Accelerating with Low Rank Updates: RLS Framework for Segmentation of the Forgetting Profile

Alexander Stotsky
Independent Researcher
Gothenburg, Sweden

ABSTRACT

This paper introduces a novel regularization method that leverages segmentation of the forgetting profile for more robust modeling of data aging in sliding window least squares estimation. Each segment is designed to enforce specific desirable properties of the estimator such as rapidity, desired condition number of the information matrix, accuracy, numerical stability, etc. The forgetting profile is structured in three segments, where the first segment enables rapid estimation via fast exponential forgetting of recent data. The second segment features a decline in the profile and marks the transition to the third segment, which is characterized by slow exponential forgetting aimed at reducing the condition number of the information matrix using earlier measurements within the moving window. Condition number reduction mitigates error propagation, thereby enhancing accuracy and stability. This approach facilitates the incorporation of a priori information regarding signal characteristics (i.e., the expected behavior of the signal) into the estimator.
The main contribution of this paper is the framework for development of a new family of recursive, computationally efficient algorithms with low rank updates, based on a novel matrix inversion lemma for moving windows and tailored to this regularization approach.
New algorithms significantly improve the approximation accuracy of low resolution daily temperature measurements obtained at the Stockholm Old Astronomical Observatory, thereby enhancing the reliability of temperature predictions.

Keywords

segmentation of the forgetting profile, RLSLR: recursive least squares algorithm with low rank updates, matrix inversion lemma for moving window, integration of the segmented profile into the RLS framework, finite & infinite windows, temperature predictions based on low resolution daily measurements at the Stockholm Old Astronomical Observatory

1. FINITE & INFINITE WINDOWS: COMPARISONS & CHALLENGES

Recursive Least Squares (RLS) algorithms with exponential forgetting, [1],[2] are often preferred in real time applications due to their recursive structure and quadratic computational complexity, in contrast to sliding window least squares estimation, which requires repeated matrix inversions and has cubic complexity. Nevertheless, estimation over the finite window with exponential forgetting offers key advantages over classical RLS algorithms with infinite memory, as outlined below:
(1) *Faster Adaptation to Non-Stationary Environments*. Real world signals are often non-stationary due to changing parameters and noise. Classical RLS algorithms with exponential forgetting retain infinite memory, where old data is progressively downweighted but never completely discarded, which can lead to slow parameter convergence. In contrast, sliding window methods with exponential forgetting apply hard cutoff, enabling faster convergence and improved tracking of rapid signal changes.
(2) *Suppression of Long-Term Influence from Outliers and Past Noise*. In infinite memory RLS estimation, past outliers retain influence due to exponential weighting. Sliding window methods remove them after fixed time, improving robustness and accuracy.
(3) *Improved Numerical Stability*. Infinite memory RLS algorithms can accumulate errors and cause numerical instability over time. Sliding window methods limit memory, reducing error buildup and improving stability.
(4) *Greater Flexibility in Memory Control*. Sliding window methods allow separate tuning of window length and forgetting factor, offering finer control over estimator responsiveness than classical RLS algorithms, which depend only on forgetting factor.
(5) *Reduction of Bias from Initial Conditions*. Classical RLS algorithms can retain bias from poor initial conditions. Sliding window methods avoid this by discarding old data, ensuring estimates reflect recent information.
(6) *Facilitated Theoretical Analysis in Non-Ideal Conditions*. Sliding window methods simplify analysis by using fixed data horizon, unlike classical RLS algorithms whose infinite memory complicates evaluation in non-stationary settings.

