{\mathbf{min}}(\bar{\phi}_k R_k \bar{\phi}_k^{\mathrm{T}}) \geq \varepsilon \boldsymbol{\sigma}_{\mathbf{min}}(R_k) \geq \varepsilon \min\{\frac{\varepsilon}{1-\lambda}, \boldsymbol{\lambda}_{\mathbf{min}}(R_0)\}$.

Substituting the upper bound of $\boldsymbol{\sigma}_{\mathbf{max}}(\bar{\phi}_k R_k \bar{\phi}_k^{\mathrm{T}})$ and lower bound of $\boldsymbol{\sigma}_{\mathbf{min}}(\bar{\phi}_k R_k \bar{\phi}_k^{\mathrm{T}})$ into definition (2) yields (38). ■

5.2 Stability

Next, for the analysis of this section, we make the assumption that there exist fixed parameters $\theta \in \mathbb{R}^n$ such that, for all $k \geq 0$,

$$y_k = \phi_k \theta. \tag{39}$$

Furthermore, for all $k \geq 0$, we define the parameter estimation error $\tilde{\theta}_k \in \mathbb{R}^n$ by

$$\tilde{\theta}_k \triangleq \theta_k - \theta. \tag{40}$$

Substituting (40) into (30), it then follows that, for all $k \geq 0$,

$$\tilde{\theta}_{k+1} = (I_n - P_{k+1}\bar{\phi}_k^{\mathrm{T}}\bar{\phi}_k)\tilde{\theta}_k. \tag{41}$$

Note that (41) is a linear time-varying system with an equilibrium point $\tilde{\theta}_k \equiv 0$. Thus, we can study the stability of the estimation error equilibrium point $\tilde{\theta}_k \equiv 0$. Theorem 4 uses the results of [18] to show that this the equilibrium point $\tilde{\theta}_k \equiv 0$ is uniformly asymptotically stable without further assumptions and globally uniformly exponentially stable under persistent excitation. For definitions of uniform asymptotic stability and global uniform exponential stability, see Definition 13.7 in [12, pp. 783, 784].

Theorem 4 *Let $\lambda \in (0,1)$, let $\varepsilon > 0$, let $R_0 \in \mathbb{R}^{n\times n}$ be positive definite, and let $\theta_0 \in \mathbb{R}^n$. For all $k \geq 0$, let $\phi_k \in \mathbb{R}^{p\times n}$, let $\bar{\phi}_k \in \mathbb{R}^{q_k\times n}$ be defined by (22), let $\bar{R}_k \in \mathbb{R}^{n\times n}$ be defined by (27), let $R_{k+1} \in \mathbb{R}^{n\times n}$ be defined by (29), and let $\theta_{k+1} \in \mathbb{R}^n$ be defined by (30). Finally, assume there exists $\theta \in \mathbb{R}^n$ such that, for all $k \geq 0$, (39) holds. Then, the following two statements hold:*

(1) The equilibrium $\tilde{\theta}_k \equiv 0$ of (41) is uniformly Lyapunov stable.
(2) If $\{\bar{\phi}_k\}_{k=0}^{\infty}$ is persistently exciting, then the equilibrium $\tilde{\theta}_k \equiv 0$ of (41) is globally uniformly exponentially stable.

Proof. This result is based on Theorem 2 of [18]. Note that, for all $k \geq 0$, (29) can be written as

$$R_{k+1} = R_k - F_k + \bar{\phi}_k^{\mathrm{T}}\bar{\phi}_k,$$

where $F_k \triangleq (1-\lambda)R_k^{\parallel}$. Note that, for all $k \geq 0$, $1-\lambda > 0$ and that by Theorem 1, $R_k^{\parallel} \succeq 0$, hence $F_k \succeq 0$. Therefore, condition *A1)* of [18] is satisfied.

Next, note that, for all $k \geq 0$, $R_k - F_k = \lambda R_k^{\parallel} + R_k^{\perp}$. Then, since $R_k^{\perp} \succeq 0$ by Theorem 1, it follows that

$$R_k - F_k \succeq \lambda R_k^{\parallel} + \lambda R_k^{\perp} = \lambda R_k.$$

Furthermore, it follows from (34) of Theorem 3 that, for all $k \geq 0$,

$$R_k - F_k \succeq \lambda \min\{\frac{\varepsilon}{1-\lambda}, \boldsymbol{\lambda}_{\mathbf{min}}(R_0)\} I_n,$$

and hence

$$(R_k - F_k)^{-1} \preceq \frac{1}{\lambda} \max\{\frac{1-\lambda}{\varepsilon}, \boldsymbol{\lambda}_{\mathbf{max}}(P_0)\} I_n.$$

Therefore, condition *A2)* of [18] is satisfied. Finally, (36) of Theorem 3 implies that, for all $k \geq 0$,

$$\min\{\frac{1-\lambda}{\beta}, \boldsymbol{\lambda}_{\mathbf{min}}(P_0)\} I_n \preceq P_k.$$

Hence, condition *A2)* of [18] is satisfied. Thus, by statement 2) of Theorem 2 of [18], the equilibrium $\tilde{\theta}_k \equiv 0$ of (41) is uniformly Lyapunov stable.

Moreover, if $\{\bar{\phi}_k\}_{k=0}^{\infty}$ is persistently exciting, then condition *A4)* of [18] is satisfied. It then follows from statement 4) of Theorem 2 of [18], the equilibrium $\tilde{\theta}_k \equiv 0$ of (41) is globally uniformly exponentially stable. ∎

6 Numerical Example

We consider the follow example of identifying $n = 4$ parameters with $p = 2$ measurements at each step. For all $0 \leq k \leq 1200$, consider the true parameters $\theta_{\mathrm{true},k} \in \mathbb{R}^4$, defined

$$\begin{aligned}\theta_{\mathrm{true},k} &\triangleq \begin{bmatrix}\theta^1_{\mathrm{true},k} & \theta^2_{\mathrm{true},k} & \theta^3_{\mathrm{true},k} & \theta^4_{\mathrm{true},k}\end{bmatrix}^{\mathrm{T}} \\ &= \begin{bmatrix}\sin(\frac{\pi}{60}k) & \cos(\frac{k}{60}) & \sin(\frac{k}{225}) & \cos(\frac{k}{225})\end{bmatrix}^{\mathrm{T}}. \quad (42)\end{aligned}$$

For all $0 \leq k < 400$, let both rows of $\phi_k \in \mathbb{R}^{2\times 4}$ be i.i.d. sampled from $\mathcal{N}(0, \mathrm{diag}(1, 1, 10^{-4}, 10^{-4}))$ and, for all $400 \leq k < 800$, let both rows of $\phi_k \in \mathbb{R}^{2\times 4}$ be i.i.d. sampled from $\mathcal{N}(0, \mathrm{diag}(10^{-4}, 10^{-4}, 1, 1))$. Lastly, for all $800 \leq k \leq 1200$, let $\phi_k = \sigma_k[0\ 2\ 1\ 0] + \nu_k$ where σ_k is i.i.d. sampled from $\mathcal{N}(0, I_2)$ and both rows of $\sigma_k \in \mathbb{R}^{2\times 4}$ are i.i.d. sampled from $\mathcal{N}(0, 10^{-4} I_4)$. Furthermore, for all $0 \leq k < 1200$, let $y_k = (\phi_k + v_k)\theta_{\mathrm{true},k} + w_k$, where both rows of $v_k \in \mathbb{R}^{2\times 4}$ are i.i.d. sampled from $\mathcal{N}(0, \frac{1}{100} I_4)$ and w_k is i.i.d. sampled from $\mathcal{N}(0, \frac{1}{100} I_2)$, i.e. regressor noise and measurement noise.

