

Reliability-aware coupled-gradient distributed adaptive filtering for heterogeneous networks

Azam Khalili, Amir Rastegarnia

Distributed adaptive filtering over heterogeneous networks is challenging due to inconsistent local update directions, varying sensing quality, and unequal reliability across nodes. Conventional diffusion-based adaptive algorithms typically promote cooperation through parameter exchange, which may be insufficient to regulate discrepancies in local learning dynamics under heterogeneous conditions. This paper proposes a reliability-aware coupled-gradient distributed adaptive filtering framework in which cooperation is introduced through two complementary mechanisms: reliability-aware gradient coupling before adaptation and reliability-aware post-update smoothing after intermediate estimation. The proposed approach enables nodes with more reliable observations to exert stronger influence during both adaptation and estimate refinement. A unified network recursion is developed, and the mean as well as mean-square behavior of the resulting algorithm is analyzed using a Lyapunov-based framework. Explicit conditions for stability are established, and the analysis characterizes how gradient coupling and post-update smoothing jointly influence convergence dynamics, disagreement reduction among neighboring update directions, and steady-state estimation accuracy. Simulation results under heterogeneous sensing conditions demonstrate that the proposed strategy improves robustness relative to purely local adaptation and achieves competitive distributed estimation performance compared with conventional diffusion LMS methods. Furthermore, close agreement between theoretical predictions and simulated steady-state behavior validates the developed analytical model.

Keywords: distributed adaptive filtering, coupled gradient learning, heterogeneous networks, mean-square deviation (MSD), reliability-aware learning.

1 Introduction

Adaptive filtering plays a central role in signal processing applications including system identification, noise suppression, channel equalization, active control, and real-time estimation. Classical adaptive algorithms such as the least-mean-square (LMS) and normalized LMS (NLMS) have been extensively studied due to their low computational complexity, robustness, and well-established convergence properties [1-3]. Although these methods perform effectively in stationary and mildly nonstationary environments, their performance may degrade in distributed and multichannel systems characterized by heterogeneous sensing quality, unequal measurement reliability, and inconsistent local adaptation behavior across nodes.

Distributed adaptive filtering has attracted considerable attention in networked systems such as wireless sensor networks, Internet-of-Things (IoT) platforms, autonomous sensing infrastructures, and multi-agent systems [4-7]. Diffusion strategies, including adapt-then-combine (ATC) and combine-then-adapt (CTA) architectures, enable neighboring nodes to cooperate through local information exchange, leading to improved

convergence behavior and reduced steady-state estimation error compared with noncooperative adaptation. Extensions incorporating noisy communication links, privacy-preserving learning, communication delays, and asynchronous adaptation have further broadened the applicability of distributed learning approaches [8-11]. Despite these advances, most distributed adaptive filtering methods rely primarily on cooperation through parameter exchange using topology-based combination rules and do not explicitly exploit heterogeneous sensing reliability when regulating local learning dynamics.

Meanwhile, deep learning has emerged as a powerful paradigm for data-driven optimization and nonlinear modeling in signal processing, communications, and biomedical applications [12-14]. This progress has stimulated increasing interest in neural adaptive filtering methods, where neural networks are employed to learn update rules or adaptive gradients directly from data [15-18]. Compared with conventional stochastic gradient methods, neural adaptive filtering provides increased flexibility for handling nonlinear and highly nonstationary environments. Learning-based adaptive approaches, including deep unfolding and meta-learning

strategies, have also attracted significant attention [19–22]. Although these approaches improve nonlinear modeling capability, they generally do not address reliability-aware cooperation among distributed adaptive agents operating under heterogeneous sensing quality.

The integration of adaptive filtering and distributed learning remains challenging, particularly in heterogeneous environments where sensing quality differs substantially across nodes. In practical distributed systems, some nodes may observe cleaner measurements and provide more informative adaptation directions, whereas others may be severely corrupted by noise, interference, or artifacts. A common strategy is to allow neighboring nodes to exchange parameter estimates after local adaptation. Nevertheless, parameter-level cooperation does not directly regulate the update directions driving local learning dynamics, nor does it explicitly exploit heterogeneous sensing reliability when aggregating neighboring information. Consequently, discrepancies among local gradients may persist and propagate through the network, potentially degrading convergence behavior and increasing steady-state estimation error.

These limitations become particularly important in multichannel sensing applications such as electroencephalography (EEG), electrocardiography (ECG), wearable sensing systems, and distributed physiological monitoring networks, where correlated signals are often observed under channel-dependent noise and artifact conditions [23–25]. In such environments, sensing quality may vary significantly among channels, making uniform cooperation mechanisms insufficient for robust distributed adaptation. More broadly, similar challenges arise in wireless sensor networks, industrial monitoring systems, and heterogeneous IoT infrastructures.

Motivated by these observations, this paper proposes a reliability-aware coupled-gradient distributed adaptive filtering framework in which cooperation is introduced through two complementary mechanisms: reliability-aware gradient coupling before adaptation and reliability-aware post-update smoothing after intermediate estimation. The proposed approach enables nodes with more reliable observations to exert stronger influence during both adaptation and estimate refinement. Unlike conventional diffusion strategies that primarily exchange parameter estimates, the proposed method incorporates reliability information directly into the learning dynamics, allowing adaptation directions as well as intermediate estimates to be regulated according to local sensing quality. Consequently, unreliable updates generated by severely corrupted nodes can be attenuated before propagating through the network.

From a theoretical perspective, the proposed framework admits tractable analysis through a unified recursion involving reliability-aware gradient coupling and post-update smoothing operators. Explicit conditions for mean stability are derived, while mean-square analysis characterizes steady-state performance and reveals how cooperation parameters jointly influence convergence dynamics, disagreement reduction among neighboring update directions, and Estimation accuracy. The resulting analysis establishes a direct connection between sensing reliability, distributed cooperation, and adaptive learning behavior.

Although several distributed adaptive filtering frameworks have incorporated noisy-link modeling, asynchronous adaptation, or communication-aware cooperation mechanisms [19, 21, 22], these approaches primarily introduce cooperation through parameter exchange or communication modeling. In contrast, the proposed framework incorporates reliability-aware cooperation at two complementary stages: gradient construction before adaptation and estimate refinement after intermediate adaptation. Consequently, sensing reliability influences not only exchanged information but also the adaptation dynamics themselves. This distinction enables the proposed method to regulate unreliable update directions before their propagation through the network while further improving spatial consistency through reliability-aware post-update smoothing.

