IEEE Control Systems Letters paper presented at
2023 American Control Conference (ACC)
San Diego, CA, USA. May 31 - June 2, 2023

Exponential Resetting and Cyclic Resetting Recursive Least Squares

Brian Lai and Dennis S. Bernstein

***Abstract*— We present two extensions of recursive least squares (RLS) with exponential forgetting (EF), namely, exponential resetting (ER) RLS and cyclic resetting (CR) RLS. Both methods guarantee that the covariance matrix is bounded above and below in the absence of persistent excitation. Under zero excitation, ER-RLS guarantees convergence of the covariance matrix P_k to a user-designed positive-definite matrix P_∞. However, ER-RLS is more computationally complex than EF-RLS. In contrast, CR-RLS has the same computational complexity as EF-RLS while guaranteeing that, under zero excitation, the difference between the covariance matrix P_k and P_∞ is asymptotically bounded. A numerical example shows that ER-RLS and CR-RLS both perform nearly identically to EF-RLS under persistent excitation while protecting against covariance windup when persistent excitation is lost.**

I. Introduction

Recursive least squares (RLS) is a widely used algorithm for online parameter estimation, which recursively updates the minimzier of the least squares regression with Tikhonov regularization problem, also known as Ridge regression [1]. We denote $\theta_k \in \mathbb{R}^n$ to be the n estimated parameters at step k, and $y_k \in \mathbb{R}^p$ the p measurements at step k, for an accumulated kp measurements by step k. Typically, $n \gg p$ [2], in which case the complexity of RLS is $\mathcal{O}(pn^2)$ per step. In the unusual case $p \geq n$, the complexity of RLS is still $\mathcal{O}(pn^2)$ per step.

In RLS, the covariance matrix of θ_k is a positive-definite matrix denoted $P_k \in \mathbb{R}^{n\times n}$. A serious drawback of RLS is the fact that the eigenvalues of P_k decrease over each step and may become arbitrarily small, resulting in a loss of adaptation alertness after a large amount of data has been collected [3], [4]. A common technique for speeding up adaptation is to introduce a forgetting factor $\lambda \in (0,1)$, which exponentially discounts old information [5], [6, section 2.2.3], often referred to as exponential forgetting (EF) RLS [7, p. 53]. EF-RLS maintains $\mathcal{O}(pn^2)$ complexity per step.

While this addresses the issue of sluggish adaptation, a critical issue that arises is, without persistent excitation, at least one of the eigenvalues of P_k becomes arbitrarily large [8], a phenomenon known as covariance blow up [9], covariance windup, or estimator windup [7, p. 473]. Hence, it has been long accepted that one of the most important properties of RLS variations is a guaranteed upper and lower bound for P_k in the absence of persistent excitation [10]. In fact, [11] shows these bounds ensure the estimation error is bounded, normalized prediction errors are square summable, and incremental changes in parameter estimates converge to zero.

Another proposed property for RLS variations is *resetting*, where, under zero excitation, the covariance matrix P_k converges (resets) to a desired positive-definite matrix $P_\infty \in \mathbb{R}^{n\times n}$. Resetting and guaranteed covariance bounds are achieved in [10] through an algorithm inspired by analysis in continuous time. However, P_∞ cannot be selected directly and is difficult to tune. Furthermore, the time complexity per step is $\mathcal{O}(n^3)$ when $n \gg p$ since the square of the covariance matrix must be computed. Another algorithm with the resetting property is *covariance resetting* [12], where the covariance matrix is reinitialized to a desired value when it becomes too small or at preset times. While this adds little computational cost and is easily implemented, its performance is unclear [4].

A recent method [13] achieves bounded covariance, resetting, and direct selection of P_∞. However, their method is limited to the case $p = 1$ and $P_\infty = \frac{1-\lambda}{\delta} I_n$, where $\delta > 0$ is a design parameter. We generalize [13] to $p > 1$ and arbitrary positive-definite P_∞, an algorithm we call *exponential resetting RLS* (ER-RLS). Interestingly, the covariance update in ER-RLS matches the *back-to-prior forgetting* developed from a Bayesian perspective in [14]. A drawback of ER-RLS is $\mathcal{O}(n^3)$ complexity per step when $n \gg p$.

This paper proposes a novel algorithm called *cyclic resetting RLS* (CR-RLS), which has guaranteed covariance bounds, runs in $\mathcal{O}(pn^2)$ time per step for any $p \geq 1$, has similar resetting properties as exponential resetting RLS, and allows for direct selection of P_∞. The tradeoff for $\mathcal{O}(pn^2)$ complexity is, under zero excitation, only the subsequence of the covariance matrix every n steps converges to P_∞. The sequence of the covariance matrix at every step may oscillate under zero excitation, but remains bounded close to P_∞. CR-RLS should only be used in place of ER-RLS when $n \gg p$.

Notation: For $A \in \mathbb{R}^{n\times n}$, $\lambda_i(A)$ denotes the ith largest eigenvalue of A, $\boldsymbol{\lambda}_{\mathbf{max}}(A) \triangleq \lambda_1(A)$, and $\boldsymbol{\lambda}_{\mathbf{min}}(A) \triangleq \lambda_n(A)$. For $P, Q \in \mathbb{R}^{n\times n}$, let $P \preceq Q$ denote that $Q - P$ is positive semidefinite.

II. Exponential Forgetting (EF) RLS

The exponential forgetting recursive least squares (EF-RLS) algorithm is given by Proposition 1. With $\lambda = 1$, Proposition 1 gives RLS without forgetting.