The result of this sampled data is, between $0 \leq k < 400$, excitation is sufficient to identify $\theta^1_{\mathrm{true},k}$ and $\theta^2_{\mathrm{true},k}$, which form a basis for the excited subspace $S^{12} \triangleq \{[x_1\ x_2\ 0\ 0] \colon x_1, x_2 \in \mathbb{R}\}$, but insufficient to identify $\theta^3_{\mathrm{true},k}$ and $\theta^4_{\mathrm{true},k}$, which form a basis for the unexcited subspace $(S^{12})^{\perp} = S^{34} \triangleq \{[0\ 0\ x_3\ x_4] \colon x_3, x_4 \in \mathbb{R}\}$. Then, the reverse is the case between $400 \leq k < 800$. Finally, between $800 \leq k \leq 1200$, excitation is sufficient to identify $\theta^{\parallel}_{\mathrm{true},k} \in \mathbb{R}$, defined as

$$\theta^{\parallel}_{\mathrm{true},k} \triangleq \frac{2}{3}\theta^2_{\mathrm{true},k} + \frac{1}{3}\theta^3_{\mathrm{true},k}, \quad (43)$$

which forms a basis for the excited subspace $S^{\parallel} \triangleq \{[0\ 2x\ x\ 0] \colon x \in \mathbb{R}\}$. However, excitation is insufficient to identify $\theta^1_{\mathrm{true},k}$, $\theta^4_{\mathrm{true},k}$, and $\theta^{\perp}_{\mathrm{true},k} \in \mathbb{R}$, defined as

$$\theta^{\perp}_{\mathrm{true},k} \triangleq \frac{1}{3}\theta^2_{\mathrm{true},k} - \frac{2}{3}\theta^3_{\mathrm{true},k}, \quad (44)$$

which form a basis for the unexcited subspace $S^{\perp} \triangleq \{[x_1\ x\ \ -2x\ x_4] \colon x_1, x, x_4 \in \mathbb{R}\}$.

We now consider the performance of SIFt-RLS in identifying $\theta_{\mathrm{true},k}$. We initialize $P_0 = I_4$, $\theta_0 = 0_{4\times 1}$ and let $\lambda = 0.5$ and $\varepsilon = 10^{-4}$.[6] The parameter tracking of SIFT-RLS is shown in Figure 2, where, for all $k \geq 0$, θ^i_k is the i^{th} element of θ_k. Between $0 \leq k < 400$, SIFt-RLS tracks $\theta^1_{\mathrm{true},k}$ and $\theta^2_{\mathrm{true},k}$ with little change to its estimate of the $\theta^3_{\mathrm{true},k}$ and $\theta^4_{\mathrm{true},k}$ and vice versa for $400 \leq k < 800$. Similarly, between $800 \leq k \leq 1200$, SIFt-RLS tracks $\theta^{\parallel}_{\mathrm{true},k}$ with little change to its estimate of $\theta^1_{\mathrm{true},k}$, $\theta^4_{\mathrm{true},k}$, and $\theta^{\perp}_{\mathrm{true},k}$.

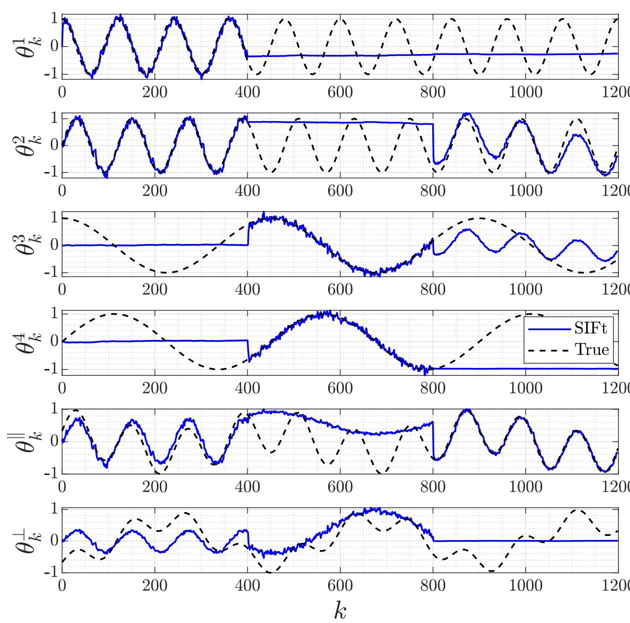

Fig. 2. Estimated parameters θ^1_k, θ^2_k, θ^3_k, θ^4_k, $\theta^{\parallel}_k$, and $\theta^{\perp}_k$ using SIFt-RLS (blue) and true parameters $\theta^1_{\mathrm{true},k}$, $\theta^2_{\mathrm{true},k}$, $\theta^3_{\mathrm{true},k}$, $\theta^4_{\mathrm{true},k}$, $\theta^{\parallel}_{\mathrm{true},k}$, and $\theta^{\perp}_{\mathrm{true},k}$ (black).

Next, we compare SIFt-RLS to RLS with no forgetting (NF), RLS with exponential forgetting (EF), and three other directional forgetting algorithms from the literature: variable direction forgetting (VD) [10], directional forgetting (DF) [3], and multiple forgetting (MF) [9,34]. EF and VD are tuned with a conservative $\lambda = 0.95$ to delay covariance windup. DF is tuned with the same

[6] We also select $q_{\max} = 1$ which only affects computation time.

$\lambda = 0.5$ of SIFt. MF is tuned with $\Lambda \triangleq [\lambda_1\ \lambda_2\ \lambda_3\ \lambda_4] = [0.9\ 0.9\ 0.95\ 0.95]$ to forget more agressively in the first and second parameters. Finally, we test MF with a priori knowledge of which subspaces will be excited (MF a priori), where time-varying forgetting factor Λ is tuned to only forget in the excited subspace. Tuning parameters are summarized in Table 1.

Table 1
RLS Algorithms Tuning Parameters

Algorithm	Tuning Parameters
SIFt	$\lambda = 0.5$, $\varepsilon = 10^{-4}$
NF	—
EF	$\lambda = 0.95$
VD	$\lambda = 0.95$, $\varepsilon = 10^{-4}$
DF	$\lambda = 0.5$
MF	$\Lambda \triangleq [\lambda_1\ \lambda_2\ \lambda_3\ \lambda_4] = [0.9\ 0.9\ 0.95\ 0.95]$
MF a priori	$\Lambda = \begin{cases} [0.5\ 0.5\ 1\ 1] & 0 \le k < 400, \\ [1\ 1\ 0.5\ 0.5] & 400 \le k < 800, \\ [1\ 0.6\ 0.85\ 1] & 800 \le k \le 1200, \end{cases}$

To more easily compare the performance of these seven different algorithms in Figure 3, we show, over $0 \le k \le 800$, the error in the S^{12} subspace, $e_k^{12} \in \mathbb{R}$, and the error in the S^{34} subspace, $e_k^{34} \in \mathbb{R}$, defined as

$$e_k^{12} \triangleq \sqrt{(\theta_k^1 - \theta_{\mathrm{true},k}^1)^2 + (\theta_k^2 - \theta_{\mathrm{true},k}^2)^2}, \tag{45}$$

$$e_k^{34} \triangleq \sqrt{(\theta_k^3 - \theta_{\mathrm{true},k}^3)^2 + (\theta_k^4 - \theta_{\mathrm{true},k}^4)^2}. \tag{46}$$

Furthermore, Figure 3 also shows, between $800 \le k \le 1200$, the error in the $S^{\parallel}$ subspace, $e_k^{\parallel} \in \mathbb{R}$, and the error in the $S^{\perp}$ subspace, $e_k^{\perp} \in \mathbb{R}$, defined as

$$e_k^{\parallel} \triangleq \sqrt{(\theta_k^{\parallel} - \theta_{\mathrm{true},k}^{\parallel})^2},$$
$$e_k^{\perp} \triangleq \sqrt{(\theta_k^1 - \theta_{\mathrm{true},k}^1)^2 + (\theta_k^4 - \theta_{\mathrm{true},k}^4)^2 + (\theta_k^{\perp} - \theta_{\mathrm{true},k}^{\perp})^2}.$$

Note that NF is unable to track changing parameters and results in the largest errors in the excited subspace. On the contrary, while EF, VD, DF, and MF are able to track the true parameters in the excited subspaces, they all result in large error in the unexcited subspaces due to covariance windup. DF performs the best of these methods, particularly between $0 \le k \le 800$, but still suffers from covariance windup between $800 \le k \le 1200$. While an aggressive $\lambda = 0.5$ is used for DF, slower windup still occurs with a more conservative λ. Finally, MF a priori achieved very comparable performance to SIFt, with tracking in the excited subspace and little change in estimates in the unexcited subspace. However, this required careful tuning and a priori knowledge of which subspaces would be excited.

