ECG Decomposition using Kalman Filtering

Mohammadsina Hassannia
Topics in Biomedical Informatics: Kalman Filtering
Fall 2025

***Abstract*—The 12-lead electrocardiogram (ECG) is one of the most important clinical tools used for cardiac diagnosis, yet its full acquisition is often impractical in wearable and resource-constrained monitoring systems. This work proposes an adaptive Kalman filter–based FIR framework for ECG lead reconstruction, capable of synthesizing a target lead from one or multiple available source leads. By modeling inter-lead relationships as time-varying linear mappings, the proposed method explicitly captures temporal dynamics and measurement uncertainty, offering a physiologically plausible and computationally efficient alternative to static linear transforms and purely data-driven approaches.**

The framework is evaluated on the large-scale PTB-XL dataset using both single-lead to single-lead and multi-lead to single-lead reconstruction scenarios. Reconstruction performance is quantified using signal-to-noise ratio (SNR), enabling systematic assessment of shared information and redundancy across ECG leads. Experimental results demonstrate that the performance closely follows known anatomical and electrophysiological principles: adjacent precordial leads exhibit strong mutual predictability due to spatial proximity, while limb leads retain greater unique information.

Beyond lead synthesis, the proposed approach provides a principled signal decomposition into shared and lead-specific components, enabling quantitative characterization of inter-lead dependencies. These findings have direct implications for optimized lead selection, ECG compression, and the design of wearable or portable cardiac monitoring systems, while preserving interpretability and adaptability in real-world clinical settings.

I. Introduction

The 12-lead electrocardiogram (ECG) remains the clinical gold standard for diagnosing cardiac conditions, as it captures the heart's electrical activity from multiple anatomical perspectives, facilitating the detection of abnormalities such as ischemia, arrhythmias, and conduction disturbances [1]–[4]. ECG lead reconstruction involves estimating missing or unrecorded leads from a reduced set of available signals. This task has gained significant attention with the rapid increase of need for wearable and portable cardiac monitors, where recording a complete 12-lead ECG is often impractical due to constraints in device size, power consumption, and patient comfort. Also, the complex interplay between shared and lead-specific components in multi-lead ECGs remains underexplored. The similarities and differences arise partly from the common cardiac rhythm, while the leads view the heart from different perspectives, resulting in both redundant and distinct patterns across the leads

The foundational idea of inter-lead relationships traces back to Dower et al., who showed that surface ECG leads can be modeled as linear projections of the cardiac vector field via the Dower transform, a fixed matrix relating the 12-lead ECG to the Frank vectorcardiogram (VCG) [5]. Early efforts built on this by employing inverse Dower transformations and linear regression for lead synthesis [6], [7]. More recent work has shifted toward nonlinear and data-driven methods, including artificial neural networks, which have demonstrated improved reconstruction accuracy [8], [9].

Kalman filtering provides a robust statistical framework for modeling temporal dynamics and measurement uncertainty in time-series data [10]. By recursively updating state estimates as new observations arrive, the Kalman filter adapts to gradual physiological changes (e.g., heart rate variability) or environmental factors while suppressing noise. using this ability, kalman filter has been used in ECG denoising and tracking applications [11]. For lead reconstruction, this approach enables direct projection from one lead to another or indirect reconstruction via a source model (e.g., the cardiac vector), offering an adaptive, computationally efficient, and physiologically plausible alternative to purely data-driven methods.

In this study, we primarily focus on reconstructing a single target ECG lead from one available source lead using a Kalman filter–based FIR formulation, which serves as a controlled and interpretable baseline for analyzing inter-lead relationships. This setting allows us to explicitly quantify the information shared between individual leads while leveraging the Kalman filter's ability to model temporal dependencies and adapt to the nonstationary nature of cardiac signals. Building on this formulation, the approach naturally extends to scenarios where multiple source leads are available, by directly incorporating them into a unified multi-lead Kalman model. This approach aligns with the inherent redundancy of the 12-lead ECG and provides a flexible framework for improving reconstruction fidelity in reduced-lead and resource-constrained acquisition settings.

