

Analysis of incremental RLS adaptive networks with noisy links

Azam Khalili, Mohammad Ali Tinati, and Amir Rastegarnia^{a)}

Faculty of Electrical and Computer Engineering, University of Tabriz

Tabriz 51664, Iran

a) a_rastegar@ieee.org

Abstract: In this paper, we study the effect of noisy links on the steady-state performance of incremental recursive least-squares (RLS) adaptive networks. In our analysis, using weighted spatial-temporal energy conservation approach, we arrive a variance relation which contains moments that represent the effects of noisy links. We evaluate these moments and derive closed-form expressions for the mean-square deviation (MSD) and excess mean-square error (EMSE) to explain the steady-state performance at each individual node. The derived expressions have good match with simulations.

Keywords: incremental networks, distributed estimation, noisy links

Classification: Science and engineering for electronics

References

- [1] C. G. Lopes and A. H. Sayed, “Incremental adaptive strategies over distributed networks,” *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4064–4077, Aug. 2007.
- [2] A. H. Sayed and C. G. Lopes, “Adaptive processing over distributed networks,” *IEICE Trans. Fundamentals*, vol. E90–A, no. 8, pp. 1504–1510, Aug. 2007.
- [3] L. Li, J. A. Chambers, C. G. Lopes, and A. H. Sayed, “Distributed estimation over an adaptive incremental network based on the affine projection algorithm,” *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 151–164, Jan. 2010.
- [4] C. G. Lopes and A. H. Sayed, “Diffusion least-mean squares over adaptive networks: Formulation and performance analysis,” *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3122–3136, July 2008.
- [5] N. Takahashi, I. Yamada, and A. H. Sayed, “Diffusion least-mean squares with adaptive combiners: Formulation and performance analysis,” *IEEE Trans. Signal Process.*, vol. 8, no. 9, pp. 4795–4810, Sept. 2010.
- [6] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, “Diffusion recursive least-squares for distributed estimation over adaptive networks,” *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1865–1877, May 2008.
- [7] A. H. Sayed, *Fundamentals of adaptive filtering*, Wiley, NJ, 2003.

1 Introduction

The problem of estimating an unknown parameter from data collected by a set of nodes arises in several applications. Among the proposed solutions in the literature, adaptive networks are appealing solutions when the statistical information about the underlying processes of interest is not available. Co-operative processing in conjunction with adaptive filtering per node allows the network to account for time variations in the signal statistics. Adaptive networks may be referred to as incremental networks or diffusion networks, depending on the manner by which the nodes communicate with each other. The incremental LMS [1], incremental RLS [2], incremental techniques based on the affine projection algorithm [3], are examples of adaptive networks with incremental mode of cooperation. When more communication and energy resources are available, a diffusion cooperative scheme can be applied. In these schemes each node updates its estimate using all available estimates from its neighbors, as well as data and its own past estimate [4, 5, 6].

In [1, 2, 3, 4, 5, 6], the communication links between nodes are assumed to be ideal. In this paper we study the steady-state performance of incremental RLS adaptive network with unreliable communications modeled by noisy links. We first show that the performance of incremental RLS adaptive network drastically deteriorates when links between nodes are noisy. Then, we use the weighted energy conservation argument and derive closed-form expressions to explain the steady-state performance at each individual node. Simulation results are also presented to clarify the derived expressions.

Notation: We adopt boldface letters for random quantities. Symbol $*$ denotes complex conjugation (scalars) and Hermitian transpose (matrices).

2 Incremental RLS with noisy links

Consider a distributed network which is deployed to estimate an unknown vector w^o from measurements collected at nodes. Each node k has access to time-realizations $\{d_k(i), u_{k,i}\}$ of zero-mean spatial data $\{d_k, \mathbf{u}_k\}$, where each $d_k(i)$ is a scalar measurement and each $u_{k,i}$ is a $1 \times M$ row regression vector. At each time instant i , the network has access to the following data

$$\mathbf{y}_i = [d_1(i) \ d_2(i) \ \cdots \ d_N(i)]^T \quad \text{and} \quad \mathbf{H}_i = [u_{1,i} \ u_{2,i} \ \cdots \ u_{N,i}]^T \quad (1)$$