As another useful metric, we define $\Delta_k^{12} \in \mathbb{R}$, $\Delta_k^{34} \in \mathbb{R}$, and $\Delta_k^{\perp} \in \mathbb{R}$ over $k \in [1, 400]$, $k \in [401, 800]$, and $k \in [801, 1200]$, respectively, as

$$\Delta_k^{12} \triangleq \sqrt{(\theta_k^1 - \theta_0^1)^2 + (\theta_k^2 - \theta_0^2)^2},$$
$$\Delta_k^{34} \triangleq \sqrt{(\theta_k^3 - \theta_{400}^3)^2 + (\theta_k^4 - \theta_{400}^4)^2},$$
$$\Delta_k^{\perp} \triangleq \sqrt{(\theta_k^1 - \theta_{800}^1)^2 + (\theta_k^4 - \theta_{800}^4)^2 + (\theta_k^{\perp} - \theta_{800}^{\perp})^2}.$$

These quantities show how much parameter estimates change in the unexcited subspaces. Figure 4 shows that NF, SIFt, and MF a priori makes only small changes to parameter estimates in the unexcited subspaces, while EF, VD, DF, and MF significantly update estimates of parameters in the unexcited subspace despite lack of useful new information in these directions.

As a final metric, Figure 5 shows the spectral radius of the covariance matrix P_k for the various RLS algorithms. The spectral radius $\rho(P_k)$ becomes very small in NF, indicating parameter estimates is no longer changing. In EF, VD, DF, and MF, the spectral radius becomes very large, indicating covariance windup and sensitivity to noise. SIFt and MF a priori both maintain a spectral radius close to 1.

7 Conclusion

This work develops SIFt-RLS, a directional forgetting algorithm which, at each step, forgets only in the information subspace. The key tool used in SIFt-RLS is a subspace decomposition of a positive-definite matrix, which was developed and analyzed in Section 3. The properties of this decomposition are used to derive explicit covariance bounds for SIFt-RLS and to develop sufficient conditions for the stability of the parameter estimation error dynamics. Moreover, an interesting parallel is made between the covariance bounds of SIFt-RLS and the covariance bounds of exponential forgetting RLS under persistent excitation with persistency window 1. A numerical example show the benefits of SIFt-RLS when the data collected is not persistently exciting.

We have presented a simple version of SIFt-RLS, in which forgetting in the information subspace is uniform and constant-rate, and no forgetting is applied in the orthogonal complement of the information subspace. However, a natural area of future interest is combining SIFt-RLS with other forgetting algorithms. One idea is to apply nonuniform forgetting and/or variable-rate forgetting to the information subspace. Another idea is to apply a separate forgetting method, for example resetting [17], to the orthogonal complement of the information subspace. While some other forgetting methods require persistent excitation to guarantee bounds on the covariance matrix [3, 7, 10], a question of interest is whether SIFt-RLS combined with these methods can guarantee similar bounds without persistent excitation, as was shown in this work for exponential forgetting.

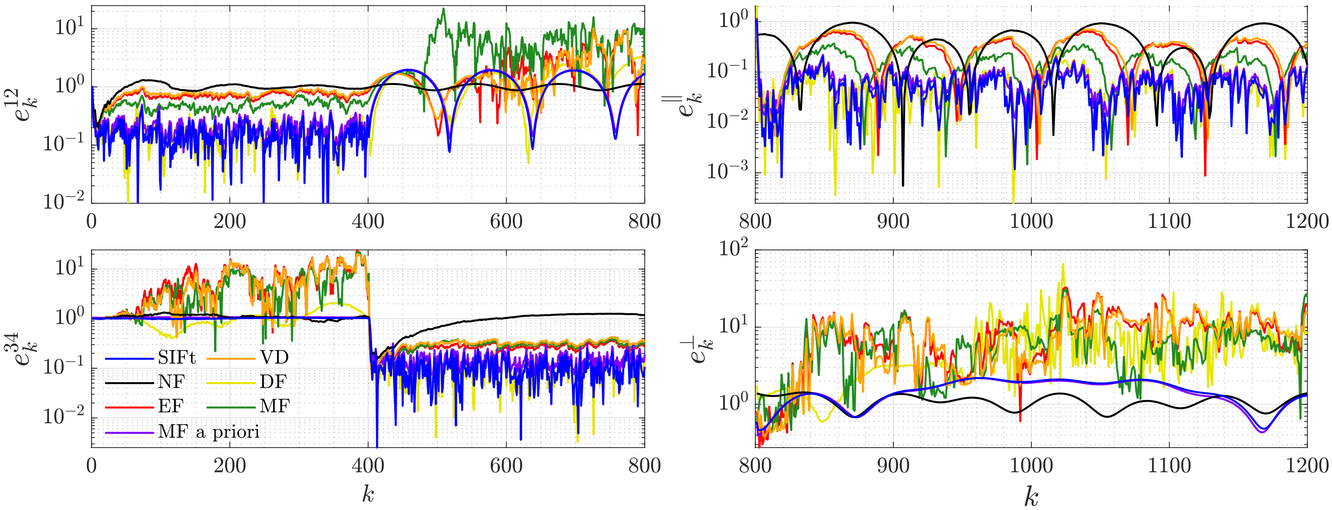

Fig. 3. Over $0 \leq k \leq 800$, e_k^{12} and e_k^{34} show the parameter estimation error in the subspaces $S^{34} \triangleq \{[x_1\ x_2\ 0\ 0]^{\mathrm{T}} : x_1, x_2 \in \mathbb{R}\}$ and $S_{34} \triangleq \{[0\ 0\ x_3\ x_4]^{\mathrm{T}} : x_3, x_4 \in \mathbb{R}\}$, respectively. Subspace S^{12} is excited over $0 \leq k < 400$ while S^{34} is not, and vice versa over $400 \leq k < 800$. Next, over $800 \leq k \leq 1200$, $e_k^{\parallel}$ and $e_k^{\perp}$ show the parameter estimation error in the subspaces $S^{\parallel} \triangleq \{[0\ 2x\ x\ 0] : x \in \mathbb{R}\}$ and $S^{\perp} \triangleq \{[x_1\ x\ \ -2x\ x_4] : x_1, x, x_4 \in \mathbb{R}\}$, respectively. Subspace $S^{\parallel}$ is excited over $800 \leq k \leq 1200$ while $S^{\perp}$ is not.

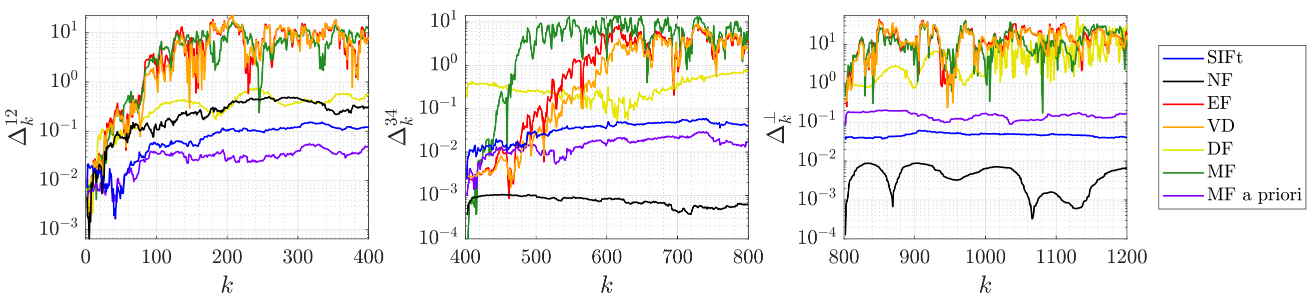

Fig. 4. Δ_k^{12}, Δ_k^{34}, and $\Delta_k^{\perp}$ show the magnitude of the change in parameter estimates in the unexcited subspace over $0 \leq k < 400$, $400 \leq k < 800$, and $800 \leq k \leq 1200$, respectively.

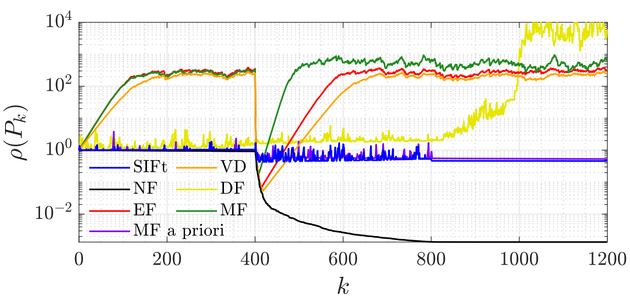

Fig. 5. Spectral radius of the covariance matrix at step k.

Acknowledgements

This work was supported by the NSF Graduate Research Fellowship under Grant DGE 1841052.

References

[1] Karl J Åström and Björn Wittenmark. Adaptive control. Courier Corporation, 2013.