The advantages of the sliding window approach, coupled with recent advancements in computationally efficient recursive algorithms exhibiting quadratic complexity, [3] - [5] render it as an increasingly attractive alternative to classical RLS methods.
Nevertheless, the pursuit of improved estimation performance in sliding window methods serves as a primary motivation for continued development, articulated below through a set of identified challenges:
(1) *Rapidity Costs*. Rapid parameter estimation comes at the cost of a high condition number of the information matrix, leading to large, imbalanced variances and potential numerical instability. Short windows and fast forgetting are required to track rapid changes in

the oscillating signals, but often fail to capture full cycles of low frequency components. As a result, low frequency harmonics are approximated by polynomials and become indistinguishable from trends. This blending of harmonic and trend components enables fast tracking but sacrifices resolution of low frequency harmonics. Regressors in short windows tend to be collinear, making the information matrix ill-conditioned or rank-deficient. Low frequency components, lacking full cycles have especially high variances of the parameter errors.
New recursive algorithms are required that simultaneously prioritize fast estimation, enhanced numerical stability, lower condition numbers, and uniform balance of the variances.
(2) *Reduction of Computational and Memory Costs for New Algorithms*. New algorithms featuring the outlined properties and quadratic complexity are needed to enable efficient parameter estimation and inversion of the information matrix, reducing computational and memory costs.

Ill-conditioning of the information matrix mentioned above can be addressed using various regularization strategies, such as adding small diagonal elements, [6], approximating low frequency components by polynomial models or selecting alternative basis functions with improved numerical stability and some others, see for example [7] and references therein. All these regularization methods entail trade-offs, potentially degrading estimation performance.

2. SEGMENTATION OF THE FORGETTING PROFILE

The most physically grounded and effective regularization method involves design/segmentation of the forgetting profile in the sliding window, where each segment assigns desired property to the estimator, [8]. Rapidity, the desired condition number of the information matrix, accuracy, and numerical stability can be mentioned among the desired properties. The proposed segmentation framework facilitates optimization of the trade-offs associated with the properties of the estimator and supports the incorporation of prior knowledge about the anticipated properties of the signal into the estimator.
Figure 1 illustrates a new method for segmenting the forgetting profile, which is introduced in this paper. The rapidity of estimation is guaranteed by fast exponential forgetting, see the red segment (red solid line) in Figure 1. Such rapid forgetting across the entire window (see red dashed line) implies ill-conditioning of the information matrix. Reduction of the condition number is associated with the blue line, which represents slow exponential forgetting. Rapid estimation can not be achieved with blue forgetting profile, but this profile mitigates ill-conditioning and improves robustness and accuracy. The segmented profile introduces the drop in the red line (which can be adjusted), followed by transition to the blue segment/tail, where the profile diminishes more slowly due to a higher forgetting factor. The condition number of the information matrix is reduced using more distant data (due to the blue segment) which implies lower variance of parameter estimates and improved numerical stability, while the rapidity is achieved by fast forgetting of recent measurements (due to the red segment).

3. RECURSIVE ESTIMATION FRAMEWORK FOR THE SEGMENTED PROFILE

3.1 Matrix Inversion Lemma for Moving Window

Consider the following generalization of the matrix inversion lemma for moving window described in [9, p. 280].
Lemma (batch update in moving window). Updates of the invertible $n \times n$ matrix A, associated with d additions and r removals of the outer products of the vectors $x_j \in \mathbb{R}^n$ and $y_j \in \mathbb{R}^n$ can be expressed in the following equivalent forms:

$$A = B + \underbrace{\sum_{j=1}^{d} x_j\, x_j^T}_{\text{updating}} - \underbrace{\sum_{j=1}^{r} y_j\, y_j^T}_{\text{downdating}} \tag{1}$$

$$A = B + Q\, D\, Q^T \tag{2}$$

where $Q = [x_1\ x_2\ ...\ x_d\ y_1\ y_2\ \cdots\ y_r]$ is the augmented regressor matrix which contains regressor column vectors, $D = \text{diag}[\underbrace{1,\ 1\ , ...\ , 1}_{\text{d additions}}, \underbrace{-1, -1\ , ...,\ -1}_{\text{r removals}}]$ is the signature matrix, which determines the sign structure of the low rank correction, $d + r \ll n$. Subsequently, the inverse of A can be efficiently obtained through the batch update by applying the Woodbury identity, [10, p. 350] to the compact form (2) as follows:

$$A^{-1} = B^{-1} - B^{-1}\, Q\, U^{-1}\, Q^T\, B^{-1} \tag{3}$$

where the update matrix $U = D + Q^T\, B^{-1}\, Q$ is invertible and the inverse B^{-1} is known.