Note that \mathbf{y}_i and \mathbf{H}_i are in fact snapshot matrices revealing the network data status at time i . Collecting all the data (available up to time i) into global matrices \mathcal{Y}_i and \mathcal{H}_i yields

$$\mathcal{Y}_i = [y_0 \ y_1 \ \cdots \ y_i]^T \quad \text{and} \quad \mathcal{H}_i = [H_0 \ H_1 \ \cdots \ H_i]^T \quad (2)$$

Now we pose regularized weighted least-squares (LS) problem to estimate w^o as follows

$$\min_w \left[\lambda^{i+1} w^* \Pi w + \|\mathcal{Y}_i - \mathcal{H}_i w\|_{\mathcal{W}_i}^2 \right] \quad (3)$$

where Π is a regularization matrix, and the weighting matrix is given by [2]

$$\mathcal{W}_i = \text{diag}\{\lambda^i D, \lambda^{i-1} D, \dots, \lambda D, D\} \quad (4)$$

with a spatial weighting factor $D = \text{diag}\{\gamma_1, \gamma_2, \dots, \gamma_N\}$, $\gamma_i \geq 0$ and (time) forgetting factor $0 \ll \lambda \leq 1$. The solution of problem (3) is given by [2]

$$w_i = P_i \mathcal{H}_i^* \mathcal{W}_i \mathcal{Y}_i \quad (5)$$

where

$$P_i = (\lambda^{i+1} \Pi + \mathcal{H}_i^* \mathcal{W}_i \mathcal{H}_i)^{-1} \quad (6)$$

In [2] the incremental RLS adaptive network is introduced to estimate w^o recursively in a distributed manner which can be summarized as follows

$$\begin{cases} \psi_0^{(i)} \leftarrow w_{i-1}; & P_{0,i} \leftarrow \lambda^{-1} P_{i-1} \\ \psi_k^{(i)} = \psi_{k-1}^{(i)} + \Gamma u_{k,i}^* (d_k(i) - u_{k,i} \psi_{k-1}^{(i)}) \\ P_{k,i} = \lambda^{-1} [P_{k,i-1} - \Gamma u_{k,i}^* u_{k,i} P_{k,i-1}] \\ w_i \leftarrow \psi_N^{(i)}; & P_i \leftarrow P_{N,i} \end{cases} \quad (7)$$

where $\psi_k^{(i)}$ is the local estimate of w^o at node k at time i and Γ is

$$\Gamma = \frac{\lambda^{-1} P_{k,i-1}}{\gamma_k^{-1} + \lambda^{-1} u_{k,i} P_{k,i-1} u_{k,i}^*} \quad (8)$$

The given algorithm in (7) works as follows [2]: at each time i , the local estimate $\psi_k^{(i)}$ at node k is the LS solution considering data blocks \mathcal{Y}_{i-1} and \mathcal{H}_{i-1} in addition to the data collected along the path. At the end of the cycle, $\psi_N^{(i)}$ will contain precisely the desired solution w_i . Note that in this implementation $\psi_k^{(i)}$ is transmitted to the next node in the path, while $P_{k,i}$ is estimated locally and independent from the neighbor nodes. In the presence of noisy links, $\psi_k^{(i)}$ in (7) is updates as

$$\psi_k^{(i)} = \psi_{k-1}^{(i)} + q_{k,i} + \Gamma u_{k,i}^* e_k(i) - \Gamma u_{k,i}^* u_{k,i} q_{k,i} \quad (9)$$

where $e_k(i) = d_k(i) - u_{k,i} \psi_{k-1}^{(i)}$ and the $M \times 1$ vector $q_{k,i}$ represents the time realization of link noise between node $k-1$ and k which is assumed to be additive, zero-mean with covariance matrix $Q_k = E(\mathbf{q}_k \mathbf{q}_k^*)$. No distributional assumptions are required on the noise sequence. The effect of noisy links on the performance of incremental RLS adaptive network is obvious from Fig. 1. (The simulation setup is described in section (4)). As it is clear from Fig. 1, the performance of distributed adaptive estimation algorithm drastically decreases when links are noisy.

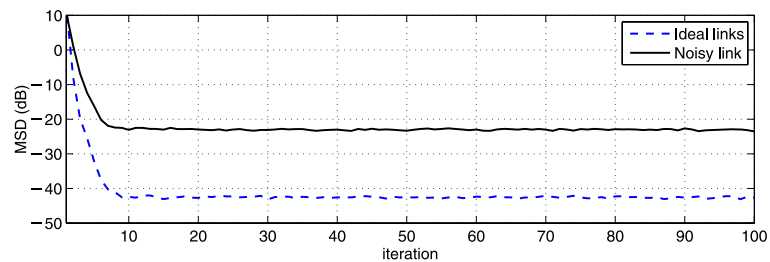


Fig. 1. The MSD learning curve for incremental RLS.