[2] Dennis S Bernstein. Matrix mathematics: theory, facts, and formulas. Princeton university press, 2009.

[3] Sergio Bittanti, Paolo Bolzern, and M Campi. Convergence and exponential convergence of identification algorithms with directional forgetting factor. Automatica, 26(5):929–932, 1990.

[4] Adam L Bruce, Ankit Goel, and Dennis S Bernstein. Convergence and consistency of recursive least squares with variable-rate forgetting. Automatica, 119:109052, 2020.

[5] Liyu Cao and Howard Schwartz. A directional forgetting algorithm based on the decomposition of the information matrix. Automatica, 36(11):1725–1731, 2000.

[6] Moody T. Chu, R. E. Funderlic, and Gene H. Golub. Rank modifications of semidefinite matrices associated with a secant update formula. SIAM Journal on Matrix Analysis and Applications, 20(2):428–436, 1998.

[7] S. Dasgupta and Yih-Fang Huang. Asymptotically convergent modified recursive least-squares with data-

dependent updating and forgetting factor for systems with bounded noise. IEEE Transactions on Information Theory, 33(3):383–392, 1987.

[8] TR Fortescue, Lester S Kershenbaum, and B Erik Ydstie. Implementation of self-tuning regulators with variable forgetting factors. Automatica, 17(6):831–835, 1981.

[9] Francesco Fraccaroli, Andrea Peruffo, and Mattia Zorzi. A new recursive least squares method with multiple forgetting schemes. In 2015 54th IEEE conference on decision and control (CDC), pages 3367–3372. IEEE, 2015.

[10] Ankit Goel, Adam L Bruce, and Dennis S Bernstein. Recursive least squares with variable-direction forgetting: Compensating for the loss of persistency [lecture notes]. IEEE Control Systems Magazine, 40(4):80–102, 2020.

[11] GC Goodwin, EK Teoh, and H Elliott. Deterministic convergence of a self-tuning regulator with covariance resetting. In IEEE Proceedings D-Control Theory and Applications, volume 130, pages 6–8, 1983.

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

[13] Roger A Horn and Charles R Johnson. Matrix analysis. Cambridge university press, 2012.

[14] Syed Aseem Ul Islam and Dennis S Bernstein. Recursive least squares for real-time implementation [lecture notes]. IEEE Control Systems Magazine, 39(3):82–85, 2019.

[15] Richard M Johnstone, C Richard Johnson Jr, Robert R Bitmead, and Brian DO Anderson. Exponential convergence of recursive least squares with exponential forgetting factor. Systems & Control Letters, 2(2):77–82, 1982.

[16] R Kulhavỳ and M Kárnỳ. Tracking of slowly varying parameters by directional forgetting. IFAC Proceedings Volumes, 17(2):687–692, 1984.

[17] Brian Lai and Dennis S. Bernstein. Exponential resetting and cyclic resetting recursive least squares. IEEE Control Systems Letters, 7:985–990, 2023.

[18] Brian Lai and Dennis S Bernstein. Generalized forgetting recursive least squares: Stability and robustness guarantees. arXiv preprint arXiv:2308.04259, 2023.

[19] Brian Lai, Syed Aseem Ul Islam, and Dennis S Bernstein. Regularization-induced bias and consistency in recursive least squares. In 2021 American Control Conference (ACC), pages 3987–3992. IEEE, 2021.

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

[21] Lennart Ljung and Torsten Söderström. Theory and practice of recursive identification. MIT press, 1983.

[22] OP Malik, GS Hope, and SJ Cheng. Some issues on the practical use of recursive least squares identification in self-tuning control. International Journal of Control, 53(5):1021–1033, 1991.

[23] Nima Mohseni and Dennis S Bernstein. Recursive least squares with variable-rate forgetting based on the f-test. In 2022 American Control Conference (ACC), pages 3937–3942. IEEE, 2022.

[24] Tam W Nguyen, Syed Aseem Ul Islam, Dennis S Bernstein, and Ilya V Kolmanovsky. Predictive cost adaptive control: A numerical investigation of persistency, consistency, and exigency. IEEE Control Systems Magazine, 41(6):64–96, 2021.

[25] Romeo Ortega, Vladimir Nikiforov, and Dmitry Gerasimov. On modified parameter estimators for identification and adaptive control. a unified framework and some new schemes. Annual Reviews in Control, 50:278–293, 2020.

[26] Constantin Paleologu, Jacob Benesty, and Silviu Ciochina. A robust variable forgetting factor recursive least-squares algorithm for system identification. IEEE Signal Processing Letters, 15:597–600, 2008.

[27] Constantin Paleologu, Jacob Benesty, and Silviu Ciochină. A practical variable forgetting factor recursive least-squares algorithm. In 2014 11th International symposium on electronics and telecommunications (ISETC), pages 1–4. IEEE, 2014.

[28] JE Parkum, Niels Kjølstad Poulsen, and Jan Holst. Recursive forgetting algorithms. International Journal of Control, 55(1):109–128, 1992.

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

[30] Shankar Sastry, Marc Bodson, and James F Bartram. Adaptive control: stability, convergence, and robustness, 1990.

[31] Vitaly Shaferman, Michael Schwegel, Tobias Glück, and Andreas Kugi. Continuous-time least-squares forgetting algorithms for indirect adaptive control. European Journal of Control, 62:105–112, 2021.

[32] Hyo-Sang Shin and Hae-In Lee. A new exponential forgetting algorithm for recursive least-squares parameter estimation. arXiv preprint arXiv:2004.03910, 2020.

[33] Lloyd N Trefethen and David Bau. Numerical linear algebra, volume 181. Siam, 2022.

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

[35] Kun Zhu, Chengpu Yu, and Yiming Wan. Recursive least squares identification with variable-direction forgetting via oblique projection decomposition. IEEE/CAA Journal of Automatica Sinica, 9(3):547–555, 2021.

A Useful Matrix Lemmas

Lemma A.1 (Matrix Inversion Lemma) *Let $A \in \mathbb{R}^{n\times n}$, $U \in \mathbb{R}^{n\times p}$, $C \in \mathbb{R}^{p\times p}$, and $V \in \mathbb{R}^{p\times n}$. If A, C, and $A + UCV$ are nonsingular, then $C^{-1} + VA^{-1}U$ is nonsingular, and $(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$.*

Proof. See Corollary 3.9.8 of [2]. ■

Lemma A.2 (min-max theorem) *Let $A \in \mathbb{R}^{n\times n}$ be symmetric. Then, for all $k = 1, \ldots, n$,*

$$\boldsymbol{\lambda_k}(A) = \max_{\substack{S\subset\mathbb{R}^n \\ \dim(S)=k}} \min_{\substack{x\in S \\ \|x\|=1}} x^{\mathrm{T}}Ax.$$

Proof. See Theorem 4.2.6 of [13]. ■

Lemma A.3 *Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times n}$ be symmetric. Then,*

$$\boldsymbol{\lambda_k}(A) + \boldsymbol{\lambda_{\min}}(B) \le \boldsymbol{\lambda_k}(A+B) \le \boldsymbol{\lambda_k}(A) + \boldsymbol{\lambda_{\max}}(B).$$

Proof. See Theorem 10.4.11 of [2]. ■

Lemma A.4 *Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times n}$. Then, the eigenvalues of $AB \in \mathbb{R}^{n\times n}$ are the eigenvalues of $BA \in \mathbb{R}^{n\times n}$*

Proof. See Theorem 1.3.22 of [13]. ■

Lemma A.5 *Let $H \in \mathbb{R}^{n\times n}$ and $M \in \mathbb{R}^{n\times n}$ by symmetric matrices. Then,* $\operatorname{rank}(H-M) = \operatorname{rank}(H) - \operatorname{rank}(M)$ *if and only if there exists $S \in \mathbb{R}^{n\times p}$ such that $M = HS(S^{\mathrm{T}}HS)^{-1}S^{\mathrm{T}}H$.*

Proof. See Lemma 2.4 in [6]. ■

Lemma A.6 *Let $H, M \in \mathbb{R}^{n\times n}$, where H is positive definite and M is positive semidefinite. Then, $H - M \succeq 0$ (respectively $H - M \succ 0$) if and only if $\rho(H^{-1}M) \le 1$ (respectively $\rho(H^{-1}M) < 1$).*

Proof. See Theorem 7.7.3 in [13]. ■

Lemma A.7 *Let $A \in \mathbb{R}^{n\times n}$ be positive definite and $b \in \mathbb{R}^{n\times p}$. Then, $b^{\mathrm{T}}Ab \in \mathbb{R}^{p\times p}$ is positive semidefinite,* $\operatorname{rank}(b^{\mathrm{T}}Rb) = \operatorname{rank}(b)$*, and $b^{\mathrm{T}}Ab$ is positive definite if and only if* $\operatorname{rank}(b) = p$.

