

Article

Adaptive Regularized Numerical Differentiation of Noisy Signals with Application to RoCoF Estimation in Power Systems

Farshad Merrikh-Bayat 

Department of Physics and Engineering, University of Scranton, 800 Linden St, Scranton, PA 18510, USA;
farshad.merrikhbayat@scranton.edu

Abstract

The total variation regularization (TVR) is one of the methods widely used for numerical differentiation of noisy signals. In this method, an appropriate value must be assigned to the regularization parameter. Although various approaches have been proposed for this purpose, their practical application may still depend on problem characteristics, prior knowledge, or additional tuning. Consequently, reliable automatic selection of the regularization parameter remains an important issue. The goal of this study is to develop a new adaptive method for estimating the derivative of a discrete-time noisy signal based on the TVR. In this method, at any moment of time, first the optimal value of the regularization parameter is calculated for a fixed-length sliding window over the most recent samples. Subsequently, for the obtained optimal regularization parameter, the numerical derivative of the samples in the sliding window are computed. The main contribution of the proposed method is in the adaptive strategy developed for automatic selection of the regularization parameter from the data which assigns small values to the regularization parameter when the signal changes rapidly, and vice versa. The proposed method is used for accurate and effective rate of change of frequency (RoCoF) estimation in power systems under different circumstances.

Keywords: numerical differentiation; noisy measurement; total variation; adaptive regularization; RoCoF estimation; low-inertia power system

MSC: 65J20; 65D25; 65F22

1. Introduction

Many algorithms used in scientific and engineering applications rely on the numerical derivative of a noisy signal. Some of the well-known applications are in image processing, biomedical signal analysis, control systems, and real-time monitoring [1–4]. Despite its apparent simplicity, numerical differentiation can be very challenging in the presence of measurement noise. The principal challenge arises from the fact that the measurement noise is typically broadband in nature and differentiation intrinsically amplifies high-frequency components. As a result, direct application of finite-difference methods to noisy signals often leads to severely degraded derivative estimates [5,6]. A comprehensive review of alternative approaches is beyond the scope of this paper, and just some of the most important numerical differentiation methods have been mentioned in the following. One of the well-known approaches for this purpose is the class of smoothing-based methods, in which the signal is first smoothed by passing it through a low-pass filter, and then its numerical derivative is computed [2]. A Butterworth filter followed by a finite difference



Academic Editor: Zhijia Zhao

Received: 14 April 2026

Revised: 9 June 2026

Accepted: 18 June 2026

Published: 2 July 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

differentiator is a widely used combination in the class of smoothing-based methods [2]. The main advantages of this method are its relatively easy implementation and low computational cost, which are obtained at the cost of lower accuracy because of suppressing high-frequency information.

Another approach for the numerical differentiation of noisy signals is to first build a local model by fitting it to the noisy signal, and then estimate the numerical derivative by computing the derivative of the model. The Savitzky–Golay (SG) filter is probably the most widely used method for modeling noisy signals. This method reduces the noise by fitting a polynomial to the samples of the signal in a sliding window using linear least-squares [7]. The order of the polynomial and the length of window are tunable parameters of the model in this approach. The key advantages of the SG filter are maintaining the relative positions of maxima, minima, and the widths (but not the heights) of the peaks. Also, the computational cost of the SG approach is significantly lower than the observer- and regularization-based approaches, which are reviewed in the following sections. The derivative obtained from the SG method is jagged because, in general, the polynomial model changes from one window to the next. The other challenge in using the SG method is the proper selection of the polynomial order and the window length. Choosing fixed values for these parameters may not lead to satisfactory results. The reason is that these parameters inherently balance noise reduction against peak distortion. Hence, if the parameters are chosen such that the result is satisfactory when the signal changes slowly, it will be unsatisfactory during fast transients, and vice versa.

Another class of differentiators is the observer-based method, where the derivative is defined as a hidden state within the state-space model of the dynamical system. For this method to be effective, an accurate model of the dynamical system must be available and also the noise characteristics must be known. Since in this paper it is assumed that neither the mathematical model of the dynamical system generating the noisy signal nor the noise characteristics are known, this approach is not useful. Finally, regularization-based formulations are another class of differentiators widely used for the numerical differentiation of non-smooth data as they preserve sharp transitions (edges) while effectively suppressing the noise [3,6,8]. One of the main challenges of using regularization-based methods is choosing an appropriate value for the regularization parameter. The ability of regularization-based methods to make a desired tradeoff between the noise attenuation and the preservation of transient features is mainly adjusted by this parameter. The common practice is to assign a constant value obtained by trial and error to this parameter.

In the modern electric power systems, numerical differentiators are used to estimate the rate of change of frequency (RoCoF) from the noisy measurements of frequency. RoCoF is defined as the time derivative of the grid frequency and is a key indicator for an imbalance between power generation and demand following disturbances. It has different applications like system inertia detection, anti-islanding, fast frequency control, stabilizing low-inertia grids, load shedding, etc. [9–12]. The growing share of distributed renewable generation in modern power systems has substantially reduced the effective synchronous rotational inertia of the grid. As a well-known fact, lower rotational inertia leads to faster frequency dynamics, and consequently, larger RoCoF following system disturbances, and highlights the importance of accurate RoCoF estimation in the monitoring and control of modern power systems more than ever [9–11].

A variety of RoCoF estimation techniques can be found in the literature [13–16]. These techniques generally involve different trade-offs between computational complexity, noise robustness, estimation delay, and transient accuracy, particularly in low-inertia power systems experiencing fast frequency dynamics. It should be noted that the techniques used to compute the time derivative of frequency from the noisy measurements are basically

the same as the techniques used to compute the derivative of any other noisy signal. The simplest approaches are based on applying finite-difference approximations directly to sampled frequency measurements. Although such approaches are computationally very effective, they are highly sensitive to measurement noise. Several filtering-based and windowed techniques have been developed to mitigate this limitation, including moving-average differentiation, polynomial fitting, and digital filtering approaches commonly used in phasor measurement units (PMUs) [10,15]. In PMU-based implementations, in addition to the measurement noise, factors such as the window length and operating conditions can also significantly affect the accuracy and practical applicability of RoCoF estimation algorithms. Other studies have explored observer-based techniques, which use the mathematical model of the system's frequency dynamics to estimate RoCoF [16]. The main problem with filtering-based and observer-based techniques is their reduced accuracy during rapid transients, which makes them inapplicable to low-inertia systems. Moreover, observer-based methods for RoCoF estimation suffer from high sensitivity to measurement noise. They also need accurate models, which are hard to obtain especially in islanded or weak grids.

Regularization-based differentiation techniques provide an effective option for computing the numerical derivative of noisy signals that contain rapid changes or sharp edges. In many existing approaches (see, for example [6]), the differentiation problem is formulated in the continuous domain as the minimization of a functional that balances data fidelity and regularization terms. Then, application of the Euler–Lagrange formalism to find the optimal solution of the functional leads to a partial differential equation (PDE), which is often nonlinear. Finally, numerical solution of the PDE yields the differentiated signal [6,17–19]. Although this functional-PDE framework is mathematically elegant, it leads to practical complications related to boundary conditions, discretization choices, and the computational cost of solving nonlinear PDEs [20].

In this paper, the numerical differentiation problem is formulated directly in discrete space. Since measured signals are inherently sampled, differentiation and integration operators can be naturally represented using matrix operators acting on the vectors containing discrete data [21–23]. This fact motivates discrete formulations in which the derivative of the noisy signal under consideration is obtained by minimizing a finite-dimensional cost function defined using discrete integration and differentiation operators. The objective of this study is to develop a discrete, regularized differentiation framework that circumvents the functional formulation and the associated Euler–Lagrange PDE machinery.

Proper selection of the regularization parameter is often the main difficulty when a regularization-based differentiation technique is used in practice. This parameter balances data fidelity with noise suppression to avoid over-smoothing. Some of the classical parameter-selection strategies that are widely used in regularization-based differentiation are the Morozov discrepancy principle, L-curve method, and Generalized Cross-Validation (GCV). Each of these strategies has its own advantages and disadvantages. For example, the limitations of the Morozov discrepancy principle include requiring an accurate estimate of the noise variance and slow convergence. Also, among others, L-curve and GCV strategies have the limitation that they are computationally expensive [20,24].

To partially address the above-mentioned issues, this article proposes an adaptive regularization parameter selection strategy based on the structural behavior of the cost function components. More specifically, instead of directly minimizing the total composite cost function, the proposed methodology examines the behavior of its individual constituent terms as functions of the regularization parameter. Numerical evidence indicates the existence of plateau regions in which certain regularization-sensitive terms exhibit

saturation-like behavior. These plateau regions provide a simple and computationally efficient indicator for selecting suitable values for the regularization parameters.