3 Performance analysis

Our analysis relies on the energy conservation approach of [7]. To carry out the performance analysis, we first need to assume a model for the data as is commonly done in the literature of adaptive algorithms. Thus, in the subsequent analysis we consider the following assumptions

A.1. Linear model $\mathbf{d}_k(i) = \mathbf{u}_{k,i}w^o + \mathbf{v}_k(i)$ where $\mathbf{v}_k(i)$ is white noise term with variance $\sigma_{v,k}^2$ and is independent of $\{\mathbf{d}_l(j), \mathbf{u}_{l,j}\}$ for all l, j .

A.2. $\mathbf{u}_{k,i}$ are spatially and temporary independent and $\{\mathbf{u}_k\}$ arise from a source with circular Gaussian distribution with covariance matrix $R_{u,k}$.

A.3. $\mathbf{q}_{k,i}$ is independent of $\{\mathbf{u}_{l,j}, \mathbf{v}_l(j), \mathbf{q}_{l,j}\}$ for all l, j .

Note that in steady-state analysis, it is desired to evaluate the MSD and EMSE for every node k which are defined as

$$\eta_k \triangleq E(\|\tilde{\boldsymbol{\psi}}_{k-1}^{(\infty)}\|^2) = E(\|\tilde{\boldsymbol{\psi}}_{k-1}^{(\infty)}\|_I^2) \text{ (MSD)} \quad (10)$$

$$\zeta_k \triangleq E(|\mathbf{e}_{a,k}(\infty)|^2) = E(\|\tilde{\boldsymbol{\psi}}_{k-1}^{(\infty)}\|_{R_{u,k}}^2) \text{ (EMSE)} \quad (11)$$

where $\mathbf{e}_{a,k}(i) \triangleq \mathbf{u}_{k,i}\tilde{\boldsymbol{\psi}}_{k-1}^{(i)}$ and $\tilde{\boldsymbol{\psi}}_k^{(i)} \triangleq w^o - \boldsymbol{\psi}_k^{(i)}$. By subtracting w^o from both sides of update equation (9) we get

$$\tilde{\boldsymbol{\psi}}_k^{(i)} = \tilde{\boldsymbol{\psi}}_{k-1}^{(i)} - \mathbf{q}_{k,i} - \Gamma \mathbf{u}_{k,i}^* \mathbf{e}_k(i) + \Gamma \mathbf{u}_{k,i}^* \mathbf{u}_{k,i} \mathbf{q}_{k,i} \quad (12)$$

Equating the weighted norm of both sides of the previous equation, take expectation of both sides and using assumptions (A.1)-(A.3) we obtain

$$E(\|\tilde{\boldsymbol{\psi}}_k\|_{\Sigma}^2) = E(\|\tilde{\boldsymbol{\psi}}_{k-1}\|_{\Sigma'}^2) + \sigma_{v,k}^2 E(\|\mathbf{u}_k\|_{\Gamma\Sigma\Gamma}^2) + E(\|\mathbf{u}_k\|_{\Gamma\Sigma\Gamma}^2 \mathbf{q}_k^* \mathbf{u}_k^* \mathbf{u}_k \mathbf{q}_k) \\ + E(\|\mathbf{q}_k\|_{\Sigma}^2) - E(\mathbf{q}_k^* \Sigma \Gamma \mathbf{u}_k^* \mathbf{u}_k \mathbf{q}_k) - E(\mathbf{q}_k^* \mathbf{u}_k^* \mathbf{u}_k \Gamma \Sigma \mathbf{q}_k) \quad (13)$$

where

$$\Sigma' \triangleq \Sigma - E(\Sigma \Gamma \mathbf{u}_k^* \mathbf{u}_k + \mathbf{u}_k^* \mathbf{u}_k \Gamma \Sigma) + E(\|\mathbf{u}_k\|_{\Gamma\Sigma\Gamma}^2 \mathbf{u}_k^* \mathbf{u}_k) \quad (14)$$