Proof. See Observation 7.1.8 in [13]. ■

B Proofs in the Subspace Decomposition of a Positive-Definite Matrix

Proof.[Proof of Theorem 1] By Lemma A.7, $v^{\mathrm{T}}Av$ is positive definite. Hence, $A^{\parallel S}$ and $A^{\perp S}$ are well defined.

Proof of (1a), (1b): Since $v^{\mathrm{T}}Av$ is positive definite, $(v^{\mathrm{T}}Av)^{-1}$ is also positive definite. Furthermore, Lemma A.7 implies that $A^{\parallel S} = (Av)(v^{\mathrm{T}}Av)^{-1}(Av)^{\mathrm{T}}$ is positive semidefinite. Next, note that by Lemma A.6, $A^{\perp S} = A - A^{\parallel S}$ is positive semidefinite if and only if $\rho(A^{-1}A^{\parallel S}) \le 1$. Notice that

$$\begin{aligned}(A^{-1}A^{\parallel S})^2 &= \left[v(v^{\mathrm{T}}Av)^{-1}v^{\mathrm{T}}A\right]^2,\\ &= v(v^{\mathrm{T}}Av)^{-1}v^{\mathrm{T}}Av(v^{\mathrm{T}}Av)^{-1}v^{\mathrm{T}}A,\\ &= v(v^{\mathrm{T}}Av)^{-1}v^{\mathrm{T}}A, = A^{-1}A^{\parallel S}.\end{aligned}$$

Therefore, $A^{-1}A^{\parallel S}$ is idempotent implying each of its n eigenvalues are either 0 or 1. Hence, $\rho(A^{-1}A^{\parallel S}) \le 1$ implying that $A^{\perp S}$ is positive semidefinite.

Proof of (2a), (2b): Define $B \triangleq (v^{\mathrm{T}}Av)^{1/2} \in \mathbb{R}^{p\times p}$. It follows that

$$A^{\parallel S} = Av(BB^{\mathrm{T}})^{-1}v^{\mathrm{T}}A = (AvB^{-1})^{\mathrm{T}}(AvB^{-1}).$$

Thus, $\operatorname{rank}(A^{\parallel S}) = \operatorname{rank}(AvB^{-1})$. Since A and B^{-1} are positive definite, it follows that $\operatorname{rank}(AvB^{-1}) = \operatorname{rank}(v)$. Hence, $\operatorname{rank}(A^{\parallel S}) = \operatorname{rank}(v) = p$. Next, Lemma A.5 implies that $\operatorname{rank}(A^{\perp S}) = \operatorname{rank}(A) - \operatorname{rank}(A^{\parallel S}) = n - p$.

Proof of (3a), (3b): Let $x \in S$. Since $v_1, \ldots, v_p$ is a basis for S, it follows that there exists $c \in \mathbb{R}^{p\times 1}$ such that $x = vc$. Hence,

$$A^{\parallel S}x = Av(v^{\mathrm{T}}Av)v^{\mathrm{T}}Avc = Avc = Ax,$$

and $A^{\perp S}x = (A - A^{\parallel S})x = Ax - Ax = 0$. ■

Proof.[Proof of Theorem 2] Let $v_1, \cdots, v_p \in \mathbb{R}^n$ be a basis for S and $v \triangleq \begin{bmatrix} v_1 & \cdots & v_p\end{bmatrix} \in \mathbb{R}^{n\times p}$. Begin by writing $\tilde{A} = AXA$, where $X \triangleq A^{-1}\tilde{A}A^{-1} \in \mathbb{R}^{n\times n}$ and define $P \triangleq v(v^{\mathrm{T}}v)^{-1}v^{\mathrm{T}} \in \mathbb{R}^{n\times n}$. Since $\tilde{A}v = Av$, it follows that

$$XAP = A^{-1}\tilde{A}v(v^{\mathrm{T}}v)^{-1}v^{\mathrm{T}} = A^{-1}Av(v^{\mathrm{T}}v)^{-1}v^{\mathrm{T}} = P.$$

Since $XAP = P$, it follows that the column space of P is a subset of the column space of X. However, note that $\operatorname{rank}(P) = \operatorname{rank}(X) = p$. Therefore, X and P have the same column space.

Hence, for all $x \in \mathbb{R}^n$, there exists $y \in \mathbb{R}^n$ such that $Xx = Py$. Moreover, since $P^2 = P$, it follows that, for all $x \in \mathbb{R}^n$, there exists $y \in \mathbb{R}^n$ such that $PXx = PPy = Py = Xx$. Hence, $PX = X$. Moreover, since P, X, and $PX = X$ are symmetric, it follows that P and X commute. Therefore,

$$PX = XP = X.$$

Next, define $Q \triangleq PAP$. Note that $XQ = XPAP = XAP = P$. Moreover, since X, Q, and $XQ = P$ are symmetric, it follows that X and Q commute. Therefore,

$$XQ = QX = P.$$

Thus, it follows that X and Q satisfy the four Moore-Penrose Conditions:

(1) $XQX = XP = X$,
(2) $QXQ = QP = PAPP = PAP = Q$,
(3) $XQ = P = P^{\mathrm{T}} = (XQ)^{\mathrm{T}}$,
(4) $QX = P = P^{\mathrm{T}} = (QX)^{\mathrm{T}}$.

Therefore, $X = Q^+$ and $\tilde{A} = AQ^+A$. Finally, define $G \triangleq v(v^{\mathrm{T}}Av)^{-1}v^{\mathrm{T}} \in \mathbb{R}^{n\times n}$. It can be easily verified that $Q = v(v^{\mathrm{T}}v)^{-1}v^{\mathrm{T}}Av(v^{\mathrm{T}}v)^{-1}v^{\mathrm{T}}$ and G satisfy the four Moore-Penrose Conditions $GQG = G$, $QGQ = Q$, $GQ = (GQ)^{\mathrm{T}}$, and $QG = (QG)^{\mathrm{T}}$. Hence, $X = Q^+ = (G^+)^+ = G$, and $\tilde{A} = AGA = Av(v^{\mathrm{T}}Av)^{-1}v^{\mathrm{T}}A = A^{\|S}$. ■

Proof.[Proof of Proposition 4] Note that since $\langle\cdot,\cdot\rangle_A$ is an inner product, it follows that $\dim(S^{\perp_A}) = n - \dim(S) = n - p$. Let $w_1, \ldots, w_{n-p} \in \mathbb{R}^n$ be a basis for $S^{\perp_A}$ and let $w \triangleq [w_1 \ \cdots \ w_{n-p}] \in \mathbb{R}^{n\times n-p}$. Then, $A^{\perp(S^{\perp_A})}$ can be written as

$$A^{\perp(S^{\perp_A})} = A - Aw(w^{\mathrm{T}}Aw)^{-1}w^{\mathrm{T}}A.$$

Note that $A^{\perp(AS)^{\perp}}$ is the difference of two symmetric matrices, A and $Aw(w^{\mathrm{T}}Aw)^{-1}w^{\mathrm{T}}A$, hence is symmetric. Next, since $\mathrm{rank}(w) = n - p$, it follows that $\mathrm{rank}(Aw(w^{\mathrm{T}}Aw)^{-1}w^{\mathrm{T}}A) = n - p$. Furthermore, it follows from Lemma A.5 that $\mathrm{rank}(A^{\perp(AS)^{\perp}}) = \mathrm{rank}(A) - \mathrm{rank}(Aw(w^{\mathrm{T}}Aw)^{-1}w^{\mathrm{T}}A) = n - (n - p) = p$. Finally, since the columns of w are elements of $S^{\perp_A}$, it follows that, for all $x \in S$, $w^{\mathrm{T}}Ax = 0_{n-p\times 1}$. Hence, for all $x \in S$,

$$A^{\perp(S^{\perp_A})}x = Ax - Aw(w^{\mathrm{T}}Aw)^{-1}w^{\mathrm{T}}Ax = Ax.$$

Thus, the three properties of Theorem 2 are satisfied and $A^{\|S} = A^{\perp(S^{\perp_A})}$ holds, proving (20).