The proposed method can be applied either globally to an entire signal or locally within sliding temporal windows. The windowed implementation, which is particularly suitable for real-time applications, enables the regularization parameter to adapt dynamically to time-varying noise conditions and transient signal characteristics. This capability is used in this study for frequency estimation in power systems following system disturbances. Several simulations are performed on synthetic signals and power system frequency disturbance scenarios using the proposed methodology. The simulation results demonstrate that the proposed methodology provides accurate derivative estimates, even in the presence of significant measurement noise.

The main contributions of this work can be summarized as follows:

- An adaptive regularization parameter selection strategy is proposed for a discrete regularized numerical differentiation framework. The proposed strategy is based on the plateau behavior of the constituent terms of the cost function and does not require prior knowledge of the noise variance.
- The numerical differentiation problem is formulated directly in discrete space using matrix representations of the integration and differentiation operators, thereby avoiding the need for solving Euler–Lagrange partial differential equations commonly encountered in variational differentiation methods.
- The effectiveness of the proposed adaptive differentiator is verified by applying it to three RoCoF estimation case studies in low-inertia power systems under different disturbance scenarios.

The rest of this paper is organized as follows. The proposed method for adaptive regularized numerical differentiation of noisy signals is presented in Section 2. Three case studies, in which the RoCoF is calculated under different operating scenarios using the proposed adaptive method, are presented in Section 3. Section 4 contains discussions on the simulation results, findings, and their implications, and finally, Section 5 concludes the paper.

2. Proposed Adaptive Regularized Differentiation Framework

Consider the model equation

$$\hat{z}(t) = z(t) + \epsilon(t), \quad (1)$$

where $\hat{z}(t)$ is the continuous-time signal obtained from measurement, $z(t)$ is the noise-free (true) signal, $\epsilon(t)$ is the measurement noise and possible sensor errors, and t is the time. The problem under consideration is to estimate $v(t) = dz(t)/dt$ from the measured noisy signal $\hat{z}(t)$ without having any information about $\epsilon(t)$. For this purpose, since all the required computations will be carried out using a digital processor, we assume in the following that the measured noisy signal, $\hat{z}(t)$, has been sampled at an appropriate frequency and that only these samples are available to us. Therefore, all of the subsequent equations will be formulated based on these samples, and the proposed algorithm will also provide the samples of $v(t)$. Another important point is that the proposed algorithm performs the computations for the samples of $\hat{z}(t)$ within a selected window of prespecified length. That is, it takes as input the samples of $\hat{z}(t)$, which are located in the window, and estimates the samples of $v(t)$ over the same window.

Denote the samples of noisy measurement within the selected window as $\hat{z}_k = \hat{z}(kT)$, $k = 0, \dots, n$, and the corresponding vector containing noisy samples as $\hat{\mathbf{z}} = [\hat{z}_0, \dots, \hat{z}_n]^\top \in \mathbb{R}^{n+1}$ where T is the sampling period. Similarly, denote the vector containing the samples of

derivative as $\mathbf{v} = [v_0, \dots, v_n]^\top \in \mathbb{R}^{n+1}$ where $v_k = v(kT)$, $k = 0, \dots, n$. We define the discrete-time derivative of $z(t)$ on $t \in [0, nT]$ as a minimizer of the cost function

$$J(\lambda, \mathbf{v}) = \lambda \|\mathbf{D}\mathbf{v}\|_2^2 + \|\mathbf{A}\mathbf{v} - \hat{\mathbf{z}} + \hat{z}_0 \mathbf{1}_{n+1}\|_2^2, \quad (2)$$

where $\mathbf{D} \in \mathbb{R}^{(n+1) \times (n+1)}$, $\mathbf{A} \in \mathbb{R}^{(n+1) \times (n+1)}$, $\lambda > 0$, and $\|\cdot\|_2$ are the discrete differentiation operator, discrete integration (anti-differentiation) operator, regularization parameter, and l^2 -norm of a vector, respectively. In this paper, the matrices $\mathbf{D}$ and $\mathbf{A}$ are defined as follows

$$\mathbf{D} = \frac{1}{T} \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 1 \\ 0 & \cdots & 0 & 0 & 0 \end{bmatrix}, \quad (3)$$

$$\mathbf{A} = T \begin{bmatrix} \frac{1}{2} & 0 & 0 & \cdots & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & \cdots & 0 \\ \frac{1}{2} & 1 & \frac{1}{2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2} & 1 & 1 & \cdots & \frac{1}{2} \end{bmatrix}, \quad (4)$$

which calculate the numerical derivative and integral of a discrete-time signal using the first-order finite-difference method and trapezoidal rule, respectively. For this purpose, these matrices must be multiplied from the left by the vector containing the samples of the (noisy) signal. Note that since no forward information is available at the final sample, the last row of $\mathbf{D}$ is set equal to zero. The trapezoidal integration operator is adopted because of its simplicity, second-order accuracy, and favorable numerical compatibility with the chosen discrete differentiation operator, resulting in an operator pair for which $\mathbf{DA} \approx \mathbf{I}$ holds to a good approximation. The vector $\mathbf{1}_{n+1}$ in (2) is defined as

$$\mathbf{1}_{n+1} := [1 \ 1 \ \dots \ 1]^\top \in \mathbb{R}^{n+1}. \quad (5)$$

Note that in (2), the vector $\hat{\mathbf{z}} - \hat{z}_0 \mathbf{1}_{n+1}$ has the same variations as $\hat{\mathbf{z}}$ with the difference that its first sample is equal to zero. The term $\hat{z}_0 \mathbf{1}_{n+1}$ should be considered in (2) since the numerical integral of $\mathbf{v}$ obtained from $\mathbf{A}\mathbf{v}$ also starts from zero.

In the absence of noise, λ in (2) can be set to zero. In this case, the $\mathbf{v}$ that minimizes J in (2) actually makes the anti-differential of $\mathbf{v}$ as close as possible to $\hat{\mathbf{z}} - \hat{z}_0 \mathbf{1}_{n+1}$. Therefore, such an optimal $\mathbf{v}$ must contain the samples of a numerical derivative of $z(t) - z(0)$, or equivalently, a numerical derivative of $z(t)$. On the other hand, $\|\mathbf{D}\mathbf{v}\|_2^2$ in (2) penalizes irregularity in $\mathbf{v}$. More precisely, in dealing with noisy data, by setting $\lambda = 0$, the resulting $\mathbf{v}$ will contain undesired large spikes. Increasing λ makes the resulting $\mathbf{v}$ smoother at the cost of a larger discrepancy between $\mathbf{A}\mathbf{v}$ and $\hat{\mathbf{z}} - \hat{z}_0 \mathbf{1}_{n+1}$.

In this paper, to find the $\mathbf{v}$ and λ that minimize the cost function in (2), we solve the following optimization problem (for the sake of simplicity in notation, $\mathbf{1}_{n+1}$ is replaced with $\mathbf{1}$ in the subsequent):

$$\min_{\lambda, \mathbf{v}} J(\lambda, \mathbf{v}) = \lambda (\mathbf{D}\mathbf{v})^\top (\mathbf{D}\mathbf{v}) + (\mathbf{A}\mathbf{v} - \hat{\mathbf{z}} + \hat{z}_0 \mathbf{1})^\top (\mathbf{A}\mathbf{v} - \hat{\mathbf{z}} + \hat{z}_0 \mathbf{1}) \quad (6)$$

$$= \mathbf{v}^\top (\lambda \mathbf{D}^\top \mathbf{D} + \mathbf{A}^\top \mathbf{A}) \mathbf{v} - 2(\hat{\mathbf{z}} - \hat{z}_0 \mathbf{1})^\top \mathbf{A}\mathbf{v} + (\hat{\mathbf{z}} - \hat{z}_0 \mathbf{1})^\top (\hat{\mathbf{z}} - \hat{z}_0 \mathbf{1}). \quad (7)$$

The optimal $\mathbf{v}$ must make the gradient of the cost function zero, that is,

$$\nabla_{\mathbf{v}} J(\lambda, \mathbf{v}) = 2(\lambda \mathbf{D}^{\top} \mathbf{D} + \mathbf{A}^{\top} \mathbf{A}) \mathbf{v} - 2\mathbf{A}^{\top} (\hat{\mathbf{z}} - \hat{z}_0 \mathbf{1}) = 0. \quad (8)$$