II. Dataset Description

The PTB-XL dataset is a large, publicly available electrocardiogram (ECG) database developed by the Physikalisch-Technische Bundesanstalt (PTB) in Germany and hosted on PhysioNet. It stands out as one of the most comprehensive open-access ECG datasets specifically curated for machine learning and signal processing research [12]. The dataset comprises 21,837 clinical 12-lead ECG recordings collected from 18,885 patients between 1989 and 1996. Each recording lasts 10 seconds and is provided in two sampling rates: the original 500 Hz version and a downsampled 100 Hz version, offering flexibility depending on the computational requirements of the analysis. The ECGs were carefully annotated by up to

two cardiologists using the SCP-ECG standard, resulting in 71 diagnostic statements that are grouped into five major categories: normal ECG (NORM), myocardial infarction (MI), ST/T changes (STTC), conduction disturbances (CD), and hypertrophy (HYP). All signals are stored in the standard WFDB format and are freely accessible.

In this project, we will use the high-resolution 500 Hz version of the data. The combination of excellent signal quality, standardized and reliable annotations, and a broad spectrum of cardiac pathologies makes PTB-XL an ideal benchmark for developing and evaluating ECG lead reconstruction algorithms.

III. Proposed Methodology

A. Single-Lead to Single-Lead Kalman FIR Model

Let $x_t \in \mathbb{R}$ denote the source ECG lead at time t, and $y_t \in \mathbb{R}$ denote the target ECG lead.

a) Regressor vector: For an FIR filter of order q, define the regressor vector

$$\boldsymbol{\phi}_t = \begin{bmatrix} x_t & x_{t-1} & \cdots & x_{t-q} \end{bmatrix}^\top \in \mathbb{R}^{q+1}.$$

b) State vector: The state vector consists of the time-varying FIR coefficients

$$\mathbf{h}_t = \begin{bmatrix} h_{0,t} & h_{1,t} & \cdots & h_{q,t} \end{bmatrix}^\top \in \mathbb{R}^{q+1}.$$

c) State evolution model: The coefficients follow a random-walk model

$$\mathbf{h}_{t+1} = \mathbf{h}_t + \mathbf{w}_t, \qquad \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}),$$

with

$$\mathbf{Q} = q_{\text{var}}\, \mathbf{I}_{q+1}.$$

d) Measurement model: The observed target lead is generated as

$$y_t = \boldsymbol{\phi}_t^\top \mathbf{h}_t + v_t, \qquad v_t \sim \mathcal{N}(0, R),$$

where

$$R = r_{\text{var}}.$$

e) Kalman filter recursion: Given initial estimates $\hat{\mathbf{h}}_{0|0}$ and $\mathbf{P}_{0|0}$, the Kalman filter proceeds as follows.

Prediction step

$$\hat{\mathbf{h}}_{t|t-1} = \hat{\mathbf{h}}_{t-1|t-1},$$
$$\mathbf{P}_{t|t-1} = \mathbf{P}_{t-1|t-1} + \mathbf{Q}.$$

Innovation

$$\hat{y}_{t|t-1} = \boldsymbol{\phi}_t^\top \hat{\mathbf{h}}_{t|t-1},$$
$$e_t = y_t - \hat{y}_{t|t-1},$$
$$S_t = \boldsymbol{\phi}_t^\top \mathbf{P}_{t|t-1} \boldsymbol{\phi}_t + R.$$