Recursion (13) is a variance relation that can be used to infer the steady-state performance at every node k . Note that Σ' is solely regressors-dependent and, therefore, decoupled from the weight error vector. Now using the eigen-decomposition $R_{u,k} = U_k \Lambda_k U_k^*$, (where Λ_k is a diagonal matrix with the eigenvalues of $R_{u,k}$ and U_k is unitary) we define the transformed quantities

$$\bar{\boldsymbol{\psi}}_k \triangleq U_k^* \tilde{\boldsymbol{\psi}}_k, \quad \bar{\boldsymbol{\psi}}_{k-1} \triangleq U_k^* \tilde{\boldsymbol{\psi}}_{k-1}, \quad \bar{\mathbf{u}}_k \triangleq \mathbf{u}_k U_k, \quad \bar{\mathbf{q}}_k \triangleq U_k^* \mathbf{q}_k \\ \bar{\Sigma} \triangleq U_k^* \Sigma U_k, \quad \bar{\Sigma}' \triangleq U_k^* \Sigma' U_k, \quad \bar{\Gamma} \triangleq U_k^* \Gamma U_k \quad (15)$$

Using the definitions in (15), Eqs.(13) and (14) can be rewritten as

$$E(\|\bar{\boldsymbol{\psi}}_k\|_{\bar{\Sigma}}^2) = E(\|\bar{\boldsymbol{\psi}}_{k-1}\|_{\bar{\Sigma}'}^2) + \sigma_{v,k}^2 E(\|\bar{\mathbf{u}}_k\|_{\bar{\Gamma}\bar{\Sigma}\bar{\Gamma}}^2) + E(\|\bar{\mathbf{u}}_k\|_{\bar{\Gamma}\bar{\Sigma}\bar{\Gamma}}^2 \bar{\mathbf{q}}_k^* \bar{\mathbf{u}}_k^* \bar{\mathbf{u}}_k \bar{\mathbf{q}}_k) \\ + E(\|\bar{\mathbf{q}}_k\|_{\bar{\Sigma}}^2) - E(\bar{\mathbf{q}}_k^* \bar{\Sigma} \bar{\Gamma} \bar{\mathbf{u}}_k^* \bar{\mathbf{u}}_k \bar{\mathbf{q}}_k) - E(\bar{\mathbf{q}}_k^* \bar{\mathbf{u}}_k^* \bar{\mathbf{u}}_k \bar{\Gamma} \bar{\Sigma} \bar{\mathbf{q}}_k) \quad (16)$$

$$\bar{\Sigma}' \triangleq \bar{\Sigma} - E(\bar{\Sigma} \bar{\Gamma} \bar{\mathbf{u}}_k^* \bar{\mathbf{u}}_k + \bar{\mathbf{u}}_k^* \bar{\mathbf{u}}_k \bar{\Gamma} \bar{\Sigma}) + E(\|\bar{\mathbf{u}}_k\|_{\bar{\Gamma}\bar{\Sigma}\bar{\Gamma}}^2 \bar{\mathbf{u}}_k^* \bar{\mathbf{u}}_k) \quad (17)$$

To proceed, we evaluate the moments in (16) and (17) which are

$$\begin{aligned} E(\|\bar{\mathbf{q}}_k\|_{\bar{\Sigma}}^2) &= \text{Tr} [B_k \bar{\Sigma}] \\ E(\|\bar{\mathbf{u}}_k\|_{\bar{\Gamma} \bar{\Sigma} \bar{\Gamma}}^2) &= \text{Tr} [\Lambda_k \bar{\Gamma} \bar{\Sigma} \bar{\Gamma}] \\ E(\|\bar{\mathbf{u}}_k\|_{\bar{\Gamma} \bar{\Sigma} \bar{\Gamma}}^2 \bar{\mathbf{q}}_k^* \bar{\mathbf{u}}_k^* \bar{\mathbf{u}}_k \bar{\mathbf{q}}_k) &= \text{Tr}[B_k (\Lambda_k \text{Tr}[\bar{\Gamma} \bar{\Sigma} \bar{\Gamma} \Lambda_k] + \delta \Lambda_k \bar{\Gamma} \bar{\Sigma} \bar{\Gamma} \Lambda_k)] \\ E(\bar{\mathbf{q}}_k^* \bar{\Sigma} \bar{\Gamma} \bar{\mathbf{u}}_k^* \bar{\mathbf{u}}_k \bar{\mathbf{q}}_k) &= \text{Tr}[\Lambda_k B_k \bar{\Sigma} \bar{\Gamma}] \\ E(\|\bar{\mathbf{u}}_k\|_{\bar{\Gamma} \bar{\Sigma} \bar{\Gamma}}^2 \bar{\mathbf{u}}_k^* \bar{\mathbf{u}}_k) &= (\Lambda_k \text{Tr}[\bar{\Gamma} \bar{\Sigma} \bar{\Gamma} \Lambda_k] + \delta \Lambda_k \bar{\Gamma} \bar{\Sigma} \bar{\Gamma} \Lambda_k) \end{aligned} \quad (18)$$