Corollary. Rank two updates of the matrix A is the special case of (1),(2) with $d = r = 1$, which can be presented in the form (2) with the following regressor matrix $Q = [x_1\ y_1]$ and the signature matrix $D = \text{diag}[1, -1]$, [3] - [5].

The inverse A^{-1} can also be computed iteratively using the Sherman Morrison formula, applying it once per rank one update and using each intermediate inverse in the subsequent step, [9], [11]. However, this approach has several limitations: it is prone to significant round off error accumulation, may destroy symmetry and positive definiteness, and can encounter singular matrices during intermediate steps, requiring additional assumptions about invertibility. Furthermore, the method can be computationally inefficient due to poor cache performance and repeated memory access, and it offers limited opportunities for parallelization due to the inherently sequential nature of the updates.
The inversion method presented in the lemma applies the correction in a single pass, thereby reducing the number of matrix multiplications and mitigating the propagation of numerical errors. It preserves matrix symmetry, improves the conditioning of the resulting inverse, and is inherently more parallelizable. In addition, the method exhibits enhanced memory efficiency, making it particularly effective for deployment in high performance computing architectures.

3.2 Recursive Estimation Algorithms with Low Rank Updates

Assume that the measured oscillating signal y_k and its corresponding model $\hat{y}_k$ are represented as follows:

$$y_k = \varphi_k^T \theta_* + \xi_k \tag{4}$$

$$\hat{y}_k = \varphi_k^T \theta_k \tag{5}$$

$$\varphi_k^T = [1\ cos(q_0 k)\ sin(q_0 k)\ \cdots\ cos(q_h k)\ sin(q_h k)]$$

where θ_* is the vector of unknown parameters, θ_k is the vector of adjustable parameters calculated via minimization of the loss function (6), φ_k is the harmonic regressor, $q_0, ...q_h$ are the frequencies, and ξ_k is white Gaussian zero mean noise uncorrelated with φ_k, $k = 1, 2, ...$

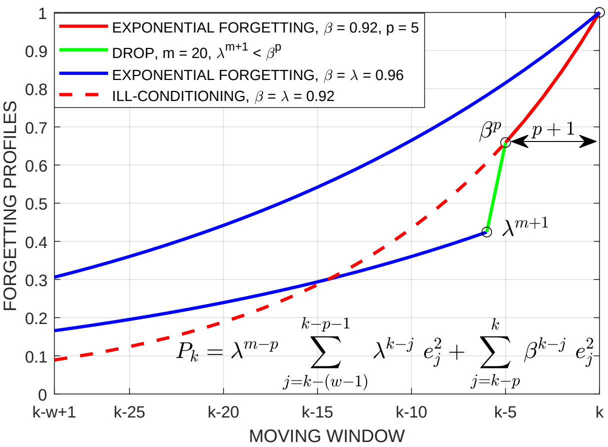

Fig. 1. The Figure shows the forgetting profile in the moving window of the size w segmented by red, green and blue lines (RGB lines). The rapidity of estimation is guaranteed by rapid exponential forgetting with small forgetting factor $\beta = 0.92$ associated with the red line (red segment). Such rapid forgetting over the whole window implies ill-conditioning of the information matrix, see red dashed line. Reduction of the condition number is associated with the blue line and the forgetting factor $\lambda = 0.96$. The resulting profile is segmented by the red segment, drop (with the magnitude determined by a positive integer m) which is plotted with the green line and blue segment with larger forgetting factor. Past data (associated with blue segment) forms the basis for reduction of the condition number of the information matrix and the variances of parameter estimates, while rapidity is achieved by fast forgetting of recent measurements. Segmented profile is associated with the cost function P_k which minimises estimation error $e_j = y_j - \varphi_j^T \theta_k$.