Kalman gain

$$\mathbf{K}_t = \frac{\mathbf{P}_{t|t-1}\boldsymbol{\phi}_t}{S_t}.$$

Update step (R is scalar)

$$\hat{\mathbf{h}}_{t|t} = \hat{\mathbf{h}}_{t|t-1} + \mathbf{K}_t e_t,$$
$$\mathbf{P}_{t|t} = (\mathbf{I}_{q+1} - \mathbf{K}_t \boldsymbol{\phi}_t^\top)\mathbf{P}_{t|t-1}(\mathbf{I}_{q+1} - \mathbf{K}_t \boldsymbol{\phi}_t^\top)^\top + R\,\mathbf{K}_t\mathbf{K}_t^\top.$$

f) Filtered output: The reconstructed target lead is

$$\hat{y}_t = \boldsymbol{\phi}_t^\top \hat{\mathbf{h}}_{t|t}.$$

B. Multi-Lead to Single-Lead Kalman FIR Model

We extend the single-lead Kalman FIR formulation to the case where a single target ECG lead is reconstructed from multiple simultaneously recorded source leads.

Let $\mathbf{x}_t \in \mathbb{R}^L$ denote the vector of L source ECG leads at time t,

$$\mathbf{x}_t = \begin{bmatrix} x_t^{(1)} & x_t^{(2)} & \cdots & x_t^{(L)} \end{bmatrix}^\top,$$

and let $y_t \in \mathbb{R}$ denote the target ECG lead.

a) Regressor vector: For an FIR filter of order q, the regressor vector is constructed by stacking delayed versions of all input leads:

$$\boldsymbol{\phi}_t = \begin{bmatrix} \mathbf{x}_t^\top & \mathbf{x}_{t-1}^\top & \cdots & \mathbf{x}_{t-q}^\top \end{bmatrix}^\top \in \mathbb{R}^{L(q+1)}.$$

b) State vector: The state vector contains the time-varying FIR coefficients associated with all leads and delays,

$$\mathbf{h}_t = \begin{bmatrix} \mathbf{h}_t^{(0)} & \mathbf{h}_t^{(1)} & \cdots & \mathbf{h}_t^{(q)} \end{bmatrix}^\top \in \mathbb{R}^{L(q+1)},$$

where $\mathbf{h}_t^{(k)} \in \mathbb{R}^L$ represents the contribution of the k-sample delayed input vector $\mathbf{x}_{t-k}$.

c) State evolution model: The FIR coefficients are modeled as a random walk,

$$\mathbf{h}_{t+1} = \mathbf{h}_t + \mathbf{w}_t, \qquad \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}),$$

with

$$\mathbf{Q} = q_{\text{var}}\, \mathbf{I}_{L(q+1)}.$$

d) Measurement model: The observed target lead is generated according to

$$y_t = \boldsymbol{\phi}_t^\top \mathbf{h}_t + v_t, \qquad v_t \sim \mathcal{N}(0, R),$$

where

$$R = r_{\text{var}}.$$

e) Kalman filter recursion: Given initial estimates $\hat{\mathbf{h}}_{0|0}$ and $\mathbf{P}_{0|0} \in \mathbb{R}^{L(q+1)\times L(q+1)}$, the Kalman filter proceeds as follows.

Prediction step

$$\hat{\mathbf{h}}_{t|t-1} = \hat{\mathbf{h}}_{t-1|t-1},$$
$$\mathbf{P}_{t|t-1} = \mathbf{P}_{t-1|t-1} + \mathbf{Q}.$$

Innovation

$$\hat{y}_{t|t-1} = \boldsymbol{\phi}_t^\top \hat{\mathbf{h}}_{t|t-1},$$
$$e_t = y_t - \hat{y}_{t|t-1},$$
$$S_t = \boldsymbol{\phi}_t^\top \mathbf{P}_{t|t-1} \boldsymbol{\phi}_t + R.$$

Kalman gain

$$\mathbf{K}_t = \frac{\mathbf{P}_{t|t-1}\boldsymbol{\phi}_t}{S_t}.$$

Update step

$$\hat{\mathbf{h}}_{t|t} = \hat{\mathbf{h}}_{t|t-1} + \mathbf{K}_t e_t,$$
$$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t \boldsymbol{\phi}_t^\top)\mathbf{P}_{t|t-1}(\mathbf{I} - \mathbf{K}_t \boldsymbol{\phi}_t^\top)^\top + R\,\mathbf{K}_t\mathbf{K}_t^\top.$$