Proposition 1. *For all $k \geq 0$, let $\phi_k \in \mathbb{R}^{p\times n}$ and $y_k \in \mathbb{R}^p$. Let $\theta_0 \in \mathbb{R}^n$, $R_0 \in \mathbb{R}^{n\times n}$ be positive definite, and $\lambda \in (0,1)$.*

Brian Lai and Dennis S. Bernstein are with the Department of Aerospace Engineering, University of Michigan, Ann Arbor, MI, USA. {brianlai, dsbaero}@umich.edu

For all $k \geq 0$, denote the minimizer of the function

$$J_k(\hat{\theta}) = \sum_{i=0}^{k} \lambda^{k-i}(y_i - \phi_i\hat{\theta})^{\mathrm{T}}(y_i - \phi_i\hat{\theta}) + \lambda^{k+1}(\hat{\theta} - \theta_0)^{\mathrm{T}} R_0(\hat{\theta} - \theta_0) \tag{1}$$

by $\theta_{k+1} \triangleq \arg\min_{\hat{\theta}\in\mathbb{R}^n} J_k(\hat{\theta})$. Then, for all $k \geq 0$,

$$R_{k+1} = \lambda R_k + \phi_k^{\mathrm{T}}\phi_k, \tag{2}$$

$$\theta_{k+1} = \theta_k + P_{k+1}\phi_k^{\mathrm{T}}(y_k - \phi_k\theta_k), \tag{3}$$

where, for all $k \geq 0$, $P_k \in \mathbb{R}^{n\times n}$ and $R_k \in \mathbb{R}^{n\times n}$ are positive definite and $R_k \triangleq P_k^{-1}$.

Proof. See [2]. □

We call $P_k \in \mathbb{R}^{n\times n}$ the covariance matrix and $R_k = P_k^{-1} \in \mathbb{R}^{n\times n}$ the information matrix. Note that when $n \gg p$, it is computationally beneficial to use the matrix inversion lemma (Lemma A.1) to rewrite (2) as

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

A. Summary and Computational Cost

Algorithm 1 Exponential Forgetting RLS (EF-RLS)

Initialize: $\lambda \in (0,1)$, $\theta_0 \in \mathbb{R}^n$, positive-definite $P_0 \in \mathbb{R}^{n\times n}$
for all $k \geq 0$ **do**
  if $n \gg p$ **then**
    $L_k \leftarrow P_k\phi_k^{\mathrm{T}}$ ▷ $\mathcal{O}(pn^2)$
    $P_{k+1} \leftarrow \frac{1}{\lambda}P_k - \frac{1}{\lambda}L_k(\lambda I_p + \phi_k L_k)^{-1}L_k^{\mathrm{T}}$ ▷ $\mathcal{O}(pn^2)$
    $\theta_{k+1} \leftarrow \theta_k + P_{k+1}\phi_k^{\mathrm{T}}(y_k - \phi_k\theta_k)$ ▷ $\mathcal{O}(pn^2)$
  else
    $R_{k+1} \leftarrow \lambda R_k + \phi_k^{\mathrm{T}}\phi_k$ ▷ $\mathcal{O}(pn^2)$
    $\theta_{k+1} \leftarrow \theta_k + R_{k+1}^{-1}\phi_k^{\mathrm{T}}(y_k - \phi_k\theta_k)$ ▷ $\mathcal{O}(n^3)$

EF-RLS is summarized by Algorithm 1. We introduce $L_k \in \mathbb{R}^{n\times p}$ to optimize computational efficiency when $n \gg p$. The complexity of EF-RLS is $\mathcal{O}(pn^2)$ in both variations presented ($n \gg p$ and $p \geq n$).

B. Information Matrix Bounds

Proposition 2 shows that, under persistent excitation, R_k in EF-RLS is lower bounded. Proposition 3 show that if, for all $k \geq 0$, the matrix $\phi_k^{\mathrm{T}}\phi_k$ is upper bounded, then R_k in EF-RLS is upper bounded.

Definition 1. *A sequence $(\phi_k)_{k=0}^{\infty} \subset \mathbb{R}^{p\times n}$ is* persistently exciting (PE) *if there exist $N \geq 1$ and $\alpha > 0$ such that, for all $k \geq 0$, $\alpha I_n \preceq \sum_{i=k}^{k+N}\phi_i^{\mathrm{T}}\phi_i$. α and N are, respectively, the "lower bound" and "persistency window" of $(\phi_k)_{k=0}^{\infty}$.*

Proposition 2. *(EF-RLS) If $(\phi_k)_{k=0}^{\infty}$ is PE with lower bound α and persistency window N, then, for all $k \geq N+1$,*

$$R_k \succeq \frac{\lambda^N(1-\lambda)}{1-\lambda^{N+1}}\alpha I_n. \tag{5}$$

Proof. See Proposition 4 in [15]. □

Proposition 3. *(EF-RLS) If there exists $\beta \geq 0$ such that, for all $k \geq 0$, $\phi_k^{\mathrm{T}}\phi_k \preceq \beta I$, then, for all $k \geq 0$,*

$$R_k \preceq \lambda^k R_0 + \frac{1-\lambda^k}{1-\lambda}\beta I_n. \tag{6}$$

Proof. See Proposition 1 in [13]. □

C. A new interpretation to forgetting

A well-known property of EF-RLS is covariance windup, where, without persistent excitation, the eigenvalues of the covariance matrix become unbounded [8]. In particular, with no excitation, the information matrix approaches $0_{n\times n}$. This is shown in Proposition 4.

Proposition 4. *(EF-RLS) If there exists an integer N such that, for all $k \geq N$, $\phi_k = 0$, then*

$$\lim_{k\to\infty} R_k = 0_{n\times n}. \tag{7}$$

Proof. Note that by repeated substitution of (2), it follows that, for all $k \geq 0$, $R_{k+N} = \lambda^k R_N$. Hence, (7) follows. □

To provide insight into why covariance windup occurs, note that, for all $k \geq 0$, (2) can be written as

$$R_{k+1} = R_k - (1-\lambda)(R_k - 0_{n\times n}) + \phi_k^{\mathrm{T}}\phi_k. \tag{8}$$

Notice that, for all $k \geq 0$, $(1-\lambda)(R_k - 0_{n\times n})$ is positive definite and $\phi_k^{\mathrm{T}}\phi_k$ is positive semidefinite. Hence, at each step $k \geq 0$, we see a decrease in the information matrix, R_k, proportional to $(R_k - 0_{n\times n})$ and an increase of $\phi_k^{\mathrm{T}}\phi_k$.