The segmented profile presented in Figure 1 is associated with the following loss function:

$$P_k = \sum_{j=k-(w-1)}^{k-p-1} \lambda^{k-p-j+m} \, (y_j - \varphi_j^T \theta_k)^2 + \sum_{j=k-p}^{k} \beta^{k-j} \, (y_j - \varphi_j^T \theta_k)^2 \tag{6}$$

This function consists of two terms, where the first term $\sum_{j=k-p}^{k} \beta^{k-j} \; (y_j - \varphi_j^T \theta_k)^2$ downweights rapidly recent $p+1$ ($p = 1, 2, \cdots \ll w$) measured points with the relatively small forgetting factor $0 < \beta < 1$, see Figure 1. Integer $m > 0$ determines the magnitude of the drop in the profile so that $\lambda^{m+1} < \beta^p$ and the second term $\sum_{j=k-(w-1)}^{k-p-1} \lambda^{k-p-j+m} \; (y_j - \varphi_j^T \theta_k)^2$ downweights older data with larger forgetting factor, $0 < \lambda < 1$.
Minimization of the loss function (6) results in the following rank $p+3$ updates of the information matrix:

$$A_k = \lambda \, A_{k-1} + Q_k \, D \, Q_k^T \tag{7}$$

$$Q_k = [\varphi_k \; \sqrt{|\beta - \lambda|}\varphi_{k-1} \cdots \sqrt{|\lambda^m - \beta^p|\lambda}\varphi_{k-p-1} \; \sqrt{\lambda^{m+w-p}}\varphi_{k-w}]$$

$$D = \mathrm{diag}[\underbrace{1, \; \mathrm{sign}(\beta - \lambda), \; \cdots , \; \mathrm{sign}(\lambda^m - \beta^p)}_{p+1}, \; -1]$$

where the augmented regressor matrix Q_k contains scaled column regressor vectors, $k \geq w+1$. The recursive framework (7)-(9) supports various types of segmentation parameterized by β, λ, m and p, see Figure 1. Each profile yields a distinct sign structure of the signature matrix D, which governs the addition and removal of updates.
The RLS algorithms with low rank updates which are derived by application of the matrix inversion lemma for moving window presented above can be written in the following form, [3], [5] :

$$\Gamma_k = \frac{1}{\lambda} \, [\, \Gamma_{k-1} - \Gamma_{k-1} \, Q_k \, S^{-1} \, Q_k^T \, \Gamma_{k-1} \,], \;\; \Gamma_w = A_w^{-1} \tag{8}$$

$$\theta_k = \theta_{k-1} - \Gamma_{k-1} \, Q_k \, S^{-1} \, [Q_k^T \, \theta_{k-1} - \tilde{y}_k] \tag{9}$$

$$S = \lambda \, D + Q_k^T \, \Gamma_{k-1} \, Q_k$$

$$\tilde{y}_k^T = [y_k \; \sqrt{|\beta - \lambda|}y_{k-1} \; \cdots \; \sqrt{|\lambda^m - \beta^p|\lambda}y_{k-p-1} \; \sqrt{\lambda^{m+w-p}}y_{k-w}]$$

where $\tilde{y}_k$ is the augmented output.
The unbiasedness of the algorithm (8), (9), $\mathbb{E}[\theta_k] = \theta_*$ can be formally derived using the framework presented in [5], assuming a full-rank information matrix. The variance of the parameter estimation error is inversely proportional to the eigenvalues of the information matrix. When the information matrix is ill-conditioned, some directions in parameter space which correspond to small eigenvalues lead to large estimation variances along these directions. This typically happens when the regressor signals are not sufficiently excited over the window. In contrast, well-conditioned information matrices result in more uniform and lower variance across all parameters.
Notice that applying an infinite window over a large number of iterations can introduce bias, increase the variance of parameter mismatch, and cause imbalance. These effects often arise due to ill-conditioning, numerical error accumulation, sensitivity to measurement noise, outliers, and unmodeled disturbances.
In contrast, segmented forgetting in the finite window alleviates ill-conditioning and reduces bias, variance, and imbalance in parameter estimation, thereby improving the robustness and performance of the algorithms, see Section 1.

4. TEMPERATURE FORECASTS

Temperature forecasts covering a period longer than one week are essential for providing early insights into weather trends that facilitate effective planning and preparedness. Such forecasts contribute to improved decision making across multiple sectors by offering a comprehensive outlook on forthcoming climate conditions.
Using periodicity is essential for accurately predicting temperature because temperature naturally follows repeating patterns such as daily, weekly, and seasonal cycles. These cycles reflect natural phenomena like the daily rise and fall of temperature, weekly patterns influenced by human activity, and seasonal changes. These patterns can be effectively identified using system identification techniques, and leveraging periodicity in this manner can greatly enhance forecasting accuracy.