By solving the above equation, the optimal $\mathbf{v}$ is obtained as follows

$$\mathbf{v}_{\text{opt}} = (\lambda \mathbf{D}^{\top} \mathbf{D} + \mathbf{A}^{\top} \mathbf{A})^{\dagger} \mathbf{A}^{\top} (\hat{\mathbf{z}} - \hat{z}_0 \mathbf{1}), \quad (9)$$

where $\dagger$ denotes the Moore–Penrose pseudo-inverse of a matrix. By substituting $\mathbf{v}_{\text{opt}}$ from (9) in (2), the single variable cost function is obtained as follows:

$$J(\lambda) = \lambda f(\lambda) + g(\lambda), \quad (10)$$

where

$$f(\lambda) := \|\mathbf{D} \mathbf{v}_{\text{opt}}\|_2^2 = \left\| \mathbf{D} (\lambda \mathbf{D}^{\top} \mathbf{D} + \mathbf{A}^{\top} \mathbf{A})^{\dagger} \mathbf{A}^{\top} (\hat{\mathbf{z}} - \hat{z}_0 \mathbf{1}) \right\|_2^2, \quad (11)$$

and

$$g(\lambda) := \|\mathbf{A} \mathbf{v}_{\text{opt}} - \hat{\mathbf{z}} + \hat{z}_0 \mathbf{1}\|_2^2 = \left\| \mathbf{A} (\lambda \mathbf{D}^{\top} \mathbf{D} + \mathbf{A}^{\top} \mathbf{A})^{\dagger} \mathbf{A}^{\top} (\hat{\mathbf{z}} - \hat{z}_0 \mathbf{1}) - (\hat{\mathbf{z}} - \hat{z}_0 \mathbf{1}) \right\|_2^2. \quad (12)$$

The λ is a scalar variable, and therefore one may expect that its optimal (or near-optimal) value can be obtained by a simple one-dimensional search over $J(\lambda)$ in (10). But, this approach is generally unreliable in practice. The reason is that, because of the numerical structure of regularized inverse problems, the cost functions such as $J(\lambda)$ in (10) typically exhibit a shallow and weakly localized minimum. Such a minimum happens because the two constituent terms $f(\lambda)$ and $g(\lambda)$ in (10) change in opposite ways: the regularization term $f(\lambda)$ is a monotonically decreasing function of λ , while the data fidelity term $g(\lambda)$ is a monotonically increasing function of it. Around the balancing regime, where these opposing effects approximately cancel, both functions approach saturation regions in which their variations with respect to λ become very small. As a result, $J(\lambda)$ often exhibits a wide, nearly flat valley rather than a sharp local minimum. In such a regime, direct minimization of $J(\lambda)$ becomes highly sensitive to numerical noise and the particular sampling of λ , which can in turn yield unstable or non-reproducible parameter estimates.

This phenomenon is a well-known issue in the regularization literature and motivates alternative parameter selection strategies that avoid direct minimization of the total cost function. A common practice is to exploit structural features of the constituent terms of the cost function such as the L-curve criterion [25], the discrepancy principle [26], and quasi-optimality rules [21,27]. The objective of these methodologies is to determine regions in the regularization-parameter space where further increases in the regularization parameter no longer produce meaningful changes in the regularized solution. Identifying such regions enables the selection of regularization parameter values that are both numerically stable and physically meaningful.

In this study, the point at which $f(\lambda)$ is at its flattest (plateau region) is considered as a proxy for locating the minimizer of $J(\lambda)$. Both empirical observations and theoretical analysis indicate that this plateau region corresponds to the regime in which the two competing constituent terms of $J(\lambda)$ are balanced. Consequently, it provides a robust and computationally efficient criterion for choosing an appropriate value for the regularization parameter. The existence of plateau regions in $f(\lambda)$ and $g(\lambda)$, and their relationship to the minimizer of $J(\lambda)$, are stated in the following proposition.

Proposition 1 (Plateau structure of the total cost function and its constituent terms). *Let $\mathbf{A}$ and $\mathbf{D}$ denote discrete numerical integration and differentiation operators, respectively, satisfying $\mathbf{D}\mathbf{A} \approx \mathbf{I}$. For $\lambda > 0$, define*

$$\mathbf{P}(\lambda) := (\lambda \mathbf{D}^\top \mathbf{D} + \mathbf{A}^\top \mathbf{A})^\dagger, \quad (13)$$

and let

$$\mathbf{y}(\lambda) := \mathbf{D}\mathbf{P}(\lambda)\mathbf{A}^\top (\hat{\mathbf{z}} - \hat{\mathbf{z}}_0 \mathbf{1}). \quad (14)$$

Consider the functions

$$f(\lambda) := \|\mathbf{y}(\lambda)\|_2^2, \quad g(\lambda) := \|\mathbf{A}\mathbf{P}(\lambda)\mathbf{A}^\top (\hat{\mathbf{z}} - \hat{\mathbf{z}}_0 \mathbf{1}) - (\hat{\mathbf{z}} - \hat{\mathbf{z}}_0 \mathbf{1})\|_2^2, \quad (15)$$

and the total cost

$$J(\lambda) := \lambda f(\lambda) + g(\lambda). \quad (16)$$

Then the following properties hold.

1. (Plateau behavior of f) The function $f(\lambda)$ is non-increasing in λ and exhibits a saturation (plateau) region for intermediate values of λ , in which its relative variation with respect to $\log \lambda$ becomes small.
2. (Plateau behavior of g) The function $g(\lambda)$ is non-decreasing in λ and exhibits a plateau region over approximately the same range of λ as $f(\lambda)$.
3. (Existence and localization of a minimizer of J) The total cost $J(\lambda)$ admits at least one minimizer $\lambda_{\text{opt}} > 0$. Moreover, any such minimizer lies within, or in close proximity to, the common plateau region of $f(\lambda)$ and $g(\lambda)$.

Proof. Sketch of proof and explanation: We analyze the spectral behavior of the regularized inverse operator $\mathbf{P}(\lambda)$. Let $\{\sigma_k\}$ denote the generalized singular values associated with the matrix pair $(\mathbf{D}, \mathbf{A})$. Then $\mathbf{P}(\lambda)$ acts as a spectral filter with gains of the form

$$\phi_k(\lambda) = \frac{1}{\lambda \sigma_k^2 + 1},$$

up to scaling.

- (a) The quantity $f(\lambda)$ measures the energy of the differentiated estimate and thus emphasizes high-frequency components. As λ increases, the filter $\phi_k(\lambda)$ progressively attenuates the contribution of modes associated with large values of σ_k , thereby inducing a monotonic decay of the function $f(\lambda)$. Once the noise-dominated components are sufficiently attenuated, further increases in λ affect primarily low-energy modes, resulting in a plateau region.
- (b) The function $g(\lambda)$ represents the data misfit. For small λ , the operator $\mathbf{A}\mathbf{P}(\lambda)\mathbf{A}^\top$ approximates the identity, yielding small values of $g(\lambda)$. As λ grows, regularization induces increasing bias, causing $g(\lambda)$ to increase. However, once the estimator is dominated by smooth components, further regularization changes the reconstruction marginally, producing a plateau. Since both f and g are governed by the same spectral transition of $\phi_k(\lambda)$, their plateau regions occur over comparable ranges of λ .
- (c) The cost $J(\lambda) = \lambda f(\lambda) + g(\lambda)$ balances variance-like and bias-like terms. For small λ , the term $\lambda f(\lambda)$ dominates and decreases with λ , whereas for large λ , the increase of $g(\lambda)$ dominates. Continuity of J together with these opposing trends guarantees the existence of at least one minimizer. Since both f and g vary slowly within their plateau region, the minimum of J necessarily lies within or near this region.

□

Based on the above explanations, we consider the optimal value of λ the one for which $f(\lambda)$ exhibits the maximum degree of flatness. It means that the optimal λ can be determined by computing the derivative of $f(\lambda)$ with respect to λ and determining the value of λ at which this derivative is nearest to zero. However, in practice we prefer to use the logarithmic derivative, defined as follows, instead of the ordinary derivative.

$$q(\lambda) := \frac{d}{d \log \lambda} \log f(\lambda) = \lambda \frac{d}{d \lambda} \log f(\lambda). \quad (17)$$

In our problem, this logarithmic differentiation has two main advantages compared to the ordinary differentiation. First, using the logarithm of $f(\lambda)$ decreases sensitivity to variations in its magnitude. Second, as seen in (17), the factor λ appears in the equation which rescales the derivative and balances the sensitivity across different values of λ .