The main contributions of this work are summarized as follows:

- A reliability-aware distributed adaptive filtering framework introducing cooperation at both gradient and estimate levels under heterogeneous sensing conditions, allowing nodes with higher sensing reliability to exert stronger influence during distributed adaptation.
- A unified global recursion incorporating reliability-aware gradient coupling and estimate-level smoothing operators, providing a compact representation of the resulting distributed learning dynamics.
- Mean and mean-square performance analyses establishing stability conditions and characterizing the joint influence of cooperation parameters on convergence behavior, disagreement reduction among neighboring update directions, and steady-state estimation performance.
- Validation under heterogeneous sensing scenarios together with discussion of potential relevance to distributed biomedical sensing and other heterogeneous adaptive network applications.

Notation

Lowercase bold letters such as $\mathbf{w}$ denote vectors, while uppercase bold letters such as $\mathbf{A}$ represent matrices. Symbols $(\cdot)^T$, $(\cdot)^{-1}$, and $\text{Tr}(\cdot)$ denote matrix transpose, inverse, and trace operators, respectively. Notation $E[\cdot]$ represents expectation, and $\rho(\cdot)$ denotes the spectral radius of a matrix. The operators $\text{diag}\{\cdot\}$ and $\text{blockdiag}\{\cdot\}$ generate diagonal and block-diagonal matrices, respectively, whereas $\text{col}\{\cdot\}$ denotes column-wise stacking of vectors. The notation $\text{vec}(\cdot)$ represents vectorization obtained by stacking matrix columns into a single vector. The Kronecker product is denoted by $\otimes$, and $\|\cdot\|$ denotes the Euclidean norm.

For node k , the local parameter estimate at iteration i is denoted by $\mathbf{w}_k(i)$, while $\mathbf{w}^*$ represents the unknown parameter vector to be estimated. The corresponding weight-error vector is defined as

$$\tilde{\mathbf{w}}_k(i) = \mathbf{w}^* - \mathbf{w}_k(i).$$

The regression vector and measurement noise at node k are denoted by $\mathbf{u}_k(i)$ and $v_k(i)$, respectively, and $R_{\mathbf{u},k} = E[\mathbf{u}_k(i)\mathbf{u}_k^T(i)]$ denotes the covariance matrix of the local regressor. The neighborhood of node k is represented by N_k , and $c_{\ell k}$ denotes the reliability-aware combination weight assigned by node k to information received from neighboring node ℓ . The matrix $\mathbf{C} = [c_{\ell k}]$ represents the corresponding reliability-aware combination matrix. Parameters λ and ρ control the gradient coupling strength and post-update smoothing strength, respectively.

2 Problem formulation

Consider a connected network composed of N sensing nodes indexed $k=1, 2, \dots, N$. The nodes are connected through local neighborhoods, where N_k denotes the set of nodes whose information can be accessed by node k , including node k itself. The objective of the network is to estimate a common unknown parameter vector $\mathbf{w}^* \in \mathbb{R}^M$, where M denotes the filter length. Although all nodes pursue the same estimation objective, the quality of their observations may differ substantially due to heterogeneous noise levels, sensor placement, channel conditions, artifacts, or local disturbances. Consequently, some nodes may provide highly reliable adaptation information, whereas others may generate noisy or misleading update directions.

At time instant i , node k acquires a scalar measurement $d_k(i)$ and a regression vector $\mathbf{u}_k(i) \in \mathbb{R}^M$. The local observation model is

$$d_k(i) = \mathbf{u}_k^T(i)\mathbf{w}^i + v_k(i), \quad (1)$$

where $v_k(i)$ denotes additive measurement noise. The noise process is assumed to be zero mean with variance $\sigma_{v,k}^2$, and the variances are allowed to differ from node to node. This setting captures heterogeneous sensing conditions in which low-noise nodes provide more reliable measurements, while high-noise nodes may produce less informative local updates.

Each node maintains a local estimate $\mathbf{w}_k(i)$ of $\mathbf{w}^i$. Based on the estimate from the previous iteration, node k computes the a priori estimation error

$$e_k(i) = d_k(i) - \mathbf{u}_k^T(i)\mathbf{w}_k(i-1). \quad (2)$$

Substituting (1) into (2) gives

$$e_k(i) = \mathbf{u}_k^T(i)\tilde{\mathbf{w}}_k(i-1) + v_k(i), \quad (3)$$

Equation (3) shows that the local prediction error contains one component due to parameter mismatch and another component due to measurement noise. Hence, even if two neighboring nodes estimate the same underlying parameter vector, their instantaneous errors may have significantly different reliability levels. The local stochastic gradient used by LMS-type adaptation at node k is defined as

$$\mathbf{g}_k(i) = \mathbf{u}_k(i)e_k(i). \quad (4)$$

Using (3), this gradient can be decomposed as

$$\mathbf{g}_k(i) = \mathbf{u}_k(i)\mathbf{u}_k^T(i)\tilde{\mathbf{w}}_k(i-1) + \mathbf{u}_k(i)v_k(i). \quad (5)$$

The first term in (5) represents the useful adaptation component associated with the current estimation error, whereas the second term represents gradient noise caused by measurement disturbance. Since the variance of $v_k(i)$ may vary across nodes, the stochastic gradients generated across the network can have unequal reliability. In particular, high-noise nodes may produce update directions with larger random perturbations, which can slow convergence, increase steady-state MSD, and degrade neighboring estimates if their information is combined without accounting for reliability.

This behavior gives rise to gradient inconsistency, meaning that neighboring nodes may produce conflicting or unequally reliable adaptation directions even though they seek the same parameter vector. Conventional distributed adaptive filtering strategies, such as ATC diffusion LMS, mitigate local noise effects by

allowing nodes to exchange and combine inter-mediate parameter estimates. While such parameter-level cooperation is effective in many settings, it does not explicitly regulate the quality of the stochastic gradients that drive the local adaptation step. As a result, unreliable gradient information may already affect the intermediate estimates before cooperation is applied.

The problem addressed in this work is therefore to design a distributed adaptive filtering strategy that accounts for heterogeneous sensing reliability during the learning process. More specifically, the objective is to construct a cooperation mechanism that reduces the influence of unreliable gradients before adaptation and further refines the intermediate estimates through reliability-aware smoothing. This leads to a two-level cooperation structure: gradient-level cooperation for regulating local update directions and estimate-level

cooperation for improving spatial consistency after adaptation.

To support this objective, each node assigns reliability-aware weights to the information received from its neighbors. These weights are designed so that nodes with cleaner observations have stronger influence than severely corrupted nodes. The resulting formulation preserves the distributed nature of diffusion adaptation, since each node communicates only with its immediate neighbors, while explicitly incorporating sensing reliability into the update process.

The distributed estimation problem considered in this paper is illustrated in Fig. 1. The figure highlights the relationship between heterogeneous observations, unequal gradient reliability, and the need for reliability-aware cooperation during both gradient construction and estimate refinement.