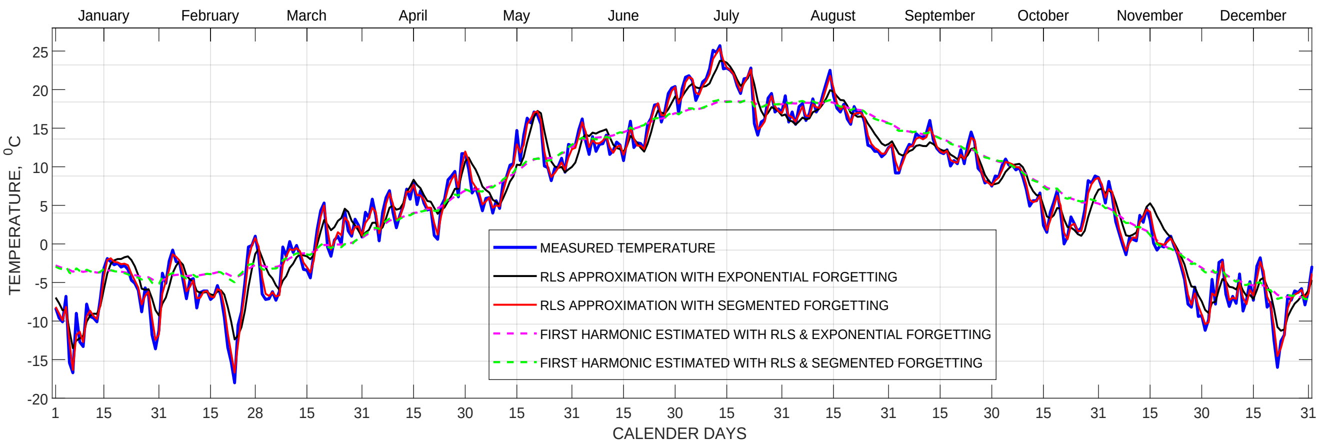

Fig. 2. Comparison of the approximation performance of segmented and exponential forgetting profiles in moving window of the size $w = 400$ is shown in this Figure. Daily temperature measurements are plotted with the blue line. The output of the RLS algorithm with rank two updates and $\lambda = 0.99$ is plotted with the black line. The output of RLS algorithm with segmented profile and rank four updates, designed for $p = 1$, $\beta = 0.89$, $\lambda = 0.99$, $m = 250$ (see Figure 1) is plotted with the red line. Histograms of approximation errors are presented in Figure 3. Estimates of the first harmonics are plotted with magenta and green lines for exponential and segmented profiles respectively.

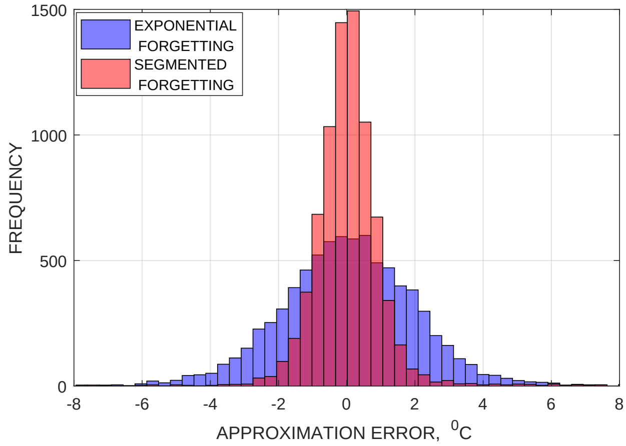

Fig. 3. The blue histogram shows the approximation error for the exponential forgetting profile, and the red histogram shows it for the segmented forgetting profile. Approximation performance is significantly improved via segmentation of the profile.