Next, it follows from definition (15) that $A^{\perp(S^{\perp_A})} = A - A^{\|(S^{\perp_A})}$ which can be rewritten as $A - A^{\perp(S^{\perp_A})} = A^{\|(S^{\perp_A})}$. It then follows from (20) that $A - A^{\|S} = A^{\|(S^{\perp_A})}$. Finally, by definition (15), it follows that $A^{\perp S} = A^{\|(S^{\perp_A})}$, proving (21). ■

C Proof of Theorem 3

We begin the proof of Theorem 3 with a useful Lemma.

Lemma 1 *For all $k \geq 0$, the $n - q_k$ nonzero eigenvalues of $R_k^{\perp}$ are bounded below by $\boldsymbol{\lambda}_{\mathbf{min}}(R_k)$ and bounded above by $\boldsymbol{\lambda}_{\mathbf{max}}(R_k)$.*

Proof. Let $k \geq 0$. To begin, note that by Theorem 1, $\mathrm{rank}(R_k^{\perp}) = n - q_k$. Therefore, $R_k^{\perp}$ has $n - q_k$ nonzero eigenvalues.

First, we show the upper bound. Since $R_k = R_k^{\|} + R_k^{\perp}$, it follows from Lemma A.3 that $\boldsymbol{\lambda}_{\mathbf{min}}(R_k^{\|}) + \boldsymbol{\lambda}_{\mathbf{max}}(R_k^{\perp}) \leq \boldsymbol{\lambda}_{\mathbf{max}}(R_k)$. Next, Theorem 1 implies that $R_k^{\|}$ is positive semidefinite, thus $0 \leq \boldsymbol{\lambda}_{\mathbf{min}}(R_k^{\|})$. It then follows that $\boldsymbol{\lambda}_{\mathbf{max}}(R_k^{\perp}) \leq \boldsymbol{\lambda}_{\mathbf{max}}(R_k)$.

Second, to show the lower bound, first let $R_k^{1/2} \in \mathbb{R}^n$ be the unique positive-definite matrix such that $R_k = R_k^{1/2}R_k^{1/2}$. Furthermore, define $\Omega_k \in \mathbb{R}^{n\times n}$ as

$$\Omega_k \triangleq I_n - R_k^{1/2}\bar{\phi}_k^{\mathrm{T}}(\bar{\phi}_k R_k \bar{\phi}_k^{\mathrm{T}})^{-1}\bar{\phi}_k R_k^{1/2}.$$

It then follows that $R_k^{\perp} = R_k^{1/2}\Omega_k R_k^{\frac{1}{2}}$. It can also be easily verified that $\Omega_k^2 = \Omega_k$, hence $R_k^{\perp} = (R_k^{\frac{1}{2}}\Omega_k)(\Omega_k R_k^{\frac{1}{2}})$. Then, it follows from Lemma A.4 that the eigenvalues of $R_k^{\perp}$ are the same as the eigenvalues of $(\Omega_k R_k^{\frac{1}{2}})(R_k^{\frac{1}{2}}\Omega_k) = \Omega_k R_k \Omega_k$.

Next, since $R_k^{\perp}$ has $n - q_k$ nonzero eigenvalues, it suffices to show that $\boldsymbol{\lambda}_{\boldsymbol{n-q_k}}(R_k^{\perp}) \geq \boldsymbol{\lambda}_{\mathbf{min}}(R_k)$. Furthermore, by the previous result, it holds that $\boldsymbol{\lambda}_{\boldsymbol{n-q_k}}(R_k^{\perp}) = \boldsymbol{\lambda}_{\boldsymbol{n-q_k}}(\Omega_k R_k \Omega_k)$. It then follows from Lemma A.2 that

$$\boldsymbol{\lambda}_{\boldsymbol{n-q_k}}(R_k^{\perp}) = \max_{\substack{S\subset\mathbb{R}^n \\ \dim(S)=n-q_k}} \ \min_{\substack{x\in S \\ \|x\|=1}} x^{\mathrm{T}}\Omega_k R_k \Omega_k x.$$

Furthermore, since $\Omega_k = R_k^{-1/2}R_k^{\perp}R_k^{-1/2}$ and $R^{-1/2}$ is positive definite, it follows that $\mathrm{rank}(\Omega_k) = n - q_k$. Hence, $\dim(\mathcal{R}(\Omega_k)) = n - q_k$. Therefore,

$$\boldsymbol{\lambda}_{\boldsymbol{n-q_k}}(R_k^{\perp}) \geq \min_{x\in\mathcal{R}(\Omega_k),\ \|x\|=1} x^{\mathrm{T}}\Omega_k R_k \Omega_k x.$$

Next, since $\Omega_k^2 = \Omega_k$, it follows that, for all $x \in \mathcal{R}(\Omega_k)$, $\Omega_k x = x$. Hence,

$$\boldsymbol{\lambda}_{\boldsymbol{n-q_k}}(R_k^{\perp}) \geq \min_{x\in\mathcal{R}(\Omega_k),\ \|x\|=1} x^{\mathrm{T}}R_k x.$$

Finally, note that

$$\min_{x\in\mathcal{R}(\Omega_k),\ \|x\|=1} x^{\mathrm{T}}R_k x \geq \min_{\|x\|=1} x^{\mathrm{T}}R_k x = \boldsymbol{\lambda}_{\mathbf{min}}(R_k). \ ■$$

We now present the proof of Theorem 3.

Proof.[Proof of Theorem 3] Proof follows by induction on $k \geq 0$. First, consider the base case $k = 0$. Note that $R_0 \succeq \boldsymbol{\lambda}_{\mathbf{min}}(R_0)I_n$ implying (34) and that $R_0 \preceq \boldsymbol{\lambda}_{\mathbf{max}}(R_0)I_n$ implying (36).

Next, let $k \geq 0$. Let $\phi_k = U_k\Sigma_k V_k^{\mathrm{T}}$ be the full SVD of ϕ_k, where $U_k \in \mathbb{R}^{p\times p}$ is orthogonal, $\Sigma_k \in \mathbb{R}^{p\times n}$ is rectangular diagonal with diagonal in descending order, and $V_k \in \mathbb{R}^{n\times n}$ is orthogonal. Moreover, denote the singular of ϕ_k as $\sigma_{1,k}, \sigma_{2,k}, \ldots, \sigma_{\min\{p,n\},k}$. Finally, since $q_k \leq \min\{p, n\}$, we can partition the columns of U_k and rows of V_k as

$$U_k = \begin{bmatrix} U_{1,k} & U_{2,k} \end{bmatrix}, \quad V_k = \begin{bmatrix} V_{1,k} & V_{2,k} \end{bmatrix}, \tag{C.1}$$

where $U_{1,k} \in \mathbb{R}^{p\times q_k}$ and $V_{1,k} \in \mathbb{R}^{n\times q_k}$ are the first q_k columns of U_k and V_k respectively, and $U_{2,k} \in \mathbb{R}^{p\times(p-q_k)}$ and $V_{2,k} \in \mathbb{R}^{n\times(n-q_k)}$ are the remaining columns of U_k and V_k respectively. Note that $\bar{\phi}_k \in \mathbb{R}^{q_k\times n}$, defined in (22), can be expressed as

$$\bar{\phi}_k = U_{1,k}^{\mathrm{T}}\phi_k. \tag{C.2}$$

Moreover, the columns of $V_{1,k}$ form an orthonormal basis for $\mathcal{R}(\bar{\phi}_k^{\mathrm{T}})$.

