13 Communication-Constrained Learning

In our discussion of decentralized learning algorithms up to this point we have generally disregarded communication constraints. Of course, in practice, we can expect that multi-agent systems will be subject to a number of limitations which introduce additional noise and asynchrony into learning recursions. In this chapter, we broadly consider two types of imperfections:

- **Partial participation or communication:** Individual agents may fail to perform a local update. They may also fail to communicate, either with all of their neighbors, or with individual neighbors. We will see that some of these behaviors can be modelled using the stochastic gradient approximation framework of Chapter 3, and performance can be recovered from the expressions in Chapters 9 and 10, while others require us to extend and tailor the performance analysis to the setting at hand.
- **Imperfect communication:** Even if agents partcipate in update and communication, these links may be noisy, either due to channel noise in an analog channel, or quantization noise arising from a digital communication scheme. We will see again how some channel and quantization models can be captured by the stochastic gradient approximation framework, while others require us to adjust convergence analysis accordingly.

We will denote as "simple" communication constraints those that can be modelled through stochastic gradient approximations satisfying our typical gradient noise conditions, while "complex" communication constraints are those that require us to adjust convergence analysis. Even in those cases, however, the general theme and techniques of Chapters 9 and 10 apply.

13.1 SIMPLE COMMUNICATION CONSTRAINTS

13.1.1 Partial Local Updates

We encountered partial local updates already in Chapter 4 in the context of fusion-center based algorithms. In a decentralized context, we may equally envision a setting any given agent k, at any given time i may be unable to compute a local model update, either because it lacks the computational resources to com-

pute a gradient step, or because it does not have access to data at that particular point in time. Then, an intermediate update would be computed according to:

$$\boldsymbol{\psi}_{k,i} = \begin{cases} \boldsymbol{w}_{k,i-1} - \mu\widehat{\nabla J}_k(\boldsymbol{w}_{k,i-1}), & \text{w.p. } \pi_k, \\ \boldsymbol{w}_{k,i-1}, & \text{otherwise.} \end{cases} \tag{13.1}$$

The intermediate estimates $\boldsymbol{\psi}_{k,i}$ can then be further processed and combined using any scheme of choices. For example, in the case of the diffusion algorithm, we could compute:

$$\boldsymbol{w}_{k,i} = \sum_{\ell\in\mathcal{N}_k} a_{\ell k}\boldsymbol{\psi}_{\ell,i} \tag{13.2}$$

As we saw in Chapter 4, we may equivalently reformulate (13.1) as:

$$\boldsymbol{\psi}_{k,i} = \boldsymbol{w}_{k,i-1} - \mu\widehat{\nabla J}_k^{\,\text{part}}(\boldsymbol{w}_{k,i-1}) \tag{13.3}$$

where $\widehat{\nabla J}_k^{\,\text{part}}(\boldsymbol{w}_{k,i-1})$ is an adjusted stochastic gradient approximation of the form:

$$\widehat{\nabla J}_k^{\,\text{part}}(\boldsymbol{w}_{k,i-1}) = \begin{cases} \frac{1}{\pi_k}\widehat{\nabla J}_k(\boldsymbol{w}_{k,i-1}), & \text{w.p. } \pi_k, \\ 0, & \text{otherwise.} \end{cases} \tag{13.4}$$

It can be verified, that if we define the gradient noise processes:

$$\boldsymbol{s}_{k,i}(\boldsymbol{w}_{k,i-1}) \triangleq \widehat{\nabla J}_k(\boldsymbol{w}_{k,i-1}) - \nabla J_k(\boldsymbol{w}_{k,i-1}) \tag{13.5}$$

$$\boldsymbol{s}_