Assume that the temperature time series is represented by equation (4), incorporating the fundamental frequency associated with the annual cycle, as well as sixteen higher order harmonics that capture additional periodic components of shorter durations. The model of the signal is presented in the form (5) with adjustable parameters (8),(9). To validate the model, low resolution daily mean temperature measurements from the Stockholm Old Astronomical Observatory, [12] are used.

Comparison of the approximation performance of segmented and exponential forgetting profiles is shown in Figure 2 and Figure 3. The measured temperature, shown by the blue line in Figure 2, is approximated using two RLS algorithms: one with rank two updates (black line), and another with segmented forgetting and rank four updates (red line). The histograms in Figure 3 illustrate that the approximation performance is significantly improved by segmenting the forgetting profile. Moreover, the first harmonic, which corresponds to the annual periodicity, is estimated more accurately, directly impacting the long term temperature forecast presented in Figure 4.

The Figure 4 presents the 30-day-ahead temperature forecast based on the first harmonic component, accompanied by a three sigma confidence interval, [13]. The variance is also estimated within the moving window. It has been demonstrated that the seasonal trend can be accurately predicted using estimates of the first harmonic obtained from low resolution temperature measurements. Nearly all observed temperature measurements fall within the established confidence intervals around the predicted values of the first harmonic, confirming the reliability of the predictions.

5. CONCLUSIONS & OUTLOOK

The regularization method, aligned with segmentation of the forgetting profile, showed strong potential in optimizing the trade-offs between rapidity, desired condition number of the information matrix, accuracy, and numerical stability. The main contribution of this work is the integration of the segmented profile into a low rank update recursive least squares framework. The development utilizes the matrix inversion lemma tailored for moving window computations.

Design flexibility in the forgetting profile enabled more accurate approximation of the frequency content in low resolution temperature measurements, thereby improving the reliability of temperature predictions.

Finally, the recursive framework with low rank updates (developed in this paper) facilitates the development and assessment of new, more advanced segmentation strategies which are anticipated to further improve estimation performance. Moreover, performance can be further enhanced through the use of Newton-Schulz and Richardson corrections,[3], [14].

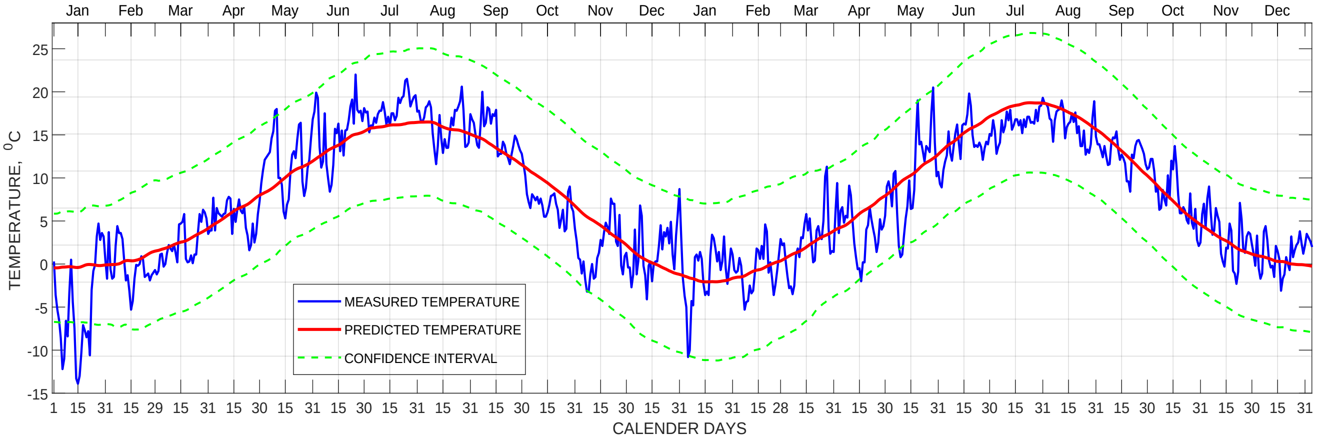

Fig. 4. The Figure shows the 30-day-ahead temperature forecast based on prediction of the first harmonic and three sigma confidence interval with estimation of the variance in moving window, [13]. It is shown that the seasonal trend can be very well predicted using estimate of the first harmonic of low resolution daily temperature measurements. The prediction estimates the mean, maximum, and minimum temperature values projected 30 days ahead. The accuracy of the prediction is assessed by checking if the actual temperature measurements lie within the confidence intervals established around the predicted first harmonic values. Most observed temperature measurements fall within the confidence intervals of the first harmonic predictions, confirming their reliability.