III. Exponential Resetting (ER) RLS

Motivated by (8), we keep the θ_k update (3) but consider a new information matrix update where, for all $k \geq 0$,

$$R_{k+1} = R_k - (1-\lambda)(R_k - R_\infty) + \phi_k^{\mathrm{T}}\phi_k, \tag{9}$$

where $R_\infty \in \mathbb{R}^{n\times n}$ is positive definite. Note that (9) can be rewritten as

$$R_{k+1} = \lambda R_k + (1-\lambda)R_\infty + \phi_k^{\mathrm{T}}\phi_k \tag{10}$$

We call this algorithm exponential resetting recursive least squares (ER-RLS). Note that [13] is equivalent to ER-RLS with $p = 1$ and $R_\infty = \frac{\delta}{1-\lambda}I_n$, where $\delta \geq 0$ is a design parameter. As shown by Proposition 5, ER-RLS retains the important property that, for all $k \geq 0$, R_k is positive definite.

Proposition 5. *(ER-RLS) If R_0 is positive definite, then, for all $k \geq 0$, R_k is positive definite.*

Proof. Proof by induction: R_0 is positive definite by assumption. Next, note that if R_k is positive definite, it follows that λR_k is positive definite. Moreover, $(1-\lambda)R_\infty$ is positive definite and $\phi_k^{\mathrm{T}}\phi_k$ is positive semidefinite. Hence, by (10), R_{k+1} is positive definite. □

A. Summary and Computational Cost

ER-RLS is summarized by Algorithm 2. Note that since the terms λR_k and $(1-\lambda)R_\infty$ in (9) are both rank n, using the matrix inversion lemma would not improve the computational cost to compute R_{k+1}^{-1}. Hence, the computational complexity of ER-RLS is $\mathcal{O}(\max\{n^3, pn^2\})$ per step.

Algorithm 2 Exponential Resetting RLS (ER-RLS)

Initialize: $\lambda \in (0,1)$, $\theta_0 \in \mathbb{R}^n$, positive-definite $P_0 \in \mathbb{R}^{n\times n}$, positive-definite $R_\infty \in \mathbb{R}^{n\times n}$

for all $k \geq 0$ **do**

$R_{k+1} \leftarrow \lambda R_k + (1-\lambda)R_\infty + \phi_k^{\mathrm{T}}\phi_k$ ▷ $\mathcal{O}(pn^2)$

$\theta_{k+1} \leftarrow \theta_k + R_{k+1}^{-1}\phi_k^{\mathrm{T}}(y_k - \phi_k\theta_k)$ ▷ $\mathcal{O}(n^3)$

B. Information Matrix Lower Bound

Proposition 6 gives a tight lower bound for R_k which depends on the step k. Corollary 1 gives a weaker lower bound that is valid for all $k \geq 0$. Note that persistent excitation is not necessary for a guaranteed R_k lower bound.

Proposition 6. *(ER-RLS) For all* $k \geq 0$,

$$R_k \succeq \lambda^k R_0 + (1-\lambda^k)R_\infty. \tag{11}$$

Proof. By repeated substitution of (9), it follows that, for all $k \geq 0$,

$$R_k = \lambda^k R_0 + (1-\lambda^k)R_\infty + \sum_{i=0}^{k-1}\lambda^i\phi_{k-i-1}^{\mathrm{T}}\phi_{k-i-1}.$$

Then, since $\sum_{i=0}^{k-1}\lambda^i\phi_{k-i-1}^{\mathrm{T}}\phi_{k-i-1}$ is positive semidefinite, (11) follows. ☐

Corollary 1. *(ER-RLS) For all* $k \geq 0$,

$$\boldsymbol{\lambda}_{\mathbf{min}}(R_k) \geq \min\{\boldsymbol{\lambda}_{\mathbf{min}}(R_0), \boldsymbol{\lambda}_{\mathbf{min}}(R_\infty)\}. \tag{12}$$

Proof. Define $r_{\min} \triangleq \min\{\boldsymbol{\lambda}_{\mathbf{min}}(R_0), \boldsymbol{\lambda}_{\mathbf{min}}(R_\infty)\}$. Applying Lemma A.2 to (11), it follows that, for all $k \geq 0$,

$$\begin{aligned}\boldsymbol{\lambda}_{\mathbf{min}}(R_k) &\geq \lambda^k\boldsymbol{\lambda}_{\mathbf{min}}(R_0) + (1-\lambda^k)\boldsymbol{\lambda}_{\mathbf{min}}(R_\infty),\\ &\geq \lambda^k r_{\min} + (1-\lambda^k)r_{\min} = r_{\min}.\end{aligned}$$ ☐

C. Information Matrix Upper Bound

Proposition 7 gives a tight upper bound for R_k which depends on the step k. Corollary 2 gives a weaker upper bound that is valid for all $k \geq 0$.

Proposition 7. *(ER-RLS) If there exists* $\beta \geq 0$ *such that, for all* $k \geq 0$, $\phi_k^{\mathrm{T}}\phi_k \preceq \beta I$, *then, for all* $k \geq 0$,

$$R_k \preceq \lambda^k R_0 + (1-\lambda^k)R_\infty + \frac{1-\lambda_k}{1-\lambda}\beta I_n. \tag{13}$$

Proof. By repeated substitution of (9), for all $k \geq 0$,

$$R_k = \lambda^k R_0 + (1-\lambda^k)R_\infty + \sum_{i=0}^{k-1}\lambda^i\phi_{k-i-1}^{\mathrm{T}}\phi_{k-i-1}.$$

Since $\sum_{i=0}^{k-1}\lambda^i\phi_{k-i-1}^{\mathrm{T}}\phi_{k-i-1} \preceq \sum_{i=0}^{k-1}\lambda^i\beta I_n$, (13) follows by geometric series. ☐

Corollary 2. *(ER-RLS) If there exists* $\beta \geq 0$ *such that, for all* $k \geq 0$, $\phi_k^{\mathrm{T}}\phi_k \preceq \beta I$, *then, for all* $k \geq 0$,

$$\boldsymbol{\lambda}_{\mathbf{max}}(R_k) \preceq \max\{\boldsymbol{\lambda}_{\mathbf{max}}(R_0), \boldsymbol{\lambda}_{\mathbf{max}}(R_\infty)\} + \frac{\beta}{1-\lambda}. \tag{14}$$

Proof. Define $r_{\max} \triangleq \max\{\boldsymbol{\lambda}_{\mathbf{max}}(R_0), \boldsymbol{\lambda}_{\mathbf{max}}(R_\infty)\}$. Note that, for all $k \geq 0$,

$$\begin{aligned}&\lambda^k\boldsymbol{\lambda}_{\mathbf{max}}(R_0) + \left(1-\lambda^k\right)\boldsymbol{\lambda}_{\mathbf{max}}(R_\infty) + \frac{1-\lambda^k}{1-\lambda}\beta\\ &\leq \lambda^k r_{\max} + \left(1-\lambda^k\right)r_{\max} + \frac{\beta}{1-\lambda} = r_{\max} + \frac{\beta}{1-\lambda}.\end{aligned}$$

Hence, by Lemma A.2, (14) follows from (13). ☐

D. Resetting Property

Proposition 8 gives the resetting property of ER-RLS, which states that, under no excitation, the sequence $\{R_k\}_{k=0}^\infty$ converges to R_∞.