According to (15), $f(\lambda)$ in (17) can be replaced by $\|\mathbf{y}(\lambda)\|_2^2$, where, according to (13) and (14), $\mathbf{y}(\lambda)$ is defined as follows:

$$\mathbf{y}(\lambda) := \mathbf{D}\mathbf{P}(\lambda)\mathbf{A}^\top (\hat{\mathbf{z}} - \hat{\mathbf{z}}_0\mathbf{1}) = \mathbf{D}(\lambda\mathbf{D}^\top\mathbf{D} + \mathbf{A}^\top\mathbf{A})^\dagger \mathbf{A}^\top (\hat{\mathbf{z}} - \hat{\mathbf{z}}_0\mathbf{1}). \quad (18)$$

Now, taking the logarithmic derivative of $f(\lambda)$ yields

$$\frac{d}{d \lambda} \log f(\lambda) = \frac{d}{d \lambda} \log \|\mathbf{y}(\lambda)\|_2^2 = \frac{2\mathbf{y}(\lambda)^\top \mathbf{y}'(\lambda)}{\|\mathbf{y}(\lambda)\|_2^2}, \quad (19)$$

where

$$\mathbf{y}'(\lambda) = \mathbf{D} \frac{d}{d \lambda} (\lambda\mathbf{D}^\top\mathbf{D} + \mathbf{A}^\top\mathbf{A})^\dagger \mathbf{A}^\top (\hat{\mathbf{z}} - \hat{\mathbf{z}}_0\mathbf{1})^\top. \quad (20)$$

Using the known formula for the derivative of the Moore–Penrose pseudo-inverse:

$$\frac{d}{d \lambda} (\lambda\mathbf{D}^\top\mathbf{D} + \mathbf{A}^\top\mathbf{A})^\dagger = -(\lambda\mathbf{D}^\top\mathbf{D} + \mathbf{A}^\top\mathbf{A})^\dagger \mathbf{D}^\top\mathbf{D} (\lambda\mathbf{D}^\top\mathbf{D} + \mathbf{A}^\top\mathbf{A})^\dagger, \quad (21)$$

we obtain

$$\mathbf{y}'(\lambda) = -\mathbf{D}\mathbf{P}(\lambda)\mathbf{D}^\top\mathbf{D}\mathbf{P}(\lambda)\mathbf{A}^\top (\hat{\mathbf{z}} - \hat{\mathbf{z}}_0\mathbf{1})^\top, \quad (22)$$

which according to (18) can be written as

$$\mathbf{y}'(\lambda) = -\mathbf{D}\mathbf{P}(\lambda)\mathbf{D}^\top\mathbf{y}(\lambda). \quad (23)$$

By substituting $\mathbf{y}'(\lambda)$ from (23) in (19) we get

$$\frac{d}{d \lambda} \log f(\lambda) = -\frac{2\mathbf{y}(\lambda)^\top \mathbf{D}\mathbf{P}(\lambda)\mathbf{D}^\top\mathbf{y}(\lambda)}{\|\mathbf{y}(\lambda)\|_2^2} \quad (24)$$

Finally, by substituting $d \log f(\lambda)/d \lambda$ from (24) in (17) we obtain

$$q(\lambda) = \frac{d}{d \log \lambda} \log f(\lambda) = -2\lambda \frac{\mathbf{y}(\lambda)^\top \mathbf{D}\mathbf{P}(\lambda)\mathbf{D}^\top\mathbf{y}(\lambda)}{\|\mathbf{y}(\lambda)\|_2^2}. \quad (25)$$

where $\mathbf{y}(\lambda)$ and $\mathbf{P}(\lambda)$ are defined in (18) and (13), respectively.

For any first-order differentiation matrix $\mathbf{D}$, the condition $\mathbf{D}\mathbf{x} = \mathbf{0}$ implies that $\mathbf{x}$ is a constant vector. On the other hand, for the integration matrix defined in (4), $\mathbf{A}\mathbf{x} \neq \mathbf{0}$ for every nonzero constant vector $\mathbf{x}$. Therefore, there exists no nonzero vector $\mathbf{x}$ such that $\mathbf{D}\mathbf{x} = \mathbf{0}$ and $\mathbf{A}\mathbf{x} = \mathbf{0}$ simultaneously. It follows that $\mathbf{x}^\top (\lambda\mathbf{D}^\top\mathbf{D} + \mathbf{A}^\top\mathbf{A})\mathbf{x} > 0$ for every nonzero vector $\mathbf{x}$, which proves that $\lambda\mathbf{D}^\top\mathbf{D} + \mathbf{A}^\top\mathbf{A}$ is positive definite (PD) and, consequently, so is

the matrix $\mathbf{P}(\lambda) = (\lambda \mathbf{D}^\top \mathbf{D} + \mathbf{A}^\top \mathbf{A})^\dagger$. Therefore, $q(\lambda) < 0$ for every $\lambda > 0$, and the optimal value of λ can be found from solving the following optimization problem

$$\lambda_{\text{opt}} = \arg \min_{\lambda} |q(\lambda)|. \quad (26)$$

Note that λ_{opt} is the value of λ for which $|q(\lambda)|$ is closest to zero, and consequently, $q(\lambda)$ attains its maximum degree of flatness. Given that the above optimization problem is single-variable, the simplest way to find λ_{opt} is to evaluate $q(\lambda)$ in (25) for several (logarithmically spaced) values of λ and consider the one for which the absolute value of $q(\lambda)$ is closest to zero as the optimal solution. More details are provided in the following section.

Algorithmic Summary of the Proposed Method

For ease of implementation, the proposed adaptive regularized differentiation method is summarized in Algorithm 1.

Algorithm 1 Adaptive regularized numerical differentiation within a moving window.

Require: Noisy samples $\hat{\mathbf{z}} = [\hat{z}_0, \hat{z}_1, \dots, \hat{z}_n]^\top \in \mathbb{R}^{n+1}$, sampling period T , matrix integration and differentiation operators $\mathbf{A}$ and $\mathbf{D}$, search bounds $\lambda_{\min}$ and $\lambda_{\max}$, and number of logarithmically spaced candidates N_λ .

Ensure: Estimated derivative vector $\mathbf{v}_{\text{opt}} \in \mathbb{R}^{n+1}$.

1: Generate logarithmically spaced candidate values

$$\Lambda \leftarrow \{\lambda_1, \lambda_2, \dots, \lambda_{N_\lambda}\} \subset [\lambda_{\min}, \lambda_{\max}].$$

2: **for** $i = 1, 2, \dots, N_\lambda$ **do**

3: Set $\lambda \leftarrow \lambda_i$.

4: Compute

$$\mathbf{P}(\lambda) \leftarrow (\lambda \mathbf{D}^\top \mathbf{D} + \mathbf{A}^\top \mathbf{A})^\dagger.$$

5: Compute

$$\mathbf{y}(\lambda) \leftarrow \mathbf{D} \mathbf{P}(\lambda) \mathbf{A}^\top (\hat{\mathbf{z}} - \hat{z}_0 \mathbf{1}).$$

6: Evaluate

$$q(\lambda) \leftarrow -2\lambda \frac{\mathbf{y}(\lambda)^\top \mathbf{D} \mathbf{P}(\lambda) \mathbf{D}^\top \mathbf{y}(\lambda)}{\|\mathbf{y}(\lambda)\|_2^2}.$$

7: **end for**

8: Select

$$\lambda_{\text{opt}} \leftarrow \arg \min_{\lambda_i \in \Lambda} |q(\lambda_i)|.$$

9: Compute

$$\mathbf{v}_{\text{opt}} \leftarrow (\lambda_{\text{opt}} \mathbf{D}^\top \mathbf{D} + \mathbf{A}^\top \mathbf{A})^\dagger \mathbf{A}^\top (\hat{\mathbf{z}} - \hat{z}_0 \mathbf{1}).$$

10: **return** $\mathbf{v}_{\text{opt}}$.

The bounds $\lambda_{\min}$ and $\lambda_{\max}$ define the admissible search interval over which $q(\lambda)$ (or equivalently, slope of the tangent to $f(\lambda)$ as an indicator of the level of flatness) is evaluated. In practice, this interval is chosen on a logarithmic scale and should be wide enough to contain the plateau region of $f(\lambda)$, or equivalently the local minimum of $|q(\lambda)|$. A simple practical procedure is to evaluate $|q(\lambda)|$ over an initially broad logarithmic range using a representative window of the measured signal, and then select $[\lambda_{\min}, \lambda_{\max}]$ so that the identified minimum is well inside the interval rather than being close to its endpoints. If the minimum occurs near either $\lambda_{\min}$ or $\lambda_{\max}$, the interval is enlarged and the preliminary scan is repeated. This procedure ensures that the adaptive search is performed over a numerically meaningful range of regularization parameters.