Disclosure Statement

This research was not supported by any organization.

6. REFERENCES

[1] Fomin V., Fradkov A. & Yakubovich V. (1981). *Adaptive control of the dynamic plants*. Nauka, Moscow, in Russian.

[2] Ljung, L. & Söderström T. (1983). *Theory and practice of recursive identification*. The MIT press series in signal processing, optimization, and control. `https://mitpress.mit.edu/9780262620581/theory-and-practice-of-recursive-identification/`

[3] Stotsky A.(2023). *Recursive estimation in the moving window: efficient detection of the distortions in the grids with desired accuracy*. Journal of Advances in Applied and Computational Mathematics, 9, 181-191. `https://doi.org/10.15377/2409-5761.2022.09.14`

[4] Stotsky A. (2024). *Kaczmarz projection algorithms with rank two gain update*. Journal of Signal Processing Systems, 96(4-5), 327-332. `https://doi.org/10.1007/s11265-024-01915-w`

[5] Stotsky A. (2025). *Recursive least squares estimation with rank two updates*. Automatika, 66 (4), 619-624 . `https://doi.org/10.1080/00051144.2025.2517431`

[6] Pillonetto G., Dinuzzo F., Chen T., De Nicolao G., & Ljung, L. (2014). *Kernel methods in system identification, machine learning and function estimation: A survey*. Automatica, 50(3),657-682. `https://doi.org/10.1016/j.automatica.2014.01.001`.

[7] Lai B. & Bernstein D. (2024). *Generalized forgetting recursive least squares: stability and robustness guarantees. IEEE Transactions on Automatic Control*, 69(11), 7646 - 7661. `https://doi.org/10.1109/TAC.2024.3394351`

[8] Maouche, K. and Slock, D. (1995). *Performance analysis and FTF version of the generalized sliding window recursive least-squares (GSWRLS) algorithm*. In: Proceedings of the Asilomar Conference on Signals, Systems and Computers, Vol. 1, Pacific Grove, USA, November 1995, pp.685–689. `https://ieeexplore.ieee.org/document/540637`

[9] Papaodysseus C., Koukoutsis E. , Halkias C. & Roussopoulos G. (1998). *A parallelizable recursive least squares algorithm for adaptive filtering with very good tracking properties*. International Journal of Computer Mathematics, 67(3-4), 275-292. `https://doi.org/10.1080/00207169808804665`

[10] Bernstein D.S. (2018). *Scalar, vector, and matrix mathematics: Theory, facts, and formulas*. Revised and expanded ed. Princeton: Princeton University Press. `https://doi.org/10.2307/j.ctvc7729t`

[11] Zhang Q. (2000). *Some implementation aspects of sliding window least squares algorithms*. In: Proceedings of the 12th IFAC Symposium on System Identification (SYSID 2000), Santa Barbara, California, USA, 21–23 June 2000, pp.763–768. `https://doi.org/10.1016/S1474-6670(17)39844-0`

[12] Moberg A. (2019). *Stockholm Historical Weather Observations — Daily mean air temperatures since 1756*. Dataset version 1. Bolin Centre Database, `https://bolin.su.se/data/stockholm-historical-temps-daily`.

[13] Stotsky A. (2025). *Detection and control of credit card fraud attacks in sliding window with exponential forgetting*. International Journal of Computer Applications (0975 - 8887), 186(74),9–15. `https://doi.org/10.5120/ijca2025924619`

[14] Stotsky A. (2022). *Recursive versus nonrecursive Richardson algorithms: systematic overview, unified frameworks and application to electric grid power quality monitoring*, Automatika, 63(2),328-337. `https://doi.org/10.1080/00051144.2022.2039989`