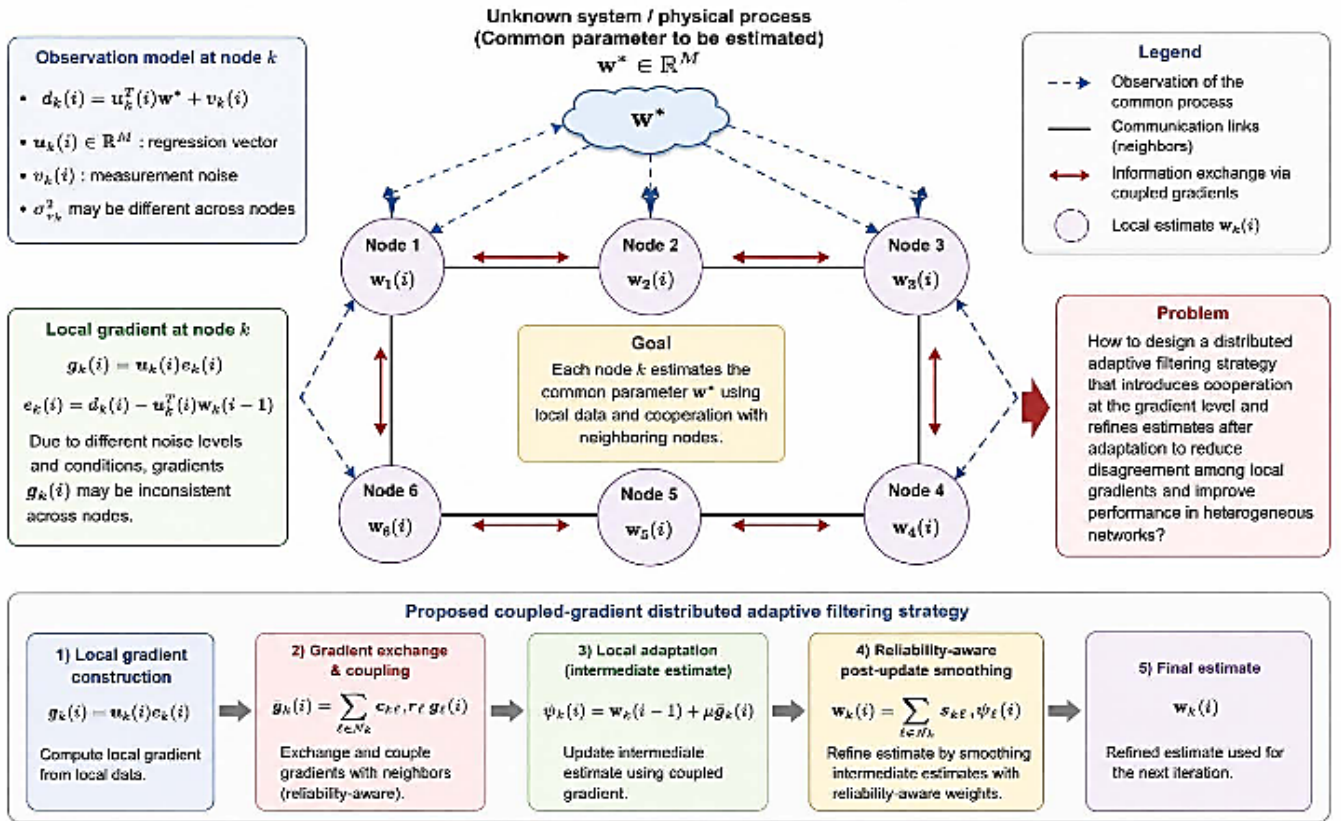


Fig. 1. Illustration of the distributed adaptive filtering problem over a heterogeneous sensing network

3 Proposed method

This section presents the proposed reliability-aware coupled-gradient distributed adaptive filtering algorithm for heterogeneous networks. The main objective is to reduce the adverse influence of unreliable nodes by allowing cooperation to occur not only through local adaptation but also through reliability-weighted gradient exchange and post-update smoothing. Unlike conventional ATC diffusion LMS, where nodes first adapt

locally and then combine intermediate parameter estimates using topology-based weights, the proposed method first constructs a reliability-aware coupled gradient and then applies a secondary reliability-aware smoothing step to the intermediate estimates. This two-stage cooperation mechanism is particularly useful in heterogeneous sensing networks, where different nodes may operate under substantially different noise conditions.

3.1 Local gradient computation

At iteration i , node k computes the local a priori estimation error as

$$e_k(i) = d_k(i) - \mathbf{u}_k^T(i) \mathbf{w}_k(i-1). \quad (6)$$

The corresponding instantaneous stochastic gradient is given by

$$\mathbf{g}_k(i) = \mathbf{u}_k(i) e_k(i). \quad (7)$$

Using the observation model (1), the local gradient can be decomposed as

$$\mathbf{g}_k(i) = \mathbf{u}_k(i) \mathbf{u}_k^T(i) \tilde{\mathbf{w}}_k(i-1) + \mathbf{u}_k(i) v_k(i). \quad (8)$$

Equation (8) shows that the local gradient contains a deterministic component associated with the current weight-error vector and a stochastic component induced by the measurement noise. Therefore, under heterogeneous sensing conditions, gradients generated by high-noise nodes may be substantially less reliable than those generated by low-noise nodes. This motivates the use of reliability-aware cooperation rather than purely topology-based averaging.

3.2 Reliability-aware combination weights

To account for heterogeneous noise conditions, each node assigns larger cooperation weights to more reliable neighbors. The weights satisfy

$$c_{\ell k} \geq 0, \quad c_{\ell k} = 0 \text{ if } \ell \notin N_k, \quad \sum_{\ell \in N_k} c_{\ell k} = 1. \quad (9)$$

A practical choice is to define the node reliability as

$$r_k = \frac{1}{\sigma_{v,k}^2 + \varepsilon},$$

where $\varepsilon > 0$ avoids division by zero, and then normalize the reliabilities over each neighborhood according to

$$c_{\ell k} = \frac{r_\ell}{\sum_{m \in N_k} r_m}, \quad \ell \in N_k. \quad (10)$$

This construction ensures that nodes with lower measurement noise variance have stronger influence during cooperation, while highly corrupted nodes contribute less to the update directions of their neighbors.

3.3 Reliability-aware coupled gradient

After computing the local gradients, neighboring nodes exchange their gradient vectors and node k constructs the reliability-aware coupled gradient

$$\mathbf{g}_k^c(i) = (1-\lambda) \mathbf{g}_k(i) + \lambda \sum_{\ell \in N_k} c_{\ell k} \mathbf{g}_\ell(i), \quad (11)$$

where $0 \leq \lambda \leq 1$ controls the strength of gradient-level cooperation. For $\lambda = 0$, the coupled gradient reduces to the purely local LMS gradient. As λ increases, node k incorporates more gradient information from its neighbors. Since the neighboring gradients are combined according to reliability-aware weights, gradients originating from cleaner nodes have stronger influence on the coupled update direction. Therefore, the coupled gradient in (11) suppresses unreliable local update directions while preserving distributed implementation.