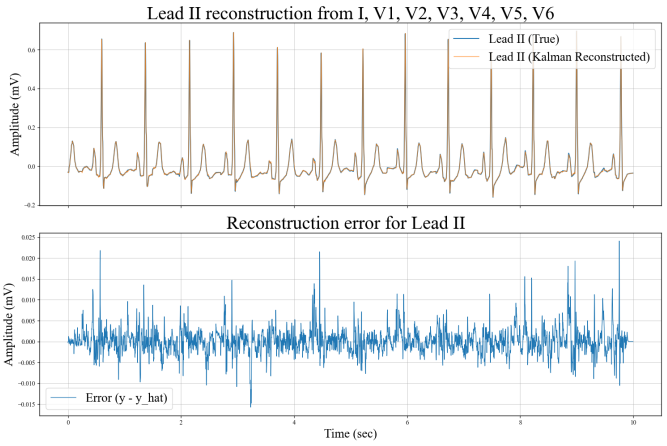

Fig. 1. Example of Lead II reconstruction (top) and reconstruction error (bottom) using a Kalman-based model (order=4, SNR≈30.43 dB).

f) Filtered output: The reconstructed target ECG lead is

$$\hat{y}_t = \boldsymbol{\phi}_t^\top \hat{\mathbf{h}}_{t|t}.$$

IV. Results

We performed an analysis using a single lead - single lead approach to assess the information shared between individual ECG leads. In this method, each target lead was reconstructed using only one source lead, and the resulting signal-to-noise ratio (SNR) was used as a quantitative measure of shared information. This pairwise evaluation provides insight into redundancy and complementarity across leads. The lead-by-lead mean SNR values obtained from this analysis are reported in Table I.

We also conducted a multi-lead to single-lead analysis to evaluate the reconstructability of individual leads. In this analysis, each lead is reconstructed using all remaining available leads.

Figure 1 shows an example of Lead II ECG reconstruction from the remaining leads using a Kalman-based model (order $= 4$, SNR ≈ 30.43 dB). In the top subplot, the reconstructed signal closely overlaps the true Lead II waveform over the full 10-second window, preserving the morphology of the P-QRS-T complexes, including the sharp R-peaks and the lower-amplitude components associated with atrial activity and repolarization. This close agreement indicates that both the timing and amplitude dynamics of the cardiac cycles are well captured by the reconstruction. The bottom subplot reports the reconstruction error, which remains centered near zero with small, noise-like fluctuations (on the order of a few tens of μV), with slightly larger deviations concentrated around high-slope regions such as the QRS complexes. Overall, the strong visual overlap and low-magnitude residuals suggest high reconstruction fidelity and demonstrate the effectiveness of the model for recovering Lead II from the available input leads.

Additionally, the SNR boxplots provide a quantitative assessment of the information shared across leads. This analysis enables the identification of leads that contain more unique information, since lower reconstruction performance suggests reduced redundancy with respect to other leads. The distributions of the reconstructed-lead SNRs are shown in Fig. 2.

a) Kalman gain analysis: Figure 3 shows the time-varying Kalman gain magnitude associated with each input ECG lead when reconstructing target lead II using an order-4 Kalman FIR model. The gain exhibits clear temporal modulation, with pronounced peaks during high-information intervals such as the QRS complexes, indicating that the filter adaptively increases reliance on informative input samples. Moreover, the gain magnitude varies across leads, reflecting their differing contributions to the reconstruction of the target lead. This behavior highlights the ability of the Kalman filter to automatically exploit inter-lead redundancy and time-varying correlations in multi-lead ECG signals.