Proposition 8. *(ER-RLS) If there exists an integer* $N > 0$ *such that, for all* $k \geq N$, $\phi_k = 0$, *then*

$$\lim_{k\to\infty} R_k = R_\infty. \tag{15}$$

Proof. By repeated substitution of (9), it follows that, for all $k \geq 0$, $R_{k+N} = \lambda^k R_N + (1-\lambda^k)R_\infty$, implying (15). ☐

IV. Cyclic Resetting (CR) RLS

The motivation for cyclic resetting recursive least squares (CR-RLS) is to develop an RLS extension with guaranteed covariance bounds and similar resetting properties to ER-RLS, but which runs in $\mathcal{O}(pn^2)$ time complexity per step.

To begin, we choose a positive-definite $R_\infty \in \mathbb{R}^{n\times n}$ and write an orthogonal diagonalization of R_∞ as

$$R_\infty = V_\infty D_\infty V_\infty^{\mathrm{T}}, \tag{16}$$

where $V_\infty \in \mathbb{R}^{n\times n}$ is orthogonal and $D_\infty \in \mathbb{R}^{n\times n}$ is diagonal. Next, we write V_∞ and D_∞ as

$$V_\infty = \begin{bmatrix} v_{\infty,0} & \cdots & v_{\infty,n-1}\end{bmatrix}, \tag{17}$$

$$D_\infty = \operatorname{diag}\left(\begin{bmatrix} d_{\infty,0} & \cdots & d_{\infty,n-1}\end{bmatrix}\right), \tag{18}$$

where, for all $i = 0,\ldots,n-1$, $v_{\infty,i} \in \mathbb{R}^n$ is an eigenvector of R_∞ and $d_{\infty,i} \in \mathbb{R}_{>0}$ is its associated eigenvector. Next, for all $i = 0,\ldots,n-1$, define $R_{\infty,i} \in \mathbb{R}^{n\times n}$ by

$$R_{\infty,i} \triangleq d_{\infty,i}v_{\infty,i}v_{\infty,i}^{\mathrm{T}}. \tag{19}$$

Note that, for all $i = 0,\ldots,n-1$, $\operatorname{rank}(R_{\infty,i}) = 1$ and $R_{\infty,i}$ is positive semidefinite. Furthermore,

$$R_\infty = R_{\infty,0} + \cdots + R_{\infty,n-1}. \tag{20}$$

Then, for all $k \geq 0$, we consider the following R_k update:

$$R_{k+1} = \lambda R_k + \frac{1-\lambda^n}{\lambda^{n-(k \bmod n)-1}}R_{\infty,(k \bmod n)} + \phi_k^{\mathrm{T}}\phi_k, \tag{21}$$

It then follows by repeated substitution of (21) that, for all $k \geq 0$, and for all $i = 1,\ldots,n$

$$R_{kn+i} = \lambda^i R_{kn} + \frac{1-\lambda^n}{\lambda^{n-i}}\sum_{j=0}^{i}R_{\infty,j} + \sum_{j=1}^{i}\lambda^{j-1}\phi_{kn+i-j}^{\mathrm{T}}\phi_{kn+i-j}, \tag{22}$$

and hence, for all $k \geq 0$,

$$R_{kn+n} = \lambda^n R_{kn} + (1-\lambda^n) R_\infty + \sum_{j=1}^{n} \lambda^{j-1} \phi_{kn+n-j}^{\mathrm{T}} \phi_{kn+n-j}. \tag{23}$$

Note that the n step update for CR-RLS, given by (23), is structured similarly to the 1 step update for ER-RLS, given by (10). Proposition 9 shows the important property that, for all $k \geq 0$, R_k is positive definite.

Proposition 9. *(CR-RLS) If R_0 is positive definite, then, for all $k \geq 0$, R_k is positive definite.*

Proof. Proof by induction: R_0 is positive definite by assumption. Next, if R_k is positive definite, it follows that $\lambda^n R_k$ is positive definite. Moreover, $\frac{1-\lambda^n}{\lambda^{n-(k \bmod n)-1}} R_{\infty,(k \bmod n)} + \phi_k^{\mathrm{T}} \phi_k$ is positive semidefinite. Hence, it follows from (21) that R_{k+1} is positive definite. □

A. Matrix Inversion Lemma for Efficient Computation

The main improvement of CR-RLS over ER-RLS is the ability to use the matrix inversion lemma (Lemma A.1) for efficient computation when $n \gg p$. First, note that, for all $k \geq 0$, (21) can be written as

$$R_{k+1} = \lambda R_k + \bar{\phi}_k^{\mathrm{T}} \bar{\phi}_k, \tag{24}$$

where

$$\bar{\phi}_k \triangleq \begin{bmatrix} \phi_k \\ \bar{v}_{\infty,(k \bmod n)}^{\mathrm{T}} \end{bmatrix} \in \mathbb{R}^{(p+1)\times n}, \tag{25}$$

and, for all $i = 0, \ldots, n-1$,

$$\bar{v}_{\infty,i} \triangleq \sqrt{\frac{(1-\lambda^n) d_{\infty,i}}{\lambda^{n-i-1}}}\, v_{\infty,i} \in \mathbb{R}^n. \tag{26}$$

Then, it follows from Lemma A.1 that, for all $k \geq 0$,

$$P_{k+1} = \frac{1}{\lambda} P_k - \frac{1}{\lambda} P_k \bar{\phi}_k^{\mathrm{T}} (\lambda I_{p+1} + \bar{\phi}_k P_k \bar{\phi}_k^{\mathrm{T}})^{-1} \bar{\phi}_k P_k, \tag{27}$$

where $P_k \triangleq R_k^{-1}$,

B. Summary and Computational Cost

Algorithm 3 Cyclic Resetting RLS (CR-RLS)