The form in (11) can also be interpreted as a reliability-aware gradient smoothing operation. Unlike conventional diffusion LMS, which combines parameter estimates after local adaptation, the proposed method modifies the adaptation direction before the intermediate estimate is produced. This is important in heterogeneous networks because unreliable gradients can be corrected before they propagate into the local parameter update.

3.4 Intermediate adaptation step

Using the coupled gradient, node k first computes an intermediate estimate

$$\psi_k(i) = \mathbf{w}_k(i-1) + \mu \mathbf{g}_k^c(i), \quad (12)$$

where $\mu > 0$ denotes the adaptation step-size. Substituting (11) into (12) gives

$$\begin{aligned} \psi_k(i) = & \mathbf{w}_k(i-1) \\ & + \mu(1-\lambda) \mathbf{g}_k(i) + \mu\lambda \sum_{\ell \in N_k} c_{\ell k} \mathbf{g}_\ell(i) \end{aligned} \quad (13)$$

Equation (13) shows that the intermediate estimate is driven by a combination of the local gradient and reliability-weighted neighboring gradients. Hence, the adaptation direction is no longer determined solely by the local observation quality of node k .

3.5 Reliability-aware post-update smoothing

After the intermediate adaptation step, each node applies a reliability-aware smoothing operation to the intermediate estimates:

$$\mathbf{w}_k(i) = (1-\rho) \psi_k(i) + \rho \sum_{\ell \in N_k} c_{\ell k} \psi_\ell(i), \quad (14)$$

where $0 \leq \rho \leq 1$ controls the strength of post-update smoothing. The parameter ρ introduces a second cooperation mechanism at the estimate level. When $\rho = 0$, the method reduces to reliability-aware coupled-gradient

adaptation without post-update smoothing. For $\rho > 0$, the final estimate is obtained by combining intermediate estimates using the same reliability-aware neighborhood weights. This step further reduces the impact of severely corrupted nodes and improves the spatial consistency of the distributed estimates. The resulting update consists of two complementary operations. The first operation, controlled by λ , regularizes the stochastic gradients before adaptation. The second operation, controlled by ρ , smooths the intermediate estimates after adaptation. Together, these two operations provide a flexible mechanism for distributed learning in heterogeneous networks.

Remark 1. (Relation to ATC Diffusion LMS) The proposed method is related to ATC diffusion LMS but differs in two important aspects. In ATC diffusion LMS, each node first performs a local LMS adaptation and then combines intermediate estimates using topology-based combination weights. In contrast, the proposed method first performs reliability-aware gradient coupling, so the direction of adaptation itself is modified before the intermediate estimate is formed. In addition, the proposed method uses reliability-aware weights rather than purely uniform or topology-based weights, allowing more reliable nodes to exert greater influence. Thus, the proposed recursion can be viewed as a reliability-aware extension of diffusion adaptation in which cooperation occurs both at the gradient level and at the intermediate-estimate level. This structure is particularly suitable for heterogeneous networks where local data quality varies significantly across nodes.

3.6 Global network representation

Define the stacked vectors

$$\mathbf{w}(i) = \text{col}\{\mathbf{w}_1(i), \dots, \mathbf{w}_N(i)\}, \quad (15)$$

$$\boldsymbol{\psi}(i) = \text{col}\{\boldsymbol{\psi}_1(i), \dots, \boldsymbol{\psi}_N(i)\} \quad (16)$$

$$\mathbf{g}(i) = \text{col}\{\mathbf{g}_1(i), \dots, \mathbf{g}_N(i)\} \quad (17)$$

The coupled-gradient operation can be written as

$$\mathbf{g}^c(i) = H_g \mathbf{g}(i), \quad (18)$$

where

$$H_g = (1 - \lambda)I_{NM} + \lambda(C \otimes I_M). \quad (19)$$

The intermediate adaptation step becomes

$$\boldsymbol{\psi}(i) = \mathbf{w}(i-1) + \mu H_g \mathbf{g}(i). \quad (20)$$

Similarly, the post-update smoothing step is expressed as

$$\mathbf{w}(i) = H_p \boldsymbol{\psi}(i), \quad (21)$$

where

$$H_p = (1 - \rho)I_{NM} + \rho(C \otimes I_M). \quad (22)$$

Combining (20) and (21), the overall global recursion is

$$\mathbf{w}(i) = H_p (\mathbf{w}(i-1) + \mu H_g \mathbf{g}(i)). \quad (23)$$

Equation (23) is the compact network-level representation of the proposed reliability-aware coupled-gradient algorithm. This recursion will be used as the basis for the mean and mean-square performance analysis.

The complete proposed algorithm is summarized in Algorithm 1.

Algorithm 1.

Reliability-Aware Coupled-Gradient Diffusion MS

- 1: Initialize $\mathbf{w}_k(0)$ for all nodes $k = 1, \dots, N$
 - 2: Select $\mu > 0$, $0 \leq \lambda \leq 1$, and $0 \leq \rho \leq 1$
 - 3: Construct reliability-aware weights c_{ik} using
 - 4: **for** $i = 1, 2, \dots$ **do**
 - 5: **for each node** k **do**
 - 6: Acquire $d_k(i)$ and $\mathbf{u}_k(i)$
 - 7: Compute $e_k(i) = d_k(i) - \mathbf{u}_k^\top(i) \mathbf{w}_k(i-1)$
 - 8: Compute $\mathbf{g}_k(i) = \mathbf{u}_k(i) e_k(i)$
 - 9: Exchange $\mathbf{g}_k(i)$ with neighboring nodes $\ell \in N_k$
 - 10: Construct $\mathbf{g}_k^c(i)$ Using (11)
 - 11: Compute $\boldsymbol{\psi}_k(i) = \mathbf{w}_k(i-1) + \mu \mathbf{g}_k^c(i)$
 - 12: Exchange $\boldsymbol{\psi}_k(i)$ with neighboring nodes $\ell \in N_k$
 - 13: Update $\mathbf{w}_k(i) = (1 - \rho) \boldsymbol{\psi}_k(i) + \rho \sum_{\ell \in N_k} c_{\ell k} \boldsymbol{\psi}_\ell(i)$
 - 14: **end for**
 - 15: **end for**
-

4 Performance analysis

This section analyzes the mean and mean-square behavior of the proposed reliability-aware coupled-gradient diffusion LMS algorithm. The analysis is developed for the exact recursion introduced in Section III, where cooperation is performed in two stages: first through reliability-aware gradient coupling and then through reliability-aware post-update smoothing. This distinction is important because both stages affect the error dynamics, stability range, and steady-state MSD.