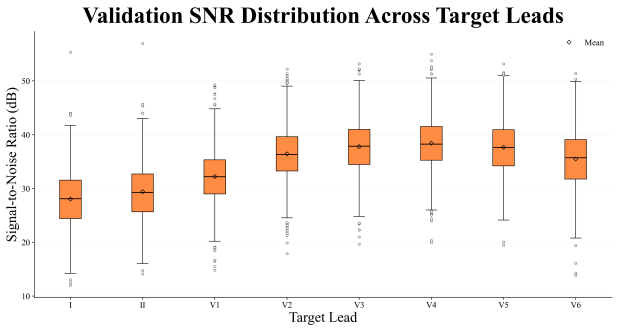

Fig. 2. Boxplot distributions of the signal-to-noise ratio (SNR) for the reconstructed leads.

TABLE I
PAIRWISE SINGLE-LEAD RECONSTRUCTION PERFORMANCE EXPRESSED AS MEAN SNR (TARGET ROWS, SOURCE COLUMNS).

T \ S	I	II	V1	V2	V3	V4	V5	V6
I	–	**10.75**	7.53	7.73	7.48	8.82	9.96	10.40
II	**11.23**	–	6.29	6.88	6.84	8.68	9.49	9.82
V1	9.47	9.97	–	**15.14**	12.60	11.81	11.71	11.71
V2	8.02	8.27	11.69	–	**18.04**	13.23	11.12	10.17
V3	8.79	8.41	10.03	**17.81**	–	16.94	12.84	11.06
V4	10.54	10.97	11.05	14.13	18.05	–	**18.45**	14.84
V5	12.57	13.64	11.88	13.31	14.82	**19.88**	–	19.76
V6	13.44	14.27	12.30	13.11	13.47	16.70	**20.22**	–

V. Discussion

From a signal processing perspective, the primary contribution of the proposed approach is its decomposition of the 12-lead ECG into shared (common) and lead-specific components, effectively establishing it as a robust signal decomposition framework. The results from Table I confirm that limb leads has more information in common, and in chest leads, leads near each other share more information, which is completely aligned with the physiology of the human heart: precordial leads capture localized electrical activity in the horizontal plane, with neighboring leads (e.g., V1 and V2, overlying the right ventricle and interventricular septum) influenced by

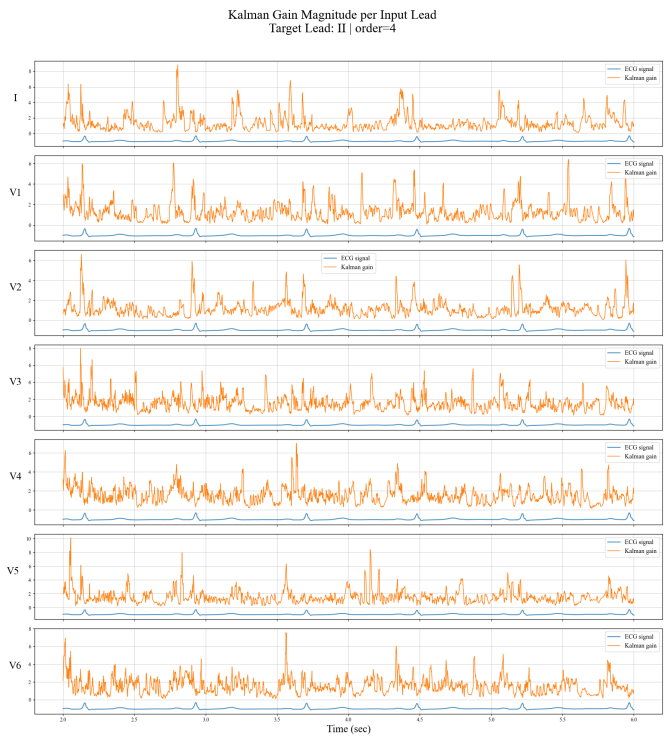

Fig. 3. Kalman gain magnitude per input lead for reconstruction of target lead II (order = 4).

similar near-field potentials due to proximity. This spatial dependency explains the superior reconstruction performance for adjacent precordial leads compared to distant ones.