where $B_k = U_k^* Q_k U_k$ and $\delta = 1$ for circular complex data and $\delta = 2$ for real data. Replacing these moments, (16) and (17) can be rewritten as

$$\begin{aligned} E(\|\bar{\boldsymbol{\psi}}_k\|_{\bar{\Sigma}}^2) &= E(\|\bar{\boldsymbol{\psi}}_{k-1}\|_{\bar{\Sigma}'}^2) + \sigma_{v,k}^2 \text{Tr}[\Lambda_k \bar{\Gamma} \bar{\Sigma} \bar{\Gamma}] + \text{Tr}[B_k \bar{\Sigma}] \\ &\quad + \text{Tr}[B_k (\Lambda_k \text{Tr}[\bar{\Gamma} \bar{\Sigma} \bar{\Gamma} \Lambda_k] + \delta \Lambda_k \bar{\Gamma} \bar{\Sigma} \bar{\Gamma} \Lambda_k)] - 2\text{Tr}[\Lambda_k B_k \bar{\Sigma} \bar{\Gamma}] \end{aligned} \quad (19)$$

$$\bar{\Sigma}' = \bar{\Sigma} - (\bar{\Sigma} \bar{\Gamma} \Lambda_k + \Lambda_k \bar{\Gamma} \bar{\Sigma}) + (\Lambda_k \text{Tr}[\bar{\Gamma} \bar{\Sigma} \bar{\Gamma} \Lambda_k] + \delta \Lambda_k \bar{\Gamma} \bar{\Sigma} \bar{\Gamma} \Lambda_k) \quad (20)$$

Note from (20) that choosing $\bar{\Sigma}$ to be diagonal, will be diagonal $\bar{\Sigma}'$ as well, suggesting a more compact notation. Thus, we introduce the $M \times 1$ vectors

$$\bar{\sigma} \triangleq \text{diag}\{\bar{\Sigma}\} \quad \bar{\sigma}' \triangleq \text{diag}\{\bar{\Sigma}'\} \quad b_k \triangleq \text{diag}\{\Lambda_k\} \quad (21)$$

Using the diagonal notation [1, 2] we obtain linear relation $\bar{\sigma}' = \bar{F}_k \bar{\sigma}$ where \bar{F}_k is a $M \times M$ matrix that includes statistics of local data as

$$\bar{F}_k = (1 - 2\beta_k + \delta\beta_k^2)I + \beta_k^2 b_k c_k^T \quad (22)$$

with $c_k = \text{diag}\{\Lambda_k^{-1}\}$ and β_k given by

$$\beta_k = \begin{cases} \frac{1-\lambda}{\gamma_k^{-1}}, & \text{for } \lambda \rightarrow 1 \\ \frac{1-\lambda}{\gamma_k^{-1}\lambda + (1-\lambda)M}, & \text{for smaller } \lambda \end{cases} \quad (23)$$

As a result, expression (19) becomes

$$E(\|\bar{\boldsymbol{\psi}}_k\|_{\text{diag}\{\bar{\sigma}\}}^2) = E(\|\bar{\boldsymbol{\psi}}_{k-1}\|_{\text{diag}\{\bar{\sigma}'\}}^2) + g_k \bar{\sigma} \quad (24)$$

where g_k is a row vector as

$$g_k = \sigma_{v,k}^2 \beta_k^2 c_k^T + (\text{diag}\{B_k\})^T \bar{F}_k \quad (25)$$