Let $\tilde{\mathbf{w}}(i) = \text{col}\{\tilde{\mathbf{w}}_1(i), \dots, \tilde{\mathbf{w}}_N(i)\}$ denote the global weight-error vector. We also define R_u as $R_u = \text{blockdiag}\{R_{u,1}, \dots, R_{u,N}\}$. The reliability-aware gradient-coupling and post-smoothing matrices are denoted by $H_g = (1-\lambda)I_{NM} + \lambda(C \otimes I_M)$ and $H_p = (1-\rho)I_{NM} + \rho(C \otimes I_M)$, respectively. Here, C is the reliability-aware combination matrix introduced in Section 3.

4.1 Assumptions

The following assumptions are used to obtain tractable performance expressions. The regression vectors $\mathbf{u}_k(i)$ are zero-mean, temporally white, and independent across nodes. Their covariance matrices $R_{u,k}$ are positive definite and may vary from node to node. The measurement noises $v_k(i)$ are zero-mean, independent of all regressors, and satisfy $E[v_k^2(i)] = \sigma_{v,k}^2$. The variances $\sigma_{v,k}^2$ are allowed to differ across the network, which models heterogeneous sensing quality. The combination matrix C is non-negative and row-stochastic, so that the reliability-aware smoothing operation preserves the desired solution $\mathbf{1}_N \otimes \mathbf{w}^\dagger$. Finally, the step-size μ is assumed sufficiently small for the standard LMS independence approximation to hold.

4.2 Global error recursion

The proposed algorithm first forms the coupled gradient $\mathbf{g}^c(i) = H_g \mathbf{g}(i)$, then computes the intermediate estimate $\boldsymbol{\psi}(i) = \mathbf{w}(i-1) + \mu H_g \mathbf{g}(i)$, and finally applies post-update smoothing according to $\mathbf{w}(i) = H_p \boldsymbol{\psi}(i)$. Since $H_p(\mathbf{1}_N \otimes \mathbf{w}^\dagger) = \mathbf{1}_N \otimes \mathbf{w}^\dagger$, the global weight-error recursion becomes

$$\tilde{\mathbf{w}}(i) = H_p \tilde{\mathbf{w}}(i-1) - \mu H_p H_g \mathbf{g}(i). \quad (24)$$

The local stochastic gradient can be decomposed as

$$\mathbf{g}_k(i) = \mathbf{u}_k(i) \mathbf{u}_k^T(i) \tilde{\mathbf{w}}_k(i-1) + \mathbf{u}_k(i) v_k(i). \quad (25)$$

By stacking all local gradients, we obtain $\mathbf{g}(i) = R_u(i) \tilde{\mathbf{w}}(i-1) + \mathbf{q}(i)$, where

$$R_u(i) = \text{blockdiag}\{\mathbf{u}_1(i) \mathbf{u}_1^T(i), \dots, \mathbf{u}_N(i) \mathbf{u}_N^T(i)\}$$

and

$$\mathbf{q}(i) = \text{col}\{\mathbf{u}_1(i) v_1(i), \dots, \mathbf{u}_N(i) v_N(i)\}$$

Substituting this expression into (24) gives the instantaneous global error recursion

$$\tilde{\mathbf{w}}(i) = (H_p - \mu H_p H_g R_u(i)) \tilde{\mathbf{w}}(i-1) - \mu H_p H_g \mathbf{q}(i). \quad (26)$$

Equation (26) shows that the proposed algorithm modifies the error dynamics through two distinct operators. The matrix H_g controls how local and neighboring gradients are blended before adaptation, while H_p controls how intermediate estimates are smoothed after adaptation. Therefore, the transient and steady-state behavior depend jointly on λ , ρ , the reliability matrix C , and the input covariance matrix R_u .

4.3 Mean stability

Taking expectations from both sides of (26) and using the independence assumptions yields

$$E[\tilde{\mathbf{w}}(i)] = F E[\tilde{\mathbf{w}}(i-1)], \quad (27)$$

where

$$F = H_p - \mu H_p H_g R_u \quad (28)$$

Matrix F is the mean transition matrix of the proposed algorithm. Hence, the algorithm is stable in the mean if

$$\rho(F) < 1. \quad (29)$$

This condition explicitly includes both cooperation stages. When $\rho=0$, we have $H_p = I_{NM}$ and the condition reduces to the stability condition of reliability-aware coupled-gradient LMS without post-update smoothing. When $\lambda=0$ and $\rho=0$, both cooperation mechanisms are removed and the standard noncooperative LMS mean recursion is recovered.

Remark 2: The role of λ is to regulate the amount of neighboring gradient information used before adaptation. Small values of λ preserve mostly local update directions, while larger values increase the influence of reliability-weighted neighboring gradients. The role of ρ is different: it controls the amount of smoothing applied after the intermediate estimate has been formed. Therefore, λ affects the direction of adaptation, whereas ρ affects the spatial consistency of the resulting estimates.

Remark 3: Unlike a purely Laplacian-based symmetric coupling matrix, the reliability-aware matrix C may be asymmetric because the weights depend on local neighborhoods and noise variances. Therefore, the most general mean-stability condition is the spectral-radius condition in (29). This condition is also the one used in the numerical evaluation of the theoretical MSD.

4.4 Effect on gradient disagreement

The main purpose of the gradient-coupling stage is to reduce inconsistency among neighboring stochastic gradients. To formalize this effect, define the consensus projection matrix $P_c = (1/N)\mathbf{1}_N\mathbf{1}_N^T \otimes I_M$ and the disagreement projection matrix $P_d = I_{NM} - P_c$. Vector $P_d\mathbf{g}(i)$ represents the component of the instantaneous gradient field that is not aligned with the network-wide consensus direction. The coupled gradient satisfies $\mathbf{g}^c(i) = H_g\mathbf{g}(i)$. Therefore, its disagreement component is $P_d\mathbf{g}^c(i) = P_dH_g\mathbf{g}(i)$. If the reliability-aware coupling operator satisfies $\gamma_g = \|P_dH_gP_d\|_2 < 1$, then the disagreement component is contracted according to

$$\|P_d\mathbf{g}^c(i)\|_2 \leq \gamma_g \|P_d\mathbf{g}(i)\|_2. \quad (30)$$

This condition explains why gradient coupling can improve robustness in heterogeneous networks. When high-noise nodes produce unreliable local gradients, reliability-aware averaging reduces their influence and moves the coupled gradient field toward a more consistent direction. The contraction factor depends on λ and on the structure of C . Moderate coupling can reduce gradient disagreement, but excessive coupling may over-smooth informative local variations and degrade adaptation.