Use CR-RLS only when $n \gg p$. Otherwise, use ER-RLS.
Initialize: $\lambda \in (0,1)$, $\theta_0 \in \mathbb{R}^n$, positive-definite $P_0 \in \mathbb{R}^{n\times n}$, positive-definite $R_\infty \in \mathbb{R}^{n\times n}$
Precompute $\bar{v}_{\infty,0}, \ldots, \bar{v}_{\infty,n-1} \in \mathbb{R}^n$ ▷ See (26), (16)
for all $k \geq 0$ **do**
  $\bar{\phi}_k \leftarrow \begin{bmatrix} \phi_k^{\mathrm{T}} & \bar{v}_{\infty,(k \bmod n)} \end{bmatrix}^{\mathrm{T}}$
  $\bar{L}_k \leftarrow P_k \bar{\phi}_k^{\mathrm{T}}$ ▷ $\mathcal{O}(pn^2)$
  $P_{k+1} \leftarrow \frac{1}{\lambda} P_k - \frac{1}{\lambda} \bar{L}_k (\lambda I_{p+1} + \bar{\phi}_k \bar{L}_k)^{-1} \bar{L}_k^{\mathrm{T}}$ ▷ $\mathcal{O}(pn^2)$
  $\theta_{k+1} \leftarrow \theta_k + P_{k+1} \phi_k^{\mathrm{T}} (y_k - \phi_k \theta_k)$ ▷ $\mathcal{O}(pn^2)$

CR-RLS is summarized by Algorithm 3. We introduce $\bar{L}_k \in \mathbb{R}^{n\times(p+1)}$ to optimize computational efficiency. The computational complexity of CR-RLS is $\mathcal{O}(pn^2)$ per step. In the case where $n \gg p$, CR-RLS provides similar bounds and resetting properties as ER-RLS with a significant improvement over the $\mathcal{O}(n^3)$ complexity per step of ER-RLS. If $p \geq n$, we recommend using ER-RLS instead of CR-RLS.

C. Information Matrix Lower Bound

Theorem 1 gives a tight lower bound for R_k dependent on the step k. Corollary 3 gives a weaker lower bound that is valid for all $k \geq 0$. We begin with a useful Lemma.

Lemma 1. *(CR-RLS) For all $k \geq 0$ and all $i = 1, \ldots, n-1$,*

$$R_{kn+i} \succeq \lambda^i R_{kn}. \tag{28}$$

Proof. It follows from repeated substitution of (2) that, for all $k \geq 0$, and for all $i = 1, \ldots, n-1$,

$$R_{kn+i} = \lambda^i R_{kn} + \sum_{j=0}^{i-1} \frac{1-\lambda^n}{\lambda^{n-i}} R_{\infty,j} + \lambda^j \phi_{k+n-j-1}^{\mathrm{T}} \phi_{k+n-j-1}.$$

Subtracting $\lambda^i R_{kn}$ from both sides, it follows that $R_{kn+i} - \lambda^i R_{kn}$ is positive semidefinite. □

Theorem 1. *(CR-RLS) For all $k \geq 0$,*

$$R_k \succeq \lambda^k R_0 + (\lambda^{k \bmod n} - \lambda^k) R_\infty \tag{29}$$

Proof. For all $k \geq 0$, it follows from (23) that $R_{k+n} \succeq \lambda^n R_k + (1-\lambda^n) R_\infty$. It then follows from repeated substitution that, for all $k \geq 0$,

$$R_{kn} \succeq \lambda^{kn} R_0 + (1-\lambda^{kn}) R_\infty.$$

It then follows from Lemma 1 that, for all $k \geq 0$, and for all $i = 0, \ldots, n-1$,

$$R_{kn+i} \succeq \lambda^i R_{kn} \succeq \lambda^i \left[\lambda^{kn} R_0 + (1-\lambda^{kn}) R_\infty\right],$$

which can be rewritten as (29). □

Corollary 3. *(CR-RLS) For all $k \geq 0$,*

$$\boldsymbol{\lambda}_{\mathbf{min}}(R_k) \geq \lambda^{n-1} \min\left\{\boldsymbol{\lambda}_{\mathbf{min}}(R_0), \boldsymbol{\lambda}_{\mathbf{min}}(R_\infty)\right\}. \tag{30}$$

Proof. Define $r_{\min} \triangleq \min\left\{\boldsymbol{\lambda}_{\mathbf{min}}(R_0), \boldsymbol{\lambda}_{\mathbf{min}}(R_\infty)\right\}$. Applying Lemma A.2 to the result of Theorem 1, it follows that, for all $k \geq 0$, $\boldsymbol{\lambda}_{\mathbf{min}}(R_k) \geq \lambda^k \boldsymbol{\lambda}_{\mathbf{min}}(R_0) + (\lambda^{k \bmod n} - \lambda^k) \boldsymbol{\lambda}_{\mathbf{min}}(R_\infty) \geq \lambda^k r_{\min} + (\lambda^{k \bmod n} - \lambda^k) r_{\min} \geq \lambda^{n-1} r_{\min}$.

D. Information Matrix Upper Bound

Theorem 2 gives a tight upper bound for the eigenvalues of R_k that depends on the step k. Corollary 4 gives a weaker upper bound that is valid for all $k \geq 0$.