Note that a big portion of the computational load of the above algorithm lies in computing $\mathbf{P}(\lambda)$ as it requires computing the pseudo-inverse of $\lambda \mathbf{D}^\top \mathbf{D} + \mathbf{A}^\top \mathbf{A}$. But, this computation does not need to be repeated at each iteration of the algorithm. In fact, $(\lambda \mathbf{D}^\top \mathbf{D} + \mathbf{A}^\top \mathbf{A})^\dagger$ needs to be computed only once for each λ value of interest in the range $[\lambda_{\min}, \lambda_{\max}]$, and then the result can be reused whenever it is needed. Similarly, all matrix-matrix products of the proposed algorithm need to be done only once because they are independent of the noisy signal to be differentiated, and only matrix-vector products need to be repeated at each iteration of the algorithm.

3. Numerical Experiments

To evaluate the performance of the proposed differentiation method, in each of the following case studies, a synthetic signal representing typical frequency dynamics in a power system is generated. Then, this synthetic frequency signal is discretized and a measurement noise is added. Finally, the numerical derivative of the noisy synthetic frequency signal is computed using the proposed method and the result is compared with the exact derivative. In all of the following case studies, the nominal frequency of the system is assumed to be 50 Hz. All simulations were performed using MATLAB R2023a.

3.1. Case Study 1: Generator Trip Disturbance

Assume the frequency signal is modeled as

$$f(t) = 50 + f_{\text{drop}}(t) + f_{\text{rec}}(t) + f_{\text{osc}}(t), \quad (27)$$

where the three components represent the frequency drop due to sudden loss of generation, the recovery due to primary control, and a small electromechanical oscillation, respectively. The frequency drop following the disturbance is modeled as

$$f_{\text{drop}}(t) = -0.12(1 - e^{-3(t-5)})u(t-5), \quad (28)$$

where $u(\cdot)$ denotes the unit step function. The recovery of the frequency due to primary control is captured by the following model

$$f_{\text{rec}}(t) = 0.09(1 - e^{-0.4(t-6)})u(t-6). \quad (29)$$

Finally, the following small damped oscillatory component is included to model the electromechanical modes of the power system:

$$f_{\text{osc}}(t) = 0.02e^{-0.25(t-6)} \sin(2\pi 0.8(t-6))u(t-6). \quad (30)$$

To emulate realistic measurement conditions, a zero-mean Gaussian noise with the standard deviation $\sigma = 0.002$ Hz, denoted as $\eta(t)$, is added to the frequency signal as follows

$$f_{\text{meas}}(t) = f(t) + \eta(t). \quad (31)$$

The objective is to estimate the RoCoF defined as

$$\text{RoCoF}(t) = \frac{df(t)}{dt}, \quad (32)$$

from $f_{\text{meas}}(t)$ using the proposed adaptive differentiation method.

The results of a simulation are shown in Figure 1, where Figure 1a shows the noisy measured frequency, and Figure 1b,c show the results of applying the proposed adaptive numerical differentiation algorithm. In these simulations the sampling period and length of sliding window are considered as $T = 0.01$ s and $n = 200$. It implies that at any moment of time, the algorithm uses the past $n = 200$ samples (or equivalently, the samples from the last 2 s) to estimate the derivative at that moment. That is why in Figure 1b,c, the values of RoCoF and λ are calculated from $t = 2$ s onwards.

Figure 1b demonstrates the estimated RoCoF together with the true RoCoF. As can be seen, the two graphs are quite close to each other. The corresponding λ_{opt} is shown in Figure 1c. To find λ_{opt} at each iteration of the algorithm, the value of $q(\lambda)$ in (25) is computed for 20 different values of λ that are logarithmically distributed in the range $[10^{-6}, 10^0]$. As expected, the algorithm chooses rather large values for λ when the changes in frequency are equal or close to zero, and it assigns smaller values to this parameter as the frequency changes increase.

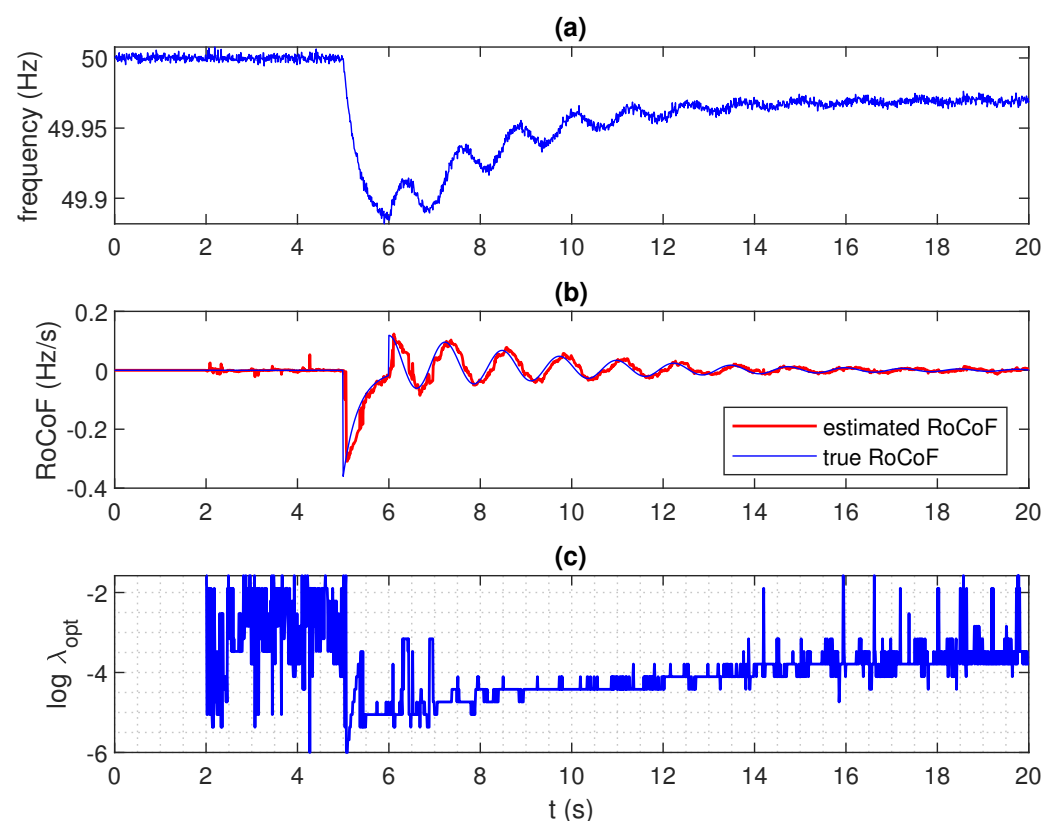


Figure 1. The simulation results of Case Study 1. (a) The measured noisy frequency, (b) true and estimated RoCoF using the adaptive method, and (c) variations of λ_{opt} .

To demonstrate the necessity of employing an adaptive strategy for tuning the value of λ , the outcomes of three simulations conducted with fixed values of the regularization parameter are presented in Figure 2. As observed, there exists no value of λ for which both the transient and steady-state values of RoCoF are fairly accurate. If, in a given problem, the temporal evolution of the frequency is such that a single fixed value of λ can effectively solve the problem, then the proposed method can be employed to determine this fixed value of λ .