Remark 4: The condition $\gamma_g < 1$ is a sufficient contraction condition. It is not required to claim that every possible reliability matrix produces strict Euclidean contraction. Instead, it clarifies the mechanism through which the proposed algorithm improves performance: the gradient-coupling stage is beneficial when the reliability-aware combination matrix reduces the disagreement component of the stochastic gradient field.

4.5 Mean-square performance

We next characterize the steady-state MSD. Define the covariance matrix

$$P_i = E[\tilde{\mathbf{w}}(i)\tilde{\mathbf{w}}^T(i)].$$

Under the small-step-size approximation, the covariance recursion associated with (26) is

$$P_i \approx FP_{i-1}F^T + \mu^2 H_p H_g C_q H_g^T H_p^T, \quad (31)$$

where

$$C_q = E[\mathbf{q}(i)\mathbf{q}^T(i)] = \text{blockdiag}\{\sigma_{v,1}^2 R_{u,1}, \dots, \sigma_{v,N}^2 R_{u,N}\}.$$

The first term in (31) describes how previous estimation errors propagate through the two cooperation stages and the adaptation step. The second term describes how

measurement noise is injected into the network through the reliability-aware coupled gradient and then shaped by the post-update smoothing operator. Therefore, both λ and ρ influence the steady-state MSD.

At steady state, $P_i = P_{i-1} = P_\infty$, and (31) becomes

$$P_\infty = FP_\infty F^T + \mu^2 H_p H_g C_q H_g^T H_p^T. \quad (32)$$

If $\rho(F) < 1$, then the discrete Lyapunov equation in (32) has a unique positive semidefinite solution. Equivalently, vectorization gives

$$\text{vec}(P_\infty) = \mu^2 (I_{N^2 M^2} - F \otimes F)^{-1} \times \text{vec}(H_p H_g C_q H_g^T H_p^T). \quad (33)$$

The steady-state network MSD is then computed as

$$\text{MSD}_\infty = \frac{1}{N} \text{Tr}(P_\infty). \quad (34)$$

Remark 5: Equation (32) is directly matched to the recursion used in the simulations. Matrix H_g appears because gradients are coupled before adaptation, and matrix H_p appears because intermediate estimates are smoothed after adaptation. Therefore, the theoretical MSD accounts for both mechanisms.

Remark 6: The coupling parameter λ and the smoothing parameter ρ should not be interpreted as parameters that always improve performance when increased. A small λ may leave unreliable gradients insufficiently corrected, whereas an excessively large λ may over-smooth local adaptation directions. Similarly, a small ρ may provide insufficient estimate-level smoothing, whereas an excessively large ρ may reduce responsiveness to local data. This explains why the best performance is typically obtained for moderate values of these parameters.

4.6 Interpretation

The analysis leads to three main conclusions. First, the proposed algorithm is governed in the mean by the transition matrix $F = H_p - \mu H_p H_g R_u$, which explicitly includes both reliability-aware gradient coupling and post-update smoothing. Second, the gradient-coupling stage can reduce disagreement among local update directions when the reliability-aware operator contracts the disagreement component of the gradient field. Third, the steady-state MSD is described by the Lyapunov equation in (32), where the measurement-noise contribution is shaped by both H_g and H_p . These results provide the theoretical basis for the simulation results presented in the next section.

5 Simulation results

This section evaluates the performance of the proposed coupled-gradient distributed adaptive filtering algorithm under heterogeneous sensing conditions. The simulations are designed to verify the theoretical findings developed in Section 4 and to investigate the influence of gradient-level cooperation on convergence behavior and steady-state performance. Comparisons are performed with noncooperative LMS and conventional adapt-then-combine (ATC) diffusion LMS strategies. Unless otherwise stated, all results are averaged over 100 independent Monte Carlo runs. The network performance is evaluated using the network mean-square deviation (MSD), defined as

$$\text{MSD}(i) = \frac{1}{N} \sum_{k=1}^N \mathbb{E} [\|w^i - w_k(i)\|^2].$$

The expectation is approximated through Monte Carlo averaging.

A connected network consisting of $N=10$ nodes is considered, where each node estimates a common unknown parameter vector

$$w^i = [0.7, -0.4, 0.3, -0.2, 0.1, -0.1, 0.25, -0.05]^T$$

of length $M=8$. The network topology follows a ring structure in which each node exchanges information with its immediate neighbors. Uniform combination weights are employed for the ATC diffusion LMS baseline, whereas the proposed method utilizes reliability-aware combination weights inversely proportional to local measurement noise variance. Consequently, nodes with more reliable observations exert stronger influence during both gradient coupling and post-update smoothing.

At iteration i , node k observes $d_k(i)$ according to (1) where the regressors are generated according to

$$u_k(i) = \alpha_k n_k(i), \quad n_k(i) \sim \mathcal{N}(0, I_M),$$

with $\alpha_k = 0.35$ for all nodes. The measurement noise variances are selected as

$$\sigma_v = [0.01, 0.012, 0.015, 0.02, 0.03, 0.12, 0.16, 0.20, 0.25, 0.30],$$

thereby creating heterogeneous sensing conditions in which some nodes experience substantially stronger disturbances than others. The adaptation step-size is fixed at $\mu=0.02$, while the gradient-coupling parameter λ and post-update smoothing parameter ρ vary depending on the experiment. For the proposed algorithm, reliability-aware cooperation is introduced using weights inversely proportional to local measurement noise variance, such that nodes with lower measurement noise exert stronger influence during both gradient exchange and estimate refinement. Specifically, the

reliability of node k is defined as $r_k = \frac{1}{\sigma_{v,k}^2 + \varepsilon}$, where $\varepsilon > 0$ is a small regularization constant introduced to avoid numerical instability. The resulting neighborhood weights are normalized to satisfy the stochasticity condition

$$\sum_{\ell \in \mathcal{N}_k} c_{\ell k} = 1,$$

ensuring distributed implementation while assigning stronger influence to more reliable sensing nodes.

Figure 2 compares the network MSD curves obtained using noncooperative LMS, ATC diffusion LMS, and the proposed coupled-gradient strategy. The noncooperative approach exhibits the largest steady-state MSD due to the absence of cooperation among nodes. ATC diffusion improves performance by combining neighboring parameter estimates, resulting in faster convergence and reduced estimation error. The proposed method introduces cooperation directly through gradient exchange while incorporating reliability-aware weighting. Consequently, adaptation becomes less sensitive to severely corrupted nodes because gradients originating from more reliable observations receive larger weights during cooperation. The results indicate that the proposed strategy consistently improves performance relative to non-cooperative adaptation and remains competitive with conventional diffusion learning under heterogeneous conditions. The performance gain becomes increasingly pronounced as the degree of heterogeneity increases.