Furthermore, based on Fig. 2 we can understand varying degrees of redundancy across leads. Certain precordial leads, such as V4 and V5, prove highly predictable from others, indicating greater redundancy, whereas limb leads like I and II contribute more unique information. These observations align with anatomical principles precordial leads, positioned closer to the heart, predominantly record near-field signals with inherent overlap, while limb leads reflect far-field projections with greater independence.

This decomposition offers an approach to quantify lead redundancy and inter-lead relationships, with direct applications to optimized lead selection, prioritization, and enhanced feature extraction in clinical diagnostic workflows. The findings carry several important implications. First, identifying leads with minimal unique information can inform the design of resource-constrained devices, such as wearable or portable ECG systems, enabling reliable monitoring without full 12-lead acquisition. Second, the innovation components may reveal differential responses to pathological conditions, yielding novel features for improved disease detection and interpretation. Third, explicit modeling of inter-lead dependencies facilitates more efficient ECG compression algorithms for storage and transmission.

VI. Conclusion

This study presented a Kalman filter–based FIR framework for ECG lead reconstruction, addressing both single-lead and multi-lead scenarios with a unified, adaptive formulation. By explicitly modeling the temporal dynamics of inter-lead relationships, the proposed approach offers a physiologically meaningful and computationally efficient alternative to static linear transforms and purely data-driven methods. The framework naturally accommodates nonstationarities in cardiac signals while maintaining interpretability through time-varying filter coefficients.

Experimental results on the PTB-XL dataset demonstrated that reconstruction performance closely follows known anatomical and electrophysiological principles. Adjacent precordial leads exhibited strong mutual predictability, reflecting their spatial proximity and shared near-field cardiac activity, whereas limb leads showed comparatively lower redundancy and contributed more unique information. The multi-lead reconstruction analysis further highlighted differential levels of lead redundancy, reinforcing the value of exploiting complementary information across leads to improve reconstruction fidelity.

Beyond lead synthesis, the proposed method provides a principled signal decomposition into shared and lead-specific components, enabling quantitative assessment of redundancy and information content across the 12-lead ECG. These insights have practical implications for optimized lead selection, ECG compression, and the design of wearable or resource-constrained monitoring systems. Future work will investigate pathology-aware modeling, nonlinear fusion strategies, and real-time deployment on embedded platforms to further extend the clinical and technological impact of this approach.

References

[1] G. D. Clifford, F. Azuaje, and P. E. McSharry, "Advanced methods and tools for ecg data analysis," *Artech House*, 2006.

[2] A. L. Goldberger, *Clinical Electrocardiography: A Simplified Approach*, 7th ed. Philadelphia, PA: Elsevier Health Sciences, 2006.

[3] A. H. Ribeiro, M. H. Ribeiro, G. M. M. Paixão, D. M. Oliveira, P. R. Gomes, J. A. Canazart, M. P. Ferreira, C. R. Andersson, P. W. Macfarlane, W. Meira, T. B. Schön, and A. L. Ribeiro, "Automatic diagnosis of the 12-lead ECG using a deep neural network," *Nat. Commun.*, vol. 11, p. 1760, Apr. 2020.

[4] D. Zhang, S. Yang, X. Yuan, and P. Zhang, "Interpretable deep learning for automatic diagnosis of 12-lead electrocardiogram," *iScience*, vol. 24, no. 4, p. 102373, Apr. 2021.

[5] G. E. Dower, "A lead synthesis for the vectorcardiogram," *American Heart Journal*, vol. 54, pp. 555–567, 1957.

[6] L. Edenbrandt and O. Pahlm, "Vectorcardiogram synthesized from a 12-lead ecg: superiority of the inverse dower matrix," *Journal of Electrocardiology*, vol. 21, no. 4, pp. 361–367, 1988.

[7] J. A. Kors *et al.*, "Reconstruction of the frank vectorcardiogram from standard ecg leads: diagnostic comparison of different methods," *Journal of Electrocardiology*, vol. 23, no. 4, pp. 313–323, 1990.