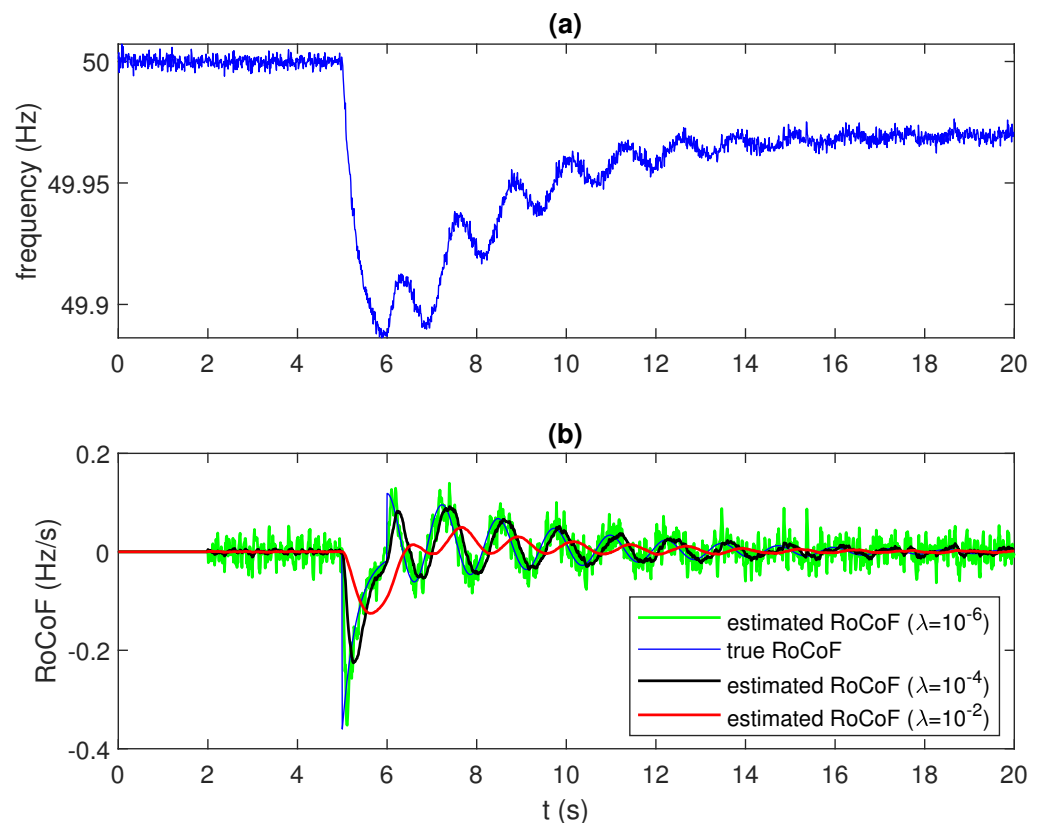


Figure 2. The simulation results of Case Study 1. (a) The measured noisy frequency, (b) true and estimated RoCoF for several fixed values of λ .

3.2. Case Study 2: Load Rejection Disturbance

In this case study, the proposed method is used to estimate RoCoF when a load rejection disturbance occurs in a power system. When a significant portion of load is suddenly disconnected from the system, the electrical power demand becomes smaller than the mechanical power provided by the generators, which results in a temporary increase in the system frequency. Then, primary frequency control mechanisms reduce the mechanical input of the generators to restore the power balance, which causes the frequency to gradually settle toward a new equilibrium. In this case, the frequency of the system can be modeled as follows

$$f(t) = 50 + f_{\text{rise}}(t) + f_{\text{corr}}(t) + f_{\text{osc}}(t), \quad (33)$$

where $f_{\text{rise}}(t)$, $f_{\text{corr}}(t)$, and $f_{\text{osc}}(t)$ represent the frequency rise due to load rejection, the corrective action of the primary frequency control, and a small damped oscillatory component, respectively. The constituent components of frequency in (33) are considered as follows

$$f_{\text{rise}}(t) = 0.10(1 - e^{-3(t-5)})u(t-5), \quad (34)$$

$$f_{\text{corr}}(t) = -0.07(1 - e^{-0.5(t-6)})u(t-6). \quad (35)$$

$$f_{\text{osc}}(t) = 0.15e^{-0.3(t-6)}\sin(2\pi 0.7(t-6))u(t-6). \quad (36)$$

A zero-mean Gaussian noise with the standard deviation $\sigma = 0.002$ Hz is also added to the frequency signal to emulate realistic measurement conditions.

The results of a simulation using the proposed adaptive numerical differentiation method are shown in Figure 3. The parameters used in this simulation are $\lambda_{\min} = 10^{-8}$,

$\lambda_{\max} = 10^{-1}$, $T = 0.01$, and $n = 200$. Moreover, to find λ_{opt} at each iteration of the algorithm, the value of $q(\lambda)$ in (25) is computed for 30 different values of λ that are logarithmically distributed in the range $[\lambda_{\min}, \lambda_{\max}]$. As seen in Figure 3b, the estimated RoCoF is quite close to the true RoCoF.

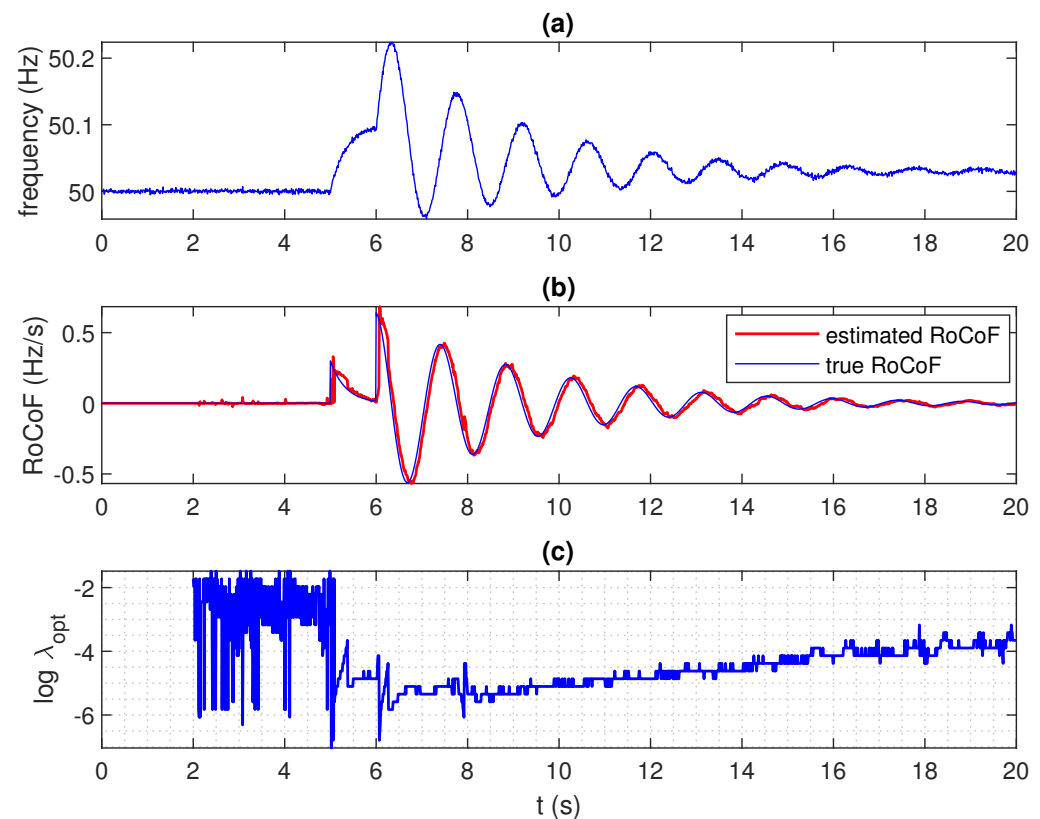


Figure 3. The simulation results of Case Study 2. (a) The measured noisy frequency, (b) true and estimated RoCoF, and (c) variations of λ_{opt} .

In Figure 4, the proposed method is compared with the SG method which estimates the derivative by performing a local least-squares polynomial fit over a moving window of data points and then analytically evaluating the derivative of that fitted polynomial [7]. To obtain this figure, the SG method was run using different values of the window length and polynomial order, and two of the results which looked better were selected. As can be seen, the proposed method produces a significantly smoother result than the SG method, though this improvement comes at the expense of increased computational cost. Moreover, the result produced by the proposed method exhibits very low latency relative to the exact derivative, and this latency decreases further as the noise level is reduced.

3.3. Case Study 3: Low-Inertia Power System Modeled by the Swing Equation

In order to evaluate the proposed differentiation method under a more realistic condition, a third case study is considered in which the system frequency dynamics are generated from the classical swing equation. This equation models the electromechanical behavior of synchronous generators and is widely used in stability analysis of power systems.

The swing equation describing the rotational dynamics of an equivalent generator can be written as

$$2H \frac{d\omega(t)}{dt} = P_m(t) - P_e(t) - D(\omega(t) - \omega_0), \quad (37)$$

where H denotes the inertia constant of the system, D is the damping coefficient, $P_m(t)$ and $P_e(t)$ represent the mechanical input power and electrical output power, respectively,

and $\omega_0 = 2\pi \times 50$ rad/s is the nominal angular speed corresponding to the rated system frequency.