Theorem 2. *(CR-RLS) If there exists $\beta \geq 0$ such that, for all $k \geq 0$, $\phi_k^{\mathrm{T}} \phi_k \preceq \beta I$, then, for all $k \geq 0$,*

$$R_k \preceq \lambda^k R_0 + \left(\frac{1}{\lambda^{n-(k \bmod n)}} - \lambda^k\right) R_\infty + \frac{1-\lambda^k}{1-\lambda} \beta I_n. \tag{31}$$

Proof. Repeated substitution of (23) implies, for all $k \geq 0$, $R_{kn} = \lambda^{kn} R_0 + (1-\lambda^{kn}) R_\infty + \sum_{j=1}^{kn} \lambda^{j-1} \phi_{kn-j}^{\mathrm{T}} \phi_{kn-j}$. Next, it follows from (22) that, for all $k \geq 0$ and all $i = 0, \ldots, n-1$,

$$R_{kn+i} = \lambda^{kn+i} R_0 + \lambda^i (1-\lambda^{kn}) R_\infty + \frac{1-\lambda^n}{\lambda^{n-i}} \textstyle\sum_{j=0}^{i} R_{\infty,j} + \sum_{j=1}^{kn+i} \lambda^{j-1} \phi_{kn-j}^{\mathrm{T}} \phi_{kn-j},$$

and note that $\sum_{j=0}^{i} R_{\infty,j} \preceq R_\infty$. Hence, for all $k \geq 0$ and all $i = 0, \ldots, n-1$, $R_{kn+i} \preceq \lambda^{kn+i} R_0 + (\lambda^i(1-\lambda^{kn}) + \frac{1-\lambda^n}{\lambda^{n-i}}) R_\infty + \sum_{j=1}^{kn+1} \lambda^{j-1}\beta I$, which simplifies to $R_{kn+i} \preceq \lambda^{kn+i} R_0 + \left(\frac{1}{\lambda^{n-i}} - \lambda^{kn+i}\right) R_\infty + \frac{1-\lambda^{kn+i}}{1-\lambda}\beta I$, which can be rewritten as (31). □

Corollary 4. *(CR-RLS) If there exists $\beta \geq 0$ such that, for all $k \geq 0$, $\phi_k^{\mathrm{T}}\phi_k \preceq \beta I$, then, for all $k \geq 0$,*

$$\boldsymbol{\lambda}_{\mathbf{max}}(R_k) \leq \frac{1}{\lambda^n} \max\{\boldsymbol{\lambda}_{\mathbf{max}}(R_0), \boldsymbol{\lambda}_{\mathbf{max}}(R_\infty)\} + \frac{\beta}{1-\lambda}. \tag{32}$$

Proof. Define $r_{\max} \triangleq \max\{\boldsymbol{\lambda}_{\mathbf{max}}(R_0), \boldsymbol{\lambda}_{\mathbf{max}}(R_\infty)\}$. Note that, for all $k \geq 0$,

$$\lambda^k \boldsymbol{\lambda}_{\mathbf{max}}(R_0) + \left(\frac{1}{\lambda^{n-(k \bmod n)}} - \lambda^k\right) \boldsymbol{\lambda}_{\mathbf{max}}(R_\infty)$$
$$\leq \lambda^k r_{\max} + \left(\frac{1}{\lambda^{n-(k \bmod n)}} - \lambda^k\right) r_{\max} \leq \frac{1}{\lambda^n} r_{\max},$$

and $\frac{1-\lambda^k}{1-\lambda}\beta \leq \frac{\beta}{1-\lambda}$. By Lemma A.2, (31) gives (32). □

E. Cyclic Resetting and Bounded Resetting

Next, Proposition 10 gives the *cyclic resetting* property of CR-RLS, which states that, under no excitation, the sequence $\{R_{kn}\}_{k=0}^\infty$ converges to R_∞.

Proposition 10. *(CR-RLS) If there exists $N > 0$ such that, for all $k \geq N$, $\phi_k = 0$, then*

$$\lim_{k\to\infty} R_{nk} = R_\infty. \tag{33}$$

Proof. Let Kn be the smallest multiple of n larger than N. By repeated substitution of (23), $R_{(K+k)n}$ can be written as

$$R_{(K+k)n} = \lambda^{kn} R_{Kn} + (1-\lambda^{kn}) R_\infty.$$

Since $\lambda \in (0,1)$, $\lim_{k\to\infty} R_{(K+k)n} = R_\infty$ follows. □

Finally, Theorem 3 and Corollary 5 give the *bounded resetting* property of CR-RLS, which states that, under no excitation, the limit inferior (respectively, limit superior) of the sequence $\{R_k\}_{k=0}^\infty$ is bounded below (respectively, above) by $\lambda^{n-1} R_\infty$ (respectively, $\frac{1}{\lambda^{n-1}} R_\infty$).

Theorem 3. *(CR-RLS) If there exists $N > 0$ such that, for all $k \geq N$, $\phi_k = 0$, then, for all $\varepsilon > 0$, there exists $M \geq 0$ such that, for all $k \geq M$,*

$$\lambda^{n-1} R_\infty - \varepsilon I_n \preceq R_k \preceq \frac{1}{\lambda^{n-1}} R_\infty + \varepsilon I_n. \tag{34}$$

Proof. Define $r_{\max} \triangleq \frac{1}{\lambda^n} \max\{\boldsymbol{\lambda}_{\mathbf{max}}(R_0), \boldsymbol{\lambda}_{\mathbf{max}}(R_\infty)\} + \frac{\beta}{1-\lambda}$, and note that, from Corollary 4, for all $k \geq 0$, $R_k \preceq r_{\max} I$. Let $\varepsilon > 0$. Let $K \geq 0$ be chosen such that $Kn \geq N$. Choose $M \geq 0$ to be an integer such that

$$M \geq \log_\lambda\left(\min\left\{\frac{\varepsilon}{r_{\max}}, \frac{\varepsilon}{\boldsymbol{\lambda}_{\mathbf{max}}(R_\infty)}\right\}\right) + (K+1)n.$$

First, we prove the lower bound. Note that since, for all $k \geq Kn$, $\phi_k = 0$, it follows from repeated substitution of (23) that, for all $k \geq 0$, $R_{(K+k)n} = \lambda^{kn} R_{Kn} + (1-\lambda^{kn}) R_\infty$. Next, by applying Lemma 1, it follows that, for all $k \geq 0$, and for all $i = 0, \ldots, n-1$, $R_{(K+k)n+i} \succeq \lambda^{kn+i} R_{Kn} + \lambda^i(1-\lambda^{kn}) R_\infty$. Furthermore, since, for all $k \geq 0$, R_∞ and R_k are positive definite, it follows that, for all $k \geq 0$ and all $i = 0, \ldots, n-1$,

$$R_{(K+k)n+i} \succeq \lambda^i R_\infty - \lambda^{kn+i} R_\infty$$
$$\succeq \lambda^{n-1} R_\infty - \lambda^{kn} \boldsymbol{\lambda}_{\mathbf{max}}(R_\infty) I_n.$$

If $(K+k)n + i \geq M$, then $kn \geq \log_\lambda(\frac{\varepsilon}{\boldsymbol{\lambda}_{\mathbf{max}}(R_\infty)})$, and thus $\lambda^{kn} \leq \frac{\varepsilon}{\boldsymbol{\lambda}_{\mathbf{max}}(R_\infty)}$. Hence, $R_{(K+k)n+i} \succeq \lambda^{n-1} R_\infty - \varepsilon I_n$.