Next, it follows from (27) and (29) that $R_{k+1} = \lambda R_k^{\parallel} + R_k^{\perp} + \bar{\phi}_k^{\mathrm{T}}\bar{\phi}_k$. It then follows from (26) that

$$R_{k+1} = \lambda R_k + (1-\lambda)R_k^{\perp} + \bar{\phi}_k^{\mathrm{T}}\bar{\phi}_k. \tag{C.3}$$

Pre-multiplying by V_k^{T} and post-multiplying by V_k, it follows that

$$V_k^{\mathrm{T}}R_{k+1}V_k = V_k^{\mathrm{T}}[\lambda R_k + (1-\lambda)R_k^{\perp} + \bar{\phi}_k^{\mathrm{T}}\bar{\phi}_k]V_k. \tag{C.4}$$

Since the columns of $V_{1,k}$ form an orthonormal basis for $\mathcal{R}(\bar{\phi}_k^{\mathrm{T}})$, it follows from Theorem 1 that $R_k^{\perp}V_{1,k} = 0_{n\times q_k}$. It then follows that

$$V_k^{\mathrm{T}}R_k^{\perp}V_k = \begin{bmatrix} 0_{q_k\times q_k} & 0_{q_k\times(n-q_k)} \\ 0_{(n-q_k)\times q_k} & V_{2,k}^{\mathrm{T}}R_k^{\perp}V_{2,k} \end{bmatrix}. \tag{C.5}$$

Moreover, note that the columns of $V_{2,k}$ are orthogonal to the columns of $V_{1,k}$, hence are also orthogonal to the rows of $\bar{\phi}_k$. Therefore, $\bar{\phi}_k V_{2,k} = 0_{q_k\times(n-q_k)}$. Thus it follows that

$$V_k^{\mathrm{T}}\bar{\phi}_k^{\mathrm{T}}\bar{\phi}_k V_k = \begin{bmatrix} V_{1,k}^{\mathrm{T}}\bar{\phi}_k^{\mathrm{T}}\bar{\phi}_k V_{1,k} & 0_{q_k\times(n-q_k)} \\ 0_{(n-q_k)\times q_k} & 0_{(n-q_k)\times(n-q_k)} \end{bmatrix}. \tag{C.6}$$

Substituting (C.5) and (C.6) into (C.4), it follows that

$$\begin{aligned} V_k^{\mathrm{T}}R_{k+1}V_k &= \lambda V_k^{\mathrm{T}}R_kV_k \\ &\quad + \begin{bmatrix} V_{1,k}^{\mathrm{T}}\bar{\phi}_k^{\mathrm{T}}\bar{\phi}_k V_{1,k} & 0_{q_k\times(n-q_k)} \\ 0_{(n-q_k)\times q_k} & (1-\lambda)V_{2,k}^{\mathrm{T}}R_k^{\perp}V_{2,k} \end{bmatrix}. \end{aligned} \tag{C.7}$$

We now prove two claims which will later be useful.

Claim 1: $V_{1,k}^{\mathrm{T}}\bar{\phi}_k^{\mathrm{T}}\bar{\phi}_kV_{1,k} = \mathrm{diag}(\sigma_{1,k}^2, \ldots, \sigma_{q_k,k}^2)$.

Proof of Claim 1: Note that $\bar{\phi}_k = U_{1,k}^{\mathrm{T}}U_k\Sigma_kV_k^{\mathrm{T}}$. Furthermore, note that, $U_{1,k}^{\mathrm{T}}U_k = [U_{1,k}^{\mathrm{T}}U_{1,k} \;\; U_{1,k}^{\mathrm{T}}U_{2,k}] = [I_{q_k} \;\; 0_{q_k\times(p-q_k)}]$. Similarly, we can express $V_{1,k}^{\mathrm{T}}V_k = [V_{1,k}^{\mathrm{T}}V_{1,k} \;\; V_{1,k}^{\mathrm{T}}V_{2,k}] = [I_{q_k} \;\; 0_{q_k\times(n-q_k)}]$. Therefore, $V_{1,k}^{\mathrm{T}}\bar{\phi}_k^{\mathrm{T}}\bar{\phi}_kV_{1,k}$ can be written as

$$\begin{aligned} V_{1,k}^{\mathrm{T}}\bar{\phi}_k^{\mathrm{T}}\bar{\phi}_kV_{1,k} &= V_{1,k}^{\mathrm{T}}V_k\Sigma_k^{\mathrm{T}}U_k^{\mathrm{T}}U_{1,k}U_{1,k}^{\mathrm{T}}U_k\Sigma_kV_k^{\mathrm{T}}V_{1,k} \\ &= \begin{bmatrix} I_{q_k} & 0 \end{bmatrix}\Sigma_k^{\mathrm{T}}\begin{bmatrix} I_{q_k} & 0 \\ 0 & 0 \end{bmatrix}\Sigma_k\begin{bmatrix} I_{q_k} \\ 0 \end{bmatrix} \\ &= \mathrm{diag}(\sigma_{1,k}^2, \ldots, \sigma_{q_k,k}^2), \end{aligned} \tag{C.8}$$

with zero matrices of appropriate sizes. □

Claim 2: *The $n - q_k$ nonzero eigenvalues of $R_k^{\perp}$ are the eigenvalues of $V_{2,k}^{\mathrm{T}}R_k^{\perp}V_{2,k} \in \mathbb{R}^{(n-q_k)\times(n-q_k)}$.*

Proof of Claim 2: First, note that, by Theorem 1, $\mathrm{rank}(R_k^{\perp}) = n - q_k$, hence $R_k^{\perp}$ has $n - q_k$ nonzero eigenvalues. Next, it follows from (C.5) that $V_k^{\mathrm{T}}R_k^{\perp}V_k$ is block diagonal. Therefore, the eigenvalues of $V_k^{\mathrm{T}}R_k^{\perp}V_k$ are q_k zeros and the $n - q_k$ eigenvalues of $V_{2,k}^{\mathrm{T}}R_k^{\perp}V_{2,k}$. Finally, since similarity transformation preserve eigenvalues, it follows that the eigenvalues of $R_k^{\perp}$ are q_k zeros and the $n - q_k$ eigenvalues of $V_{2,k}^{\mathrm{T}}R_k^{\perp}V_{2,k}$. □

We now show the induction step for statement *(1)* of Theorem 3. First, assume, for inductive hypothesis, that (34) holds $k \geq 0$. Applying Lemma A.3 to (C.7) yields

$$\begin{aligned} &\boldsymbol{\lambda}_{\mathbf{min}}(V_k^{\mathrm{T}}R_{k+1}V_k) \geq \lambda\boldsymbol{\lambda}_{\mathbf{min}}(V_k^{\mathrm{T}}R_kV_k) + \\ &\min\{\boldsymbol{\lambda}_{\mathbf{min}}(V_{1,k}^{\mathrm{T}}\bar{\phi}_k^{\mathrm{T}}\bar{\phi}_kV_{1,k}), (1-\lambda)\boldsymbol{\lambda}_{\mathbf{min}}(V_{2,k}^{\mathrm{T}}R_k^{\perp}V_{2,k})\}. \end{aligned} \tag{C.9}$$

Since similarity transformations preserve eigenvalues, it then follows that

$$\begin{aligned} &\boldsymbol{\lambda}_{\mathbf{min}}(R_{k+1}) \geq \lambda\boldsymbol{\lambda}_{\mathbf{min}}(R_k) + \\ &\min\{\boldsymbol{\lambda}_{\mathbf{min}}(V_{1,k}^{\mathrm{T}}\bar{\phi}_k^{\mathrm{T}}\bar{\phi}_kV_{1,k}), (1-\lambda)\boldsymbol{\lambda}_{\mathbf{min}}(V_{2,k}^{\mathrm{T}}R_k^{\perp}V_{2,k})\}. \end{aligned} \tag{C.10}$$

Next, applying the results of Claims 1 and 2, it follows that

$$\begin{aligned} \boldsymbol{\lambda}_{\mathbf{min}}(R_{k+1}) &\geq \lambda\boldsymbol{\lambda}_{\mathbf{min}}(R_k) \\ &\quad + \min\{\sigma_{q_k,k}^2, (1-\lambda)\boldsymbol{\lambda}_{\boldsymbol{n-q_k}}(R_k^{\perp})\}. \end{aligned} \tag{C.11}$$

Note that, by construction of q_k, it follows that $\sigma_{q_k,k} \geq \sqrt{\varepsilon}$. Moreover, by Lemma 1, $\boldsymbol{\lambda}_{\boldsymbol{n-q_k}}(R_k^{\perp}) \geq \boldsymbol{\lambda}_{\mathbf{min}}(R_k)$. Applying these two inequalities to (C.11), it follows that

$$\begin{aligned} \boldsymbol{\lambda}_{\mathbf{min}}(R_{k+1}) &\geq \lambda\boldsymbol{\lambda}_{\mathbf{min}}(R_k) + \min\{\varepsilon, (1-\lambda)\boldsymbol{\lambda}_{\mathbf{min}}(R_k)\} \\ &= \min\{\varepsilon + \lambda\boldsymbol{\lambda}_{\mathbf{min}}(R_k), \boldsymbol{\lambda}_{\mathbf{min}}(R_k)\}. \end{aligned} \tag{C.12}$$