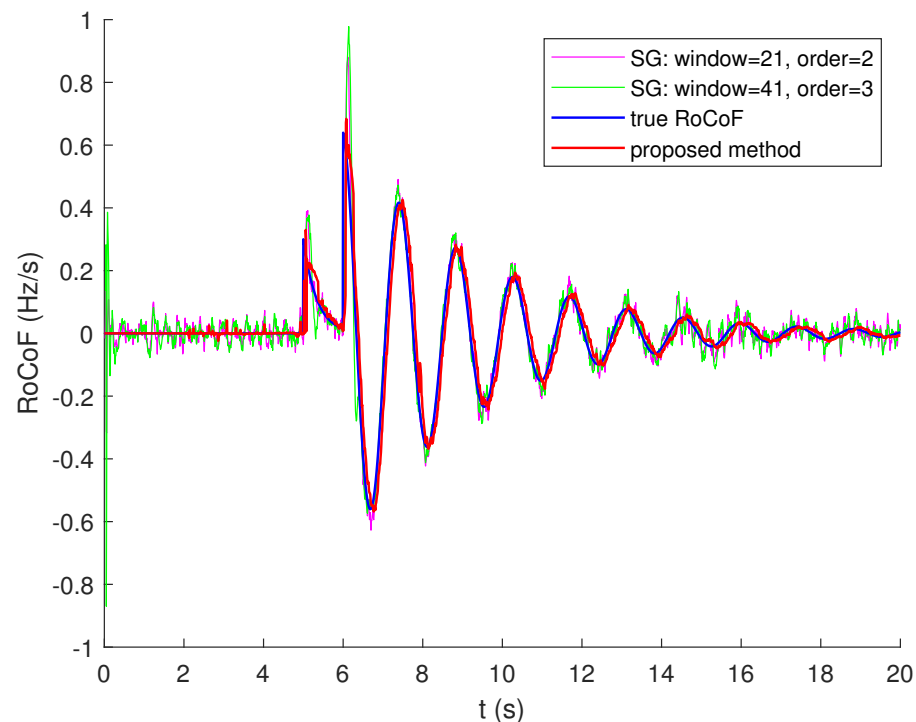


Figure 4. Comparison of the proposed method with the SG method for solving Case Study 2.

In this case study, we assume that a disturbance in the form of a sudden increase in the electrical load occurs at $t = t_d$. Such a disturbance leads to an imbalance between the mechanical input power and the electrical output power, which causes a temporary deceleration of the rotating masses and a drop in the system frequency. Then, the primary frequency control provided by turbine governors gradually increases the mechanical input power and therefore the frequency such that a new equilibrium operating point is achieved.

The mechanical input power provided by the primary frequency control to reject the electrical load disturbance is assumed to be a first-order response as follows

$$P_m(t) = P_{m0} + K_g \left(1 - e^{-(t-t_d)/T_g}\right) u(t - t_d), \quad (38)$$

where P_{m0} is the initial mechanical power, K_g is the governor gain, and T_g is the time constant of the governor response. The values of the parameters used in simulations are $K_g = 0.20$ p.u., $T_g = 2$ s, $H = 2.5$ s, $D = 1.2$ p.u./Hz, $P_{m0} = 1$ p.u., and $t_d = 5$ s. Also, the initial value of electrical load is considered as $P_{e0} = 1$ p.u., which is increased by the amount of $\Delta P = 0.6$ p.u. at $t_d = 5$ s.

After computation of $\omega(t)$ from (37), $f(t)$ is obtained as $f(t) = \omega(t)/2\pi$. Then, a zero-mean Gaussian noise with the standard deviation $\sigma = 0.001$ Hz is added to $f(t)$ to create $f_{\text{meas}}(t)$.

The proposed method is used to estimate $df(t)/dt$ from $f_{\text{meas}}(t)$. The results of a simulation assuming $\lambda_{\min} = 10^{-4}$, $\lambda_{\max} = 10^{-1}$, $T = 0.01$ s, and $n = 100$ are shown in Figure 5. To find λ_{opt} in Figure 5c, at each iteration of the algorithm, the value of $q(\lambda)$ in (25) is computed just for 10 different values of λ that are logarithmically distributed in the range $[\lambda_{\min}, \lambda_{\max}]$. As can be seen in Figure 5b, the estimated RoCoF is quite close to the true RoCoF.

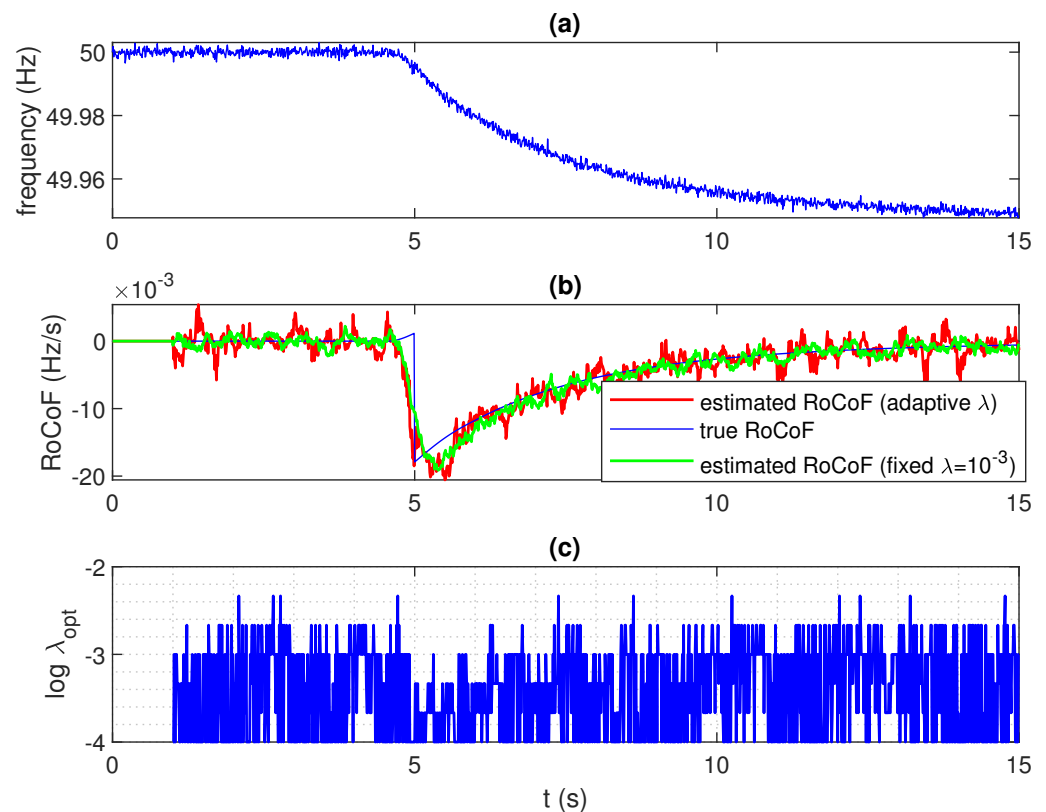


Figure 5. The simulation results of Case Study 3. (a) The measured noisy frequency, (b) true and estimated RoCoF, and (c) variations of λ_{opt} .

Unlike Figures 1c and 3c, no increasing or decreasing trend is observed in Figure 5c. This behavior may occur in problems whose noise level, signal bandwidth, and transient behavior remain relatively consistent over time, so that a single regularization parameter happens to be well suited to the entire signal. However, such a fixed value is typically not known in advance and can differ substantially under various operating conditions. The main advantage of the proposed adaptive approach is therefore not necessarily to outperform the best fixed regularization parameter selected in hindsight for every individual case study, but rather to provide an automatic parameter selection mechanism that remains applicable across different scenarios without requiring prior tuning. Based on Figure 5c, it appears that the problem can be solved using a fixed λ with a value between 10^{-3} and 10^{-4} . Note that using a fixed value of λ has the advantage of making the optimization problem in (2) convex, which results in a substantial decrease in the algorithm's computational expense. The result of a simulation for a fixed λ equal to 10^{-3} is depicted in Figure 5b. In this particular example, not only does the use of a fixed λ solve the problem, but the quality of the solution is even better than that obtained using the proposed adaptive method. This behavior can be explained as follows: in this problem, a fixed λ is close to optimal. However, in the proposed method, even if the optimal value of λ is constant, the algorithm will continuously vary λ around its optimal value due to the presence of noise. These fluctuations in λ , in turn, lead to oscillations in the estimated RoCoF.

4. Discussion

An adaptive method for estimating the derivative of discrete-time noisy measurements based on total variation regularization was developed. This method has many applications in different fields of engineering and science. For example, in control systems engineering, the time derivative of the error signal can be incorporated into the control law to enhance the system stability and performance.

In Section 3, the proposed method was used for estimating the RoCoF in power systems under diverse operating conditions. These conditions include initial frequency excursions in the upward and downward directions, frequency trajectories that are either monotonic or exhibit oscillatory behavior, and scenarios employing either a time-varying regularization parameter or one with a constant mean value. In each case study, the estimated RoCoF was compared with the exact one, and it was observed that in all cases, the estimated value is fairly accurate. It was also observed that when the eigenvalues associated with the frequency dynamics are well separated, the optimal regularization parameter exhibits significant variations and therefore the adaptive method should be applied.