Comparing (25) with g_k in [2] we can see that $(\text{diag}\{B_k\})^T \bar{F}_k$ explicitly accounts for the noisy link effects. Thus, following the similar steps given in [1, 2] we can obtain the following expressions for MSD and EMSE

$$\eta_k = a_k (I - \Pi_{k,1})^{-1} r, \quad (\text{MSD}) \quad (26)$$

$$\zeta_k = a_k (I - \Pi_{k,1})^{-1} b_k, \quad (\text{EMSE}) \quad (27)$$

where $r \triangleq \text{diag}\{I\}$ and

$$\Pi_{k,l} \triangleq \bar{F}_{k+l-1} \bar{F}_{k+l} \dots \bar{F}_N \bar{F}_1 \dots \bar{F}_{k-1}, \quad l = 1, \dots, N \quad (28)$$

$$a_k \triangleq g_k \Pi_{k,2} + g_{k+1} \Pi_{k,3} + \dots + g_{k-2} \Pi_{k,N} + g_{k-1} \quad (29)$$

4 Simulation

In this section we provide computer simulations comparing the theoretical to simulation results. To this aim we consider a network with $N = 15$ nodes where each local filter has $M = 4$ taps. Each node accesses independent Gaussian regressors $u_{k,i}$ where their eigenvalue spread is $\rho = 5$. The observation noise variances $\sigma_{v,k}^2$ and $\text{Tr}[R_{u,k}]$ are shown in Fig. 2. We also assume $Q_k = 10^{-4}I_M$ to model the covariance matrix of link noise and $\lambda = 0.997$. The system evolves for 2000 iterations and the steady-state values are obtained by averaging the last 200 time samples. The curves are obtained by averaging over 100 experiments with $\lambda = 0.997$. In Fig. 3 the steady-state of MSD and EMSE are plotted. It is clear from Fig. 3 that there is a good match between simulations and theory which is an evidence that the derived expressions can describe the steady-state performance of incremental RLS adaptive network with noisy links.

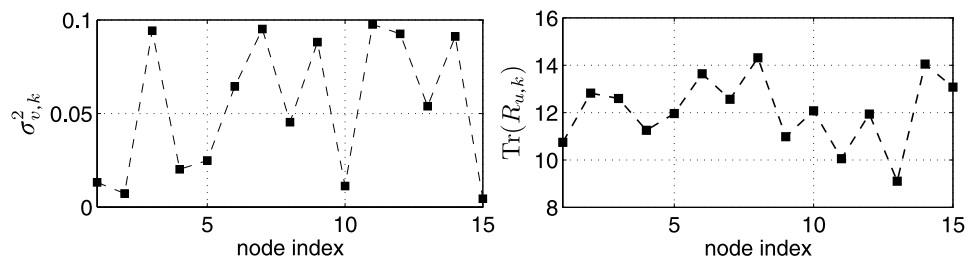


Fig. 2. The Observation noise profile, $\sigma_{v,k}^2$ and $\text{Tr}(R_{u,k})$.

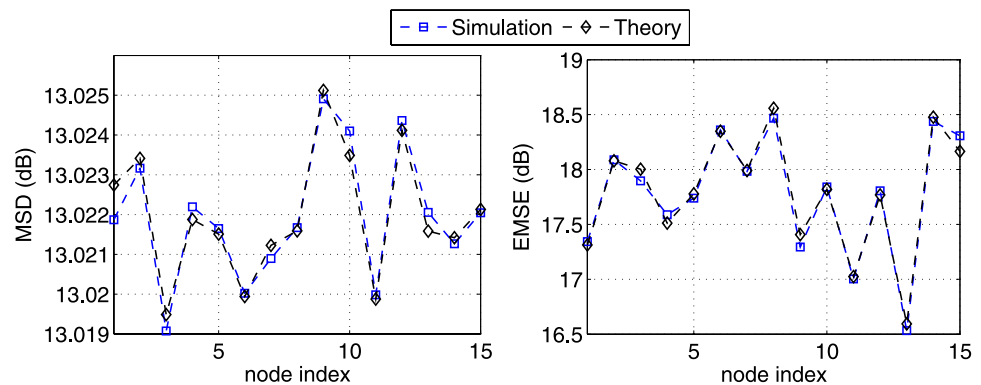


Fig. 3. The MSD and EMSE per node k – comparing simulation with theory.

5 Conclusion

In this paper, we studied the effect of noisy links on the performance of incremental RLS adaptive networks. Using weighted spatial-temporal energy conservation relation, we arrived a variance relation which contains moments that represent the effects of noisy links. We evaluated these moments and derived closed-form expressions for the MSD and EMSE to explain the steady-state performance at each individual node. The derived expressions have good match with simulations.