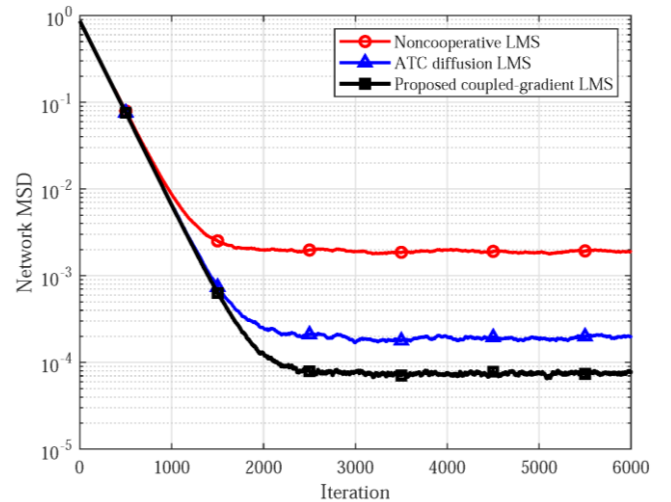


Fig. 2. Network MSD comparison among noncooperative LMS, ATC diffusion LMS, and the proposed coupled-gradient distributed adaptive filtering algorithm under heterogeneous sensing conditions

The influence of the coupling parameter is investigated for $\lambda \in \{0, 0.1, 0.2, 0.3, 0.5, 0.7\}$, where $\lambda = 0$ corresponds to purely local adaptation without gradient cooperation. Figure 3 shows that increasing λ initially reduces network MSD by suppressing disagreement among neighboring update directions. Moderate coupling therefore improves both convergence behavior and steady-state performance. However, excessively large coupling may over-constrain local adaptation and reduce responsiveness to node-specific observations. Consequently, the best performance is generally obtained for intermediate coupling strengths. This observation agrees with the theoretical analysis, which predicts a trade-off between cooperation intensity and adaptation flexibility.

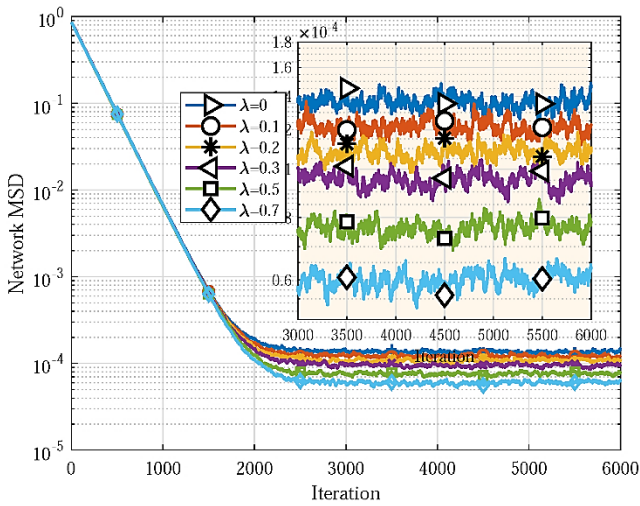


Fig. 3. Effect of the coupling parameter λ on network MSD performance

To validate the analytical model, the theoretical steady-state MSD obtained from the discrete Lyapunov equation is compared with simulated learning curves. The corresponding network MSD is computed as

$$\text{MSD}_\infty = \frac{1}{N} \text{Tr}(P_\infty).$$

Figure 4 demonstrates close agreement between theoretical prediction and simulated steady-state performance. Minor discrepancies in transient regions are expected due to finite Monte Carlo averaging and the independence assumptions commonly employed in LMS analysis. The results indicate that the derived theoretical model captures the dominant long-term behavior of the proposed algorithm.

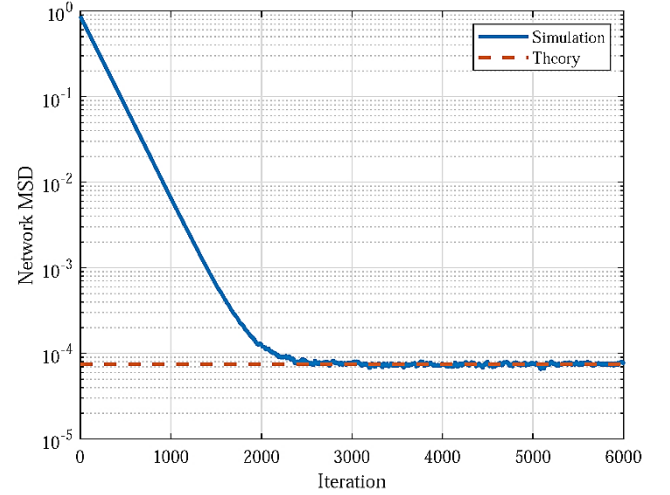


Fig. 4. Comparison between simulated network MSD and theoretical steady-state MSD predicted by the mean-square analysis

Figure 5 illustrates the steady-state MSD as a function of both adaptation step-size μ and coupling strength λ . The heatmap reveals a stable operating region where moderate values of μ and λ minimize steady-state error. Very small step-sizes slow adaptation, whereas excessively large step-sizes degrade stability. Similarly, weak coupling provides insufficient cooperation while overly strong coupling suppresses local adaptation. These results highlight the existence of an optimal trade-off between adaptation speed and cooperation intensity.

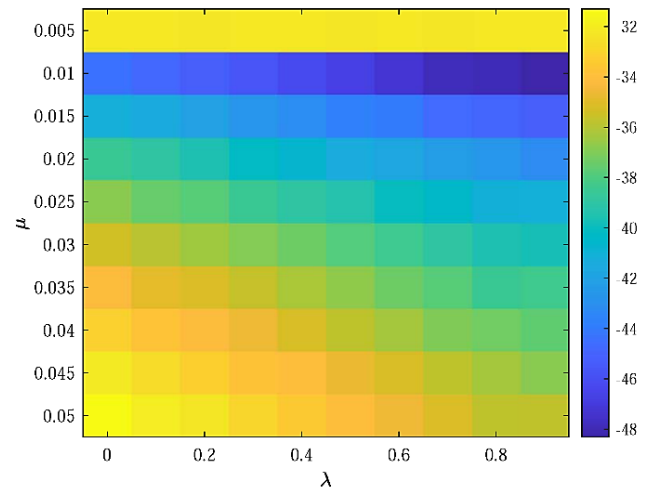


Fig. 5. Steady-state network MSD (dB) as a function of adaptation step-size μ and coupling strength λ . The heatmap reveals stable operating regions and highlights the trade-off between adaptation speed and cooperation intensity.

The simulation results support three main observations. First, gradient-level cooperation improves robustness relative to purely local adaptation. Second, reliability-aware weighting reduces the influence of severely corrupted nodes and enhances distributed estimation accuracy under heterogeneous sensing conditions. Third, the theoretical expressions accurately predict steady-state network performance. Overall, the results suggest that combining coupled-gradient adaptation with reliability-aware cooperation provides an effective framework for distributed adaptive filtering over heterogeneous networks.