Second, we prove the upper bound. Since $\lim_{k\to\infty} R_{nk} = R_\infty$, Proposition 10 implies that there exists $M \geq 0$ such that, for all $k \geq M$, and for $i = 0$, $R_{kn+i} \preceq R_\infty + \varepsilon I \preceq \frac{1}{\lambda^{n-1}} R_\infty + \varepsilon I_n$. Next, we address the cases $i = 1, \ldots, n-1$. Since, for all $k \geq Kn$, $\phi_k = 0$, it follows from (22) and repeated substitution of (23) that, for all $k \geq 0$ and all $i = 1, \ldots, n-1$, $R_{(K+k)n+i} = \lambda^{kn+i} R_{Kn} + \lambda^i(1-\lambda^{kn}) R_\infty + \frac{1-\lambda^n}{\lambda^{n-i}} \sum_{j=0}^{i} R_{\infty,j}$. Moreover, since $\sum_{j=0}^{i} R_{\infty,j} \preceq R_\infty$, it follows that, for all $k \geq 0$ and all $i = 1, \ldots, n-1$,

$$R_{(K+k)n+i} \preceq \lambda^{kn+i} R_{Kn} + \left(\lambda^i(1-\lambda^{kn}) + \frac{1-\lambda^n}{\lambda^{n-i}}\right) R_\infty,$$
$$= \lambda^{kn+i} R_{Kn} + \frac{R_\infty}{\lambda^{n-i}} - R_\infty \lambda^{kn+i} \preceq \lambda^{kn} r_{\max} I_n + \frac{R_\infty}{\lambda^{n-1}}.$$

If $(K+k)n + i \geq M$, then $kn \geq \log_\lambda(\frac{\varepsilon}{r_{\max}})$, and thus $\lambda^{kn} \leq \frac{\varepsilon}{r_{\max}}$. Hence, $R_{(K+k)n+i} \preceq \varepsilon I_n + \frac{1}{\lambda^{n-1}} R_\infty$. □

Corollary 5. *(CR-RLS) If there exists $N > 0$ such that, for all $k \geq N$, $\phi_k = 0$, then*

$$\liminf_{k\to\infty} \boldsymbol{\lambda}_{\mathbf{min}}(R_k) \geq \lambda^{n-1} \boldsymbol{\lambda}_{\mathbf{min}}(R_\infty), \tag{35}$$

$$\limsup_{k\to\infty} \boldsymbol{\lambda}_{\mathbf{max}}(R_k) \leq \frac{1}{\lambda^{n-1}} \boldsymbol{\lambda}_{\mathbf{max}}(R_\infty). \tag{36}$$

V. Numerical Example

Consider the following example with $n = 4$ and $p = 2$. Let $\lambda = 0.9$, $\theta_0 = [0\ 0\ 0\ 0]^{\mathrm{T}}$, and $P_0 = P_\infty = I_4$, where $P_\infty \triangleq R_\infty^{-1}$. For all $0 \leq k \leq 500$ and $1000 \leq k \leq 1500$, each row of $\phi_k \in \mathbb{R}^{2\times 4}$ is i.i.d. sampled from $\mathcal{N}(0, I_4)$. However, for all $500 < k < 1000$, each row of ϕ_k is i.i.d. sampled from $\mathcal{N}(0, \frac{1}{10^2} I_4)$. Also, for all $0 \leq k \leq 1500$, let $v_k \in \mathbb{R}^2$ be i.i.d. sampled from $\mathcal{N}(0, I_2)$, and $y_k \in \mathbb{R}^2$ be given by $y_k = \phi_k \theta_{\mathrm{true},k} + v_k$, where $\theta_{\mathrm{true},k} = [1\ 1\ \sin\frac{\pi k}{100}\ \cos\frac{\pi k}{100}]^{\mathrm{T}}$.

This example shows EF-RLS, ER-RLS, and CR-RLS applied to estimate fixed and time-varying parameters in a linear measurement process corrupted by Gaussian noise. Figure 1 gives, for all $0 \leq k \leq 1500$, the eigenvalues of P_k and $\|\theta_k - \theta_{\mathrm{true},k}\|_2$ to show estimation performance.

For $0 \leq k \leq 500$ and $1000 \leq k \leq 1500$, ϕ_k is persistently exciting and similar magnitude to noise v_k. During this period, EF-RLS, ER-RLS, and CR-RLS perform nearly identically. However, for $500 < k < 1000$, excitation is poor while noise is large. As a result, EF-RLS experiences covariance windup, the eigenvalues of P_k reaching above 10^2, causing sensitivity to noise and poor estimation. ER-RLS and CR-RLS prevent covariance windup by limiting the eigenvalues of P_k to 1 (Corollary 1) and $1/\lambda^3$ (Corollary 3), respectively, which limits estimation error. Notice that,

TABLE I: Summary of Computational Complexities per Step and Information Matrix Properties

	Complexity ($n \gg p$)	Complexity ($p \geq n$)	R_k Lower Bound	R_k Upper Bound	R_k Resetting
EF-RLS	$\mathcal{O}(pn^2)$	$\mathcal{O}(pn^2)$	Prop. 2 *(requires PE)*	Prop. 3	Prop. 4 *(Resetting to 0)*
ER-RLS	$\mathcal{O}(n^3)$	$\mathcal{O}(pn^2)$	Prop. 6, Cor. 1	Prop. 7, Cor. 2	Prop. 8 *(Resetting to R_∞)*
CR-RLS	$\mathcal{O}(pn^2)$	*(Use ER-RLS Instead)*	Theo. 1, Cor. 3	Theo. 2, Cor. 4	Prop. 10, Theo. 3, Cor. 5 *(Cyclic Resetting)*

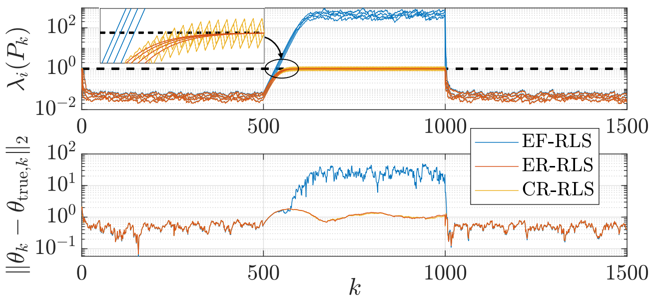

Fig. 1: Eigenvalues of P_k (top) with zoomed view and $\|\theta_k - \theta_{\mathrm{true},k}\|_2$ (bottom) for $0 \leq k \leq 1500$. No persistent excitation when $500 < k < 1000$.