[8] M. H. Ostertag and G. R. Tsouri, "Reconstructing ecg precordial leads from a reduced lead set using independent component analysis," in *Proceedings of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2011, pp. 4168–4171.

[9] D. E. P. Moghaddam, A. Banta, A. Post, M. Razavi, and B. Aazhang, "Reconstructing 12-lead ecg from reduced lead sets using an encoder–decoder convolutional neural network," *Computers in Biology and Medicine*, vol. 122, p. 103861, 2020.

[10] A. Dutta and M. Das, "Ecg signal denoising using adaptive unscented kalman filter," in *2021 International Conference on Communication, Circuits, and Systems (IC3S)*. IEEE, 2021, pp. 139–144.
[11] R. Sameni, "A linear kalman notch filter for power-line interference cancellation," in *Proceedings of the 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP)*. IEEE, 2008, pp. 1–6.
[12] P. Wagner, N. Strodthoff, R. D. Bousseljot, D. Kreiseler, F. Lunze, T. Samek, and W. Schaeffter, "Ptb-xl, a large publicly available electrocardiography dataset," *Scientific Data*, vol. 7, no. 1, p. 154, 2020.

APPENDIX

In a Kalman filter, the innovation (or residual) is defined as the difference between the observed measurement and its predicted value,

$$\nu_k = y_k - \hat{y}_k^-, \tag{1}$$

where y_k denotes the measurement at time k and $\hat{y}_k^-$ is the predicted measurement obtained from the prior state estimate. Under the standard Kalman filter assumptions of linear dynamics and zero-mean Gaussian process and measurement noise, the innovation sequence forms a zero-mean, white stochastic process.

The covariance of the innovation is given by

$$S_k = H_k P_k^- H_k^\top + R_k, \tag{2}$$

where P_k^- is the prior state covariance, H_k is the measurement matrix, and R_k is the measurement noise covariance. If the filter model and noise statistics are correctly specified, the innovation satisfies

$$\nu_k \sim \mathcal{N}(0, S_k). \tag{3}$$

A common approach for assessing Kalman filter consistency is to examine the fraction of innovations that lie within confidence bounds derived from the predicted covariance. For a scalar measurement, approximately 68% of the innovations are expected to lie within the interval $\pm\sqrt{S_k}$, corresponding to the one-standard-deviation confidence region of a Gaussian distribution. Deviations from this expected coverage indicate potential miscalibration of the assumed noise statistics.

Figure 4 presents the innovation sequence for a representative record together with the predicted $\pm 1\sigma$ innovation envelope. The innovations remain centered around zero, and approximately 68% of the samples fall within the $\pm 1\sigma$ bounds, in agreement with the theoretical expectation for Gaussian innovations under correct covariance estimation. Occasional excursions beyond the envelope correspond to transient deviations in the signal, while the overall coverage indicates that the Kalman filter is well calibrated for this example.

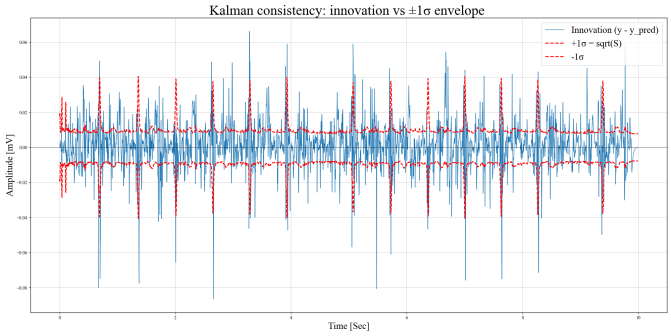

Fig. 4. Kalman filter innovation sequence overlaid with the predicted $\pm 1\sigma$ bounds, where $\sigma = \sqrt{S}$. Approximately 68% of the innovations lie within the envelope, consistent with the theoretical expectation for Gaussian innovations with correctly estimated covariance.