However, when the eigenvalues associated with the frequency dynamics are closely spaced with respect to each other, or when the frequency behavior is effectively dominated by a single eigenvalue, it is appropriate to approximate the regularization parameter by a constant value. In this case, the computational cost of the proposed algorithm is substantially reduced, as for a given value of λ , the numerical derivative can be computed directly by solving a convex quadratic optimization problem.

A substantial fraction of the computational cost of the proposed adaptive algorithm is associated with evaluating the matrix $\mathbf{P}(\lambda) = (\lambda \mathbf{D}^\top \mathbf{D} + \mathbf{A}^\top \mathbf{A})^\dagger$, which can be calculated once for the prespecified values of λ in the range $[\lambda_{\min}, \lambda_{\max}]$, and reused as needed. Similarly, all the matrix-by-matrix multiplications of the proposed algorithm need to be done only once because they are independent of the noisy signal to be differentiated. Hence, for each sliding window it is only necessary to perform matrix–vector multiplications involving the vector of noisy measurements, $\hat{\mathbf{z}} - \hat{z}_0 \mathbf{1}$, using the fixed matrices that have been precomputed.

5. Conclusions

A new method for estimating the numerical derivative of a noisy signal based on TVR was proposed. One of the novelties of the proposed method is the representation of the corresponding optimization problem directly in discrete space without the need to apply the Euler–Lagrange method and then numerically solve the resulting nonlinear PDE. However, the main novelty of the proposed method is in the adaptive algorithm proposed for automatic adjustment of the regularization parameter. The simulation results indicate that when the eigenvalues associated with the dynamics of the noisy signal are significantly different, assigning a fixed value to the regularization parameter may not lead to satisfactory results. In this case, applying the proposed method leads to much better results compared to when a fixed value is chosen for the regularization parameter.

The proposed adaptive method was successfully used for estimating the RoCoF in power systems. Diverse numerical simulations showed that the proposed method can accurately estimate RoCoF from the noisy frequency measurements while it preserves the fast transients. It was also observed that the adaptive strategy is particularly beneficial when the dominant modes of the frequency dynamics are well separated, whereas a fixed regularization parameter may be sufficient when the system behavior is dominated by a single mode. Future work may focus on reducing the computational cost of the adaptive parameter selection procedure and extending the method to real-time applications involving experimentally measured signals.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are openly available in [GitHub] at [<https://github.com/fbayat-gif/RoCoF1.git>, accessed on 13 April 2026].

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

RoCoF	Rate of Change of Frequency
PMU	Phasor Measurement Unit
PDE	Partial Differential Equation
GCV	Generalized Cross-Validation
SG	Savitzky–Golay
PD	Positive Definite
TVR	Total Variation Regularization

References

1. Pragliola, M.; Calatroni, L.; Lanza, A.; Sgallari, F. On and beyond total variation regularization in imaging: The role of space variance. *SIAM Rev.* **2023**, *65*, 601–685. [\[CrossRef\]](#)
2. Van Breugel, F.; Kutz, J.N.; Brunton, B.W. Numerical differentiation of noisy data: A unifying multi-objective optimization framework. *IEEE Access* **2020**, *8*, 196865–196877. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Liu, T.; Zhang, W.; Wang, L.; Ueland, M.; Forbes, S.L.; Zheng, W.X.; Su, S.W. Numerical differentiation from noisy signals: A kernel regularization method to improve transient-state features for the electronic nose. *IEEE Trans. Syst. Man Cybern. Syst.* **2024**, *54*, 3497–3511. [\[CrossRef\]](#)
4. Yan, H.; Yang, Z.; Gao, J.; Wang, X. QRS detection in noisy electrocardiogram using an adaptively regularized numerical differentiation method. *Biomed. Signal Process. Control* **2025**, *105*, 107666. [\[CrossRef\]](#)
5. Jauberteau, F.; Jauberteau, J.L. Numerical differentiation with noisy signal. *Appl. Math. Comput.* **2009**, *215*, 2283–2297. [\[CrossRef\]](#)
6. Chartrand, R. Numerical differentiation of noisy, nonsmooth data. *ISRN Appl. Math.* **2011**, *2011*, 164564. [\[CrossRef\]](#)
7. Schafer, R.W. What is a Savitzky–Golay filter? [lecture notes]. *IEEE Signal Process. Mag.* **2011**, *28*, 111–117. [\[CrossRef\]](#)
8. Stickel, J.J. Data smoothing and numerical differentiation by a regularization method. *Comput. Chem. Eng.* **2010**, *34*, 467–475. [\[CrossRef\]](#)
9. Ulbig, A.; Borsche, T.S.; Andersson, G. Impact of low rotational inertia on power system stability and operation. *IFAC Proc. Vol.* **2014**, *47*, 7290–7297. [\[CrossRef\]](#)
10. Milano, F.; Dörfler, F.; Hug, G.; Hill, D.J.; Verbič, G. Foundations and challenges of low-inertia systems. In *Proceedings of the Power Systems Computation Conference (PSCC), Dublin, Ireland, 2018*; IEEE: Piscataway, NJ, USA, 2018; pp. 1–25.
11. Saleem, M.I.; Saha, S.; Roy, T.K.; Ghosh, S.K. Assessment and management of frequency stability in low inertia renewable energy rich power grids. *IET Gener. Transm. Distrib.* **2024**, *18*, 1372–1390. [\[CrossRef\]](#)
12. Tan, B.; Zhao, J.; Netto, M.; Krishnan, V.; Terzija, V.; Zhang, Y. Power system inertia estimation: Review of methods and the impacts of converter-interfaced generations. *Int. J. Electr. Power Energy Syst.* **2022**, *134*, 107362. [\[CrossRef\]](#)
13. Deng, X.; Mo, R.; Wang, P.; Chen, J.; Nan, D.; Liu, M. Review of RoCoF estimation techniques for low-inertia power systems. *Energies* **2023**, *16*, 3708.
14. Yin, H.; Wu, Y.; Qiu, W.; Zeng, C.; You, S.; Tan, J.; Hoke, A.; Kruse, C.J.; Rockwell, B.W.; Kawamura, K.A.; et al. Precise ROCOF estimation algorithm for low inertia power grids. *Electr. Power Syst. Res.* **2022**, *209*, 107968. [\[CrossRef\]](#)
15. Frigo, G.; Derviškić, A.; Zuo, Y.; Paolone, M. PMU-based ROCOF measurements: Uncertainty limits and metrological significance in power system applications. *IEEE Trans. Instrum. Meas.* **2019**, *68*, 3810–3822. [\[CrossRef\]](#)
16. Mir, A.S.; Singh, A.K.; Senroy, N. Robust observer based methodology for frequency and rate of change of frequency estimation in power systems. *IEEE Trans. Power Syst.* **2021**, *36*, 5385–5395. [\[CrossRef\]](#)
17. Rudin, L.I.; Osher, S.; Fatemi, E. Nonlinear total variation based noise removal algorithms. *Physica D* **1992**, *60*, 259–268. [\[CrossRef\]](#)
18. Vogel, C.R.; Oman, M.E. Iterative methods for total variation denoising. *SIAM J. Sci. Comput.* **1996**, *17*, 227–238. [\[CrossRef\]](#)
19. Knowles, I.; Wallace, R. A variational method for numerical differentiation. *Numer. Math.* **1995**, *70*, 91–110. [\[CrossRef\]](#)
20. Hansen, P.C. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Rev.* **1992**, *34*, 561–580. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Engl, H.W.; Hanke, M.; Neubauer, A. *Regularization of Inverse Problems*; Springer: Vienna, Austria, 1996.
22. Chung, J.; Gazzola, S. Computational methods for large-scale inverse problems: A survey on hybrid projection methods. *SIAM Rev.* **2024**, *66*, 205–284. [\[CrossRef\]](#)
23. Hansen, P.C. *Discrete Inverse Problems: Insight and Algorithms*; SIAM: Philadelphia, PA, USA, 2010.
24. Lv, D.; Li, T.; Yang, P.; Zhang, C.; Li, J. Adaptive regularization parameter adjustment for total variation denoising. *Signal Process.* **2026**, *243*, 110494. [\[CrossRef\]](#)
25. Hansen, P.C. The L-curve and its use in the numerical treatment of inverse problems. In *Computational Inverse Problems in Electrocardiology*; WIT Press: Southampton, UK, 2000.

26. Morozov, V.A. *Methods for Solving Incorrectly Posed Problems*; Springer: New York, NY, USA, 2012.
27. Tikhonov, A.N.; Arsenin, V.Y. *Solutions of Ill-Posed Problems*; Winston & Sons: Washington, DC, USA, 1977.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.