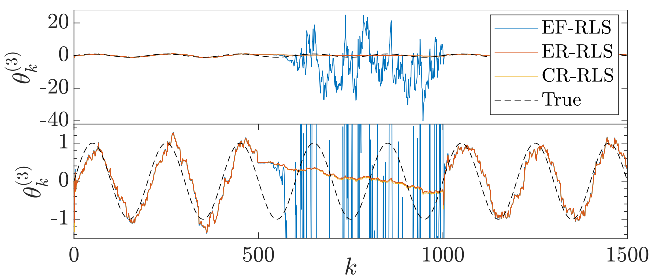

Fig. 2: The third element of θ_k and $\theta_{\mathrm{true},k}$ for $0 \leq k \leq 1500$ (bottom shows zoomed in y-axis). No persistent excitation when $500 < k < 1000$.

under little excitation, P_k nearly converges to P_∞ in ER-RLS (Proposition 8), while P_k oscillates around P_∞ in CR-RLS (Theorem 3).

For a closer look at tracking of time-varying parameters, Figure 2 shows $\theta_k^{(3)}$, the third element of θ_k. To track a quickly changing $\theta_{\mathrm{true},k}^{(3)} = \sin \frac{\pi k}{100}$, an aggressive forgetting factor $\lambda = 0.9$ is needed. However, without persistent excitation during $500 < k < 1000$, EF-RLS soon becomes sensitive to noise and the parameter estimate becomes erratic. ER-RLS and CR-RLS both limit the rate of parameter adaptation when persistent excitation is lost, resulting in robustness to measurement noise.

VI. Conclusions

This paper presents two extensions of RLS with a forgetting factor, or EF-RLS. The first, ER-RLS, which generalizes [13], is inspired by the interpretation that *forgetting* in EF-RLS is equivalent to *resetting* the information matrix to 0. ER-RLS extends EF-RLS by allowing resetting to a user-selected positive-definite matrix R_∞ but is more comp. complex than EF-RLS. The second, CR-RLS, maintains the same comp. complexity as EF-RLS when $n \gg p$, while providing similar resetting properties as ER-RLS. Computational complexities, guaranteed information matrix bounds, and resetting properties for the three algorithms are summarized in Table I. A numerical example shows how both ER-RLS and CR-RLS prevent covariance windup experienced by EF-RLS for both fixed and time-varying parameter estimation.

VII. Acknowledgements

This work is supported by the NSF Graduate Research Fellowship under Grant No. DGE 1841052.

References

[1] M. Ismail and J. Principe, "Equivalence between RLS algorithms and the ridge regression technique," in *Conf. Rec. of the Asilomar Conf. on Signals, Systems and Computers*. IEEE, 1996, pp. 1083–1087.

[2] S. A. U. Islam and D. Bernstein, "Recursive least squares for real-time implementation," *IEEE Ctrl. Sys. Mag.*, vol. 39, no. 3, pp. 82–85, 2019.

[3] B. Lai, S. A. U. Islam, and D. S. Bernstein, "Regularization-induced bias and consistency in recursive least squares," in *Proc. American Control Conference*. IEEE, 2021, pp. 3987–3992.

[4] R. Ortega, V. Nikiforov, and D. Gerasimov, "On modified parameter estimators for identification and adaptive control. a unified framework and some new schemes," *Ann. Rev. in Ctrl.*, vol. 50, pp. 278–293, 2020.

[5] K. J. Åström, U. Borisson *et al.*, "Theory and applications of self-tuning regulators," *Automatica*, vol. 13, no. 5, pp. 457–476, 1977.

[6] S. Sastry, M. Bodson, and J. F. Bartram, "Adaptive control: Stability, convergence, and robustness," *The Journal of the Acoustical Society of America*, vol. 88, no. 1, pp. 588–589, 1990.

[7] K. J. Åström and B. Wittenmark, *Adaptive control.* Courier Co., 2008.

[8] R. M. Johnstone, C. R. Johnson Jr, R. R. Bitmead, and B. D. Anderson, "Exponential convergence of recursive least squares with exponential forgetting factor," *Sys. & Ctrl. Letters*, vol. 2, no. 2, pp. 77–82, 1982.

[9] O. Malik, G. Hope, and S. Cheng, "Some issues on the practical use of recursive least squares identification in self-tuning control," *International Journal of Control*, vol. 53, no. 5, pp. 1021–1033, 1991.

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

[11] J. Parkum, N. K. Poulsen, and J. Holst, "Recursive forgetting algorithms," *Intl. Journal of Control*, vol. 55, no. 1, pp. 109–128, 1992.

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

[13] H.-S. Shin and H.-I. Lee, "A new exponential forgetting algorithm for recursive least-squares parameter estimation," *arXiv:2004.03910*, 2020.

[14] V. Vaerenbergh *et al.*, "Kernel recursive least-squares tracker for time-varying regression," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 8, pp. 1313–1326, 2012.

[15] A. Goel, A. L. Bruce, and D. S. Bernstein, "Recursive least squares with variable-direction forgetting: Compensating for the loss of persistency," *IEEE Ctrl. Syst. Mag.*, vol. 40, no. 4, pp. 80–102, 2020.

Appendix

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

Lemma A.2. *Let $A, B \in \mathbb{R}^{n \times n}$ be symmetric. Then, for all $i = 1, \ldots, n$, $\lambda_i(A) + \boldsymbol{\lambda}_{\mathbf{min}}(B) \leq \lambda_i(A+B) \leq \lambda_i(A) + \boldsymbol{\lambda}_{\mathbf{max}}(B)$.*