Next, substituting inductive hypothesis (34) into (C.12), it follows that

$$\boldsymbol{\lambda}_{\mathbf{min}}(R_{k+1}) \geq \min\Big\{\varepsilon + \lambda(\frac{\varepsilon}{1-\lambda}), \varepsilon + \lambda\boldsymbol{\lambda}_{\mathbf{min}}(R_0), \frac{\varepsilon}{1-\lambda}, \boldsymbol{\lambda}_{\mathbf{min}}(R_0)\Big\}. \tag{C.13}$$

Note that $\varepsilon + \lambda(\frac{\varepsilon}{1-\lambda})$ simplifies to $\frac{\varepsilon}{1-\lambda}$. Hence, (C.13) can be rewritten as

$$\boldsymbol{\lambda}_{\mathbf{min}}(R_{k+1}) \geq \min\{\varepsilon + \lambda\boldsymbol{\lambda}_{\mathbf{min}}(R_0), \frac{\varepsilon}{1-\lambda}, \boldsymbol{\lambda}_{\mathbf{min}}(R_0)\}. \tag{C.14}$$

In the case where $\boldsymbol{\lambda}_{\mathbf{min}}(R_0) \geq \frac{\varepsilon}{1-\lambda}$, it follows that

$$\varepsilon + \lambda\boldsymbol{\lambda}_{\mathbf{min}}(R_0) \geq \varepsilon + \lambda(\frac{\varepsilon}{1-\lambda}) = (\frac{\varepsilon}{1-\lambda}).$$

In the case where $\boldsymbol{\lambda}_{\mathbf{min}}(R_0) < \frac{\varepsilon}{1-\lambda}$, it follows that $\varepsilon - (1-\lambda)\boldsymbol{\lambda}_{\mathbf{min}}(R_0) > 0$, and hence

$$\begin{aligned}\varepsilon + \lambda\boldsymbol{\lambda}_{\mathbf{min}}(R_0) &= \boldsymbol{\lambda}_{\mathbf{min}}(R_0) + \varepsilon - (1-\lambda)\boldsymbol{\lambda}_{\mathbf{min}}(R_0) \\ &> \boldsymbol{\lambda}_{\mathbf{min}}(R_0).\end{aligned}$$

Combining these two cases implies that

$$\varepsilon + \lambda\boldsymbol{\lambda}_{\mathbf{min}}(R_0) \geq \min\{\frac{\varepsilon}{1-\lambda}, \boldsymbol{\lambda}_{\mathbf{min}}(R_0)\}. \tag{C.15}$$

Substituting (C.15) into (C.14) yields

$$\boldsymbol{\lambda}_{\mathbf{min}}(R_{k+1}) \geq \min\{\frac{\varepsilon}{1-\lambda}, \boldsymbol{\lambda}_{\mathbf{min}}(R_0)\}. \tag{C.16}$$

Therefore, by induction, (34) holds for all $k \geq 0$.

Next, we will show the induction step for statement *(2)* of Theorem 3. Assume that $(\phi_k)_{k=0}^{\infty}$ is bounded with upper bound $\beta \in (0,\infty)$ and assume, for inductive hypothesis, that (36) holds for $k \geq 0$. Next, applying Lemma A.3 to (C.7), it follows that

$$\boldsymbol{\lambda}_{\mathbf{max}}(V_k^{\mathrm{T}} R_{k+1} V_k) \leq \lambda\boldsymbol{\lambda}_{\mathbf{max}}(V_k^{\mathrm{T}} R_k V_k) + \max\{\boldsymbol{\lambda}_{\mathbf{max}}(V_{1,k}^{\mathrm{T}} \bar{\phi}_k^{\mathrm{T}} \bar{\phi}_k V_{1,k}), (1-\lambda)\boldsymbol{\lambda}_{\mathbf{max}}(V_{2,k}^{\mathrm{T}} R_k^{\perp} V_{2,k})\}. \tag{C.17}$$

By analogous reasoning to that used previously, we use the fact that similarity transformations preserve eigenvalues, Lemma 1, and Claims 1 and 2 to simplify (C.17) to

$$\begin{aligned}\boldsymbol{\lambda}_{\mathbf{max}}(R_{k+1}) &\leq \lambda\boldsymbol{\lambda}_{\mathbf{max}}(R_k) \\ &\quad + \max\{\sigma_{1,k}^2, (1-\lambda)\boldsymbol{\lambda}_{\mathbf{max}}(R_k)\}.\end{aligned} \tag{C.18}$$

Moreover since $(\phi_k)_{k=0}^{\infty}$ is bounded with upper bound β, it follows from Definition 2 that $\sigma_{1,k}^2 \leq \beta$. Thus, it follows from (C.18) that

$$\boldsymbol{\lambda}_{\mathbf{max}}(R_{k+1}) \leq \max\{\beta + \lambda\boldsymbol{\lambda}_{\mathbf{max}}(R_k), \boldsymbol{\lambda}_{\mathbf{max}}(R_k)\}. \tag{C.19}$$

Next, substituting inductive hypothesis (36) into (C.19) yields

$$\boldsymbol{\lambda}_{\mathbf{max}}(R_{k+1}) \leq \max\Big\{\beta + \lambda(\frac{\beta}{1-\lambda}), \beta + \lambda\boldsymbol{\lambda}_{\mathbf{max}}(R_0), \frac{\beta}{1-\lambda}, \boldsymbol{\lambda}_{\mathbf{max}}(R_0)\Big\}. \tag{C.20}$$

Then, by similar reasoning to before, it can be shown that (C.20) simplifies to

$$\boldsymbol{\lambda}_{\mathbf{max}}(R_{k+1}) \leq \max\{\frac{\beta}{1-\lambda}, \boldsymbol{\lambda}_{\mathbf{max}}(R_0)\}. \tag{C.21}$$

Therefore, by induction, (36) holds for all $k \geq 0$. ■

D Comparison to [35]

In [35], a different vector measurement extension of [5] is presented using an oblique projection decomposition. A tuning parameter $\varepsilon > 0$ is chosen and, for all $k \geq 0$, if $\|\phi_k\|_2 \geq \varepsilon$, then the information matrix $R_k \in \mathbb{R}^{n\times n}$ is updated as

$$R_{k+1} = R_k - (1-\lambda)R_k\phi_k^{\mathrm{T}}(\phi_k R_k \phi_k^{\mathrm{T}})^{+}\phi_k R_k + \phi_k^{\mathrm{T}}\phi_k, \tag{D.1}$$

and if $\|\phi_k\|_2 < \varepsilon$, then $R_{k+1} = R_k + \phi_k^{\mathrm{T}}\phi_k$. It is shown in Theorem 3 of [35] that, for all $k \geq 0$, R_k is positive definite. However, we show that there may not exist $\gamma > 0$ such that, for all $k \geq 0$, $\boldsymbol{\lambda}_{\mathbf{min}}(R_k) \succeq \gamma$. In particular, the following example shows that R_k may converge to a positive semidefinite matrix. Hence, Theorem 3 of SIFt-RLS provides stronger bounds on the eigenvalues of the information matrix than [35] does.

Let $\varepsilon > 0$. Let $R_0 = \mathrm{diag}(\varepsilon^2, 1)$ and, for all $k \geq 0$, let $\phi_k = \mathrm{diag}(\varepsilon, \lambda^{\frac{k+1}{2}})$. Then, for all $k \geq 0$, $\|\phi_k\|_2 \geq \varepsilon$. Moreover, for all $k \geq 0$, ϕ_k is nonsingular, hence (D.1) simplifies to

$$R_{k+1} = \lambda R_k + \phi_k^{\mathrm{T}}\phi_k = \lambda R_k + \begin{bmatrix} \varepsilon^2 & 0 \\ 0 & \lambda^{k+1} \end{bmatrix}. \tag{D.2}$$

Therefore, for all $k \geq 0$,

$$R_k = \begin{bmatrix} \varepsilon^2 \sum_{i=0}^{k} \lambda^i & 0 \\ 0 & (k+1)\lambda^k \end{bmatrix}. \tag{D.3}$$

This implies that $\lim_{k\to\infty} R_k = \mathrm{diag}(\frac{\varepsilon^2}{1-\lambda}, 0)$. Therefore, there does not exist $\gamma > 0$ such that, for all $k \geq 0$, $\boldsymbol{\lambda}_{\mathbf{min}}(R_k) \succeq \beta$.