6 Conclusion

This paper proposed a reliability-aware coupled-gradient distributed adaptive filtering framework for heterogeneous networks. Unlike conventional diffusion strategies that perform cooperation primarily through parameter exchange, the proposed approach introduces cooperation at two complementary levels: reliability-aware gradient coupling before adaptation and reliability-aware post-update smoothing after intermediate estimation. This structure enables neighboring nodes with more reliable observations to exert stronger influence during both adaptation and estimate refinement. A unified network recursion was developed, and the mean as well as mean-square behavior of the proposed algorithm was analyzed through a Lyapunov-based framework. The analysis established explicit conditions for stability and revealed how the gradient-coupling parameter and post-smoothing parameter jointly affect convergence behavior, disagreement reduction among neighboring update directions, and steady-state estimation accuracy. Simulation results under heterogeneous sensing conditions demonstrated that the proposed strategy improves robustness relative to purely local adaptation and achieves competitive distributed estimation performance in comparison with conventional diffusion LMS methods. The results further indicated that moderate cooperation strengths generally provide the most favorable balance between adaptation flexibility and distributed consistency, whereas excessively strong cooperation may oversmooth local information and degrade performance. In addition, close agreement between theoretical predictions and simulated steady-state behavior validated the developed analytical model and supported the accuracy of the derived mean-square performance expressions.

References

- [1] G. Sahaana, "Design of Adaptive FIR filter for Biomedical Signal Processing Applications," *2022 International Conference on Computer, Power and Communications (ICCCPC)*, Chennai, India, 2022, pp. 205-209.
- [2] O. Rosanne, I. Albuquerque, R. Cassani, J.-F. Gagnon, S. Tremblay and T. H. Falk, "Adaptive Filtering for Improved EEG-Based Mental Workload Assessment of Ambulant Users," *Front. Neurosci.* 2021.
- [3] A. H. Sayed, *Fundamentals of Adaptive Filtering*, Wiley, 2003.
- [4] A. H. Sayed, "Adaptation, learning, and optimization over networks," *Foundations and Trends in Machine Learning*, vol. 7, issue 4-5, pp. 311-801, NOW Publishers, Boston-Delft, 2014.
- [5] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3122-3136, July 2008.
- [6] A. Khalili, V. Vahidpour, A. Rastegarnia, W. M. Bazzi, S. Sanei, "Energy-efficient diffusion Kalman filtering for multiagent networks in IoT," *IEEE Internet of Things Journal*, 9 (8), 6277-6287.
- [7] M. Latifi, A. Khalili, A. Rastegarnia, W.M. Bazzi, S. Sanei, "A robust scalable demand-side management based on diffusion-admm strategy for smart grid," *IEEE Internet of Things Journal* 7 (4), 3363-3377.
- [8] D. Yu and L. Deng, "Deep Learning and Its Applications to Signal and Information Processing [Exploratory DSP]," in *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 145-154, Jan. 2011.
- [9] Ch. Sharma, "Deep Learning Techniques for Biomedical Imaging," in *Biomedical Imaging Technology, Signal Processing Strategies and Innovations, IEEE*, 2026, pp.25-44.
- [10] N. Tanwar and Y. Hasiija, "Deep Learning: A tool in Biomedical Science," *2022 International Conference for Advancement in Technology (ICONAT)*, Goa, India, 2022, pp. 1-6.
- [11] Z. Esmailbeig and M. Soltanian, "Deep Learning Meets Adaptive Filtering: A Stein's Unbiased Risk Estimator Approach," *2023 59th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Monticello, IL, USA, 2023, pp. 1-6.
- [12] M. J. C. S. Reis, "Deep-Learning-Driven Adaptive Filtering for Non-Stationary Signals: Theory and Simulation," *Electronics* 2026, 15, 381. <https://doi.org/10.3390/electronics15020381>.
- [13] C. Wang, S. Zhao, T. Huang, Y. Zhou, "A deep learning-enhanced adaptive filtering framework for intelligent CNS/INS navigation system," *Measurement*, vol. 270, page 120884, 2026.
- [14] B. Wu, F. Zhou, "Application of neural network adaptive filter method to simultaneous detection of polymetallic ions based on ultraviolet-visible spectroscopy," *Front Chem.* 2024 Sep 5;12:1409527.
- [15] J. R. Hershey, J. L. Roux, F. Weninger, "Deep unfolding: Model-based inspiration of novel deep architectures," *arXiv preprint arXiv:1409.2574*.
- [16] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," In *Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML'10)*. Omnipress, Madison, WI, USA, 399-406, 2010.
- [17] M. Andrychowicz et al., "Learning to learn by gradient descent by gradient descent," *NeurIPS*, 3981-3989, 2016.

- [18] Z. Sun, and Y. Yu, "Fast Approximation for Sparse Coding with Applications to Object Recognition," *Sensors* 2021, 21, 1442. <https://doi.org/10.3390/s21041442>.
- [19] A. Khalili, M. A. Tinati, A. Rastegarnia and J. A. Chambers, "Steady-State Analysis of Diffusion LMS Adaptive Networks with Noisy Links," in *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 974-979, Feb. 2012.
- [20] H. Han, Z. Zhang, H. Chen, and A. H. Sayed, "Masked diffusion strategy for privacy-preserving distributed learning," *IEEE Trans. Information Forensics and Security*, vol. 20, pp. 8402-8416, 2025.
- [21] F. Hua, R. Nassif, C. Richard, H. Wang, and A. H. Sayed, "Diffusion LMS with communication delays: Stability and performance analysis," *IEEE Signal Processing Letters*, vol. 27, pp. 730-734, 2020.
- [22] R. Nassif, C. Richard, A. Ferrari, and A. H. Sayed, "Multitask diffusion adaptation over asynchronous networks," *IEEE Trans. Signal Processing*, vol. 64, no. 11, pp. 2835-2850, June 2016.
- [23] M. K. Islam, A. Rastegarnia, Z. Yang, "Methods for artifact detection and removal from scalp EEG: A review," *Neurophysiologie Clinique/Clinical Neurophysiology* 46 (4-5), 287-305.
- [24] M. K. Islam, A. Rastegarnia, Z. Yang, "A Wavelet-Based Artifact Reduction from Scalp EEG for Epileptic Seizure Detection," *IEEE Journal of Biomedical and Health Informatics*, 20 (5), 1321 – 1332.
- [25] A. Bashashati, M. Fatourehchi, R. K. Ward, and G. E. Birch, "A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals," *J Neural Eng.* 2007 Jun; vol. 4, no.2, pp.32-57, 2007.

Received 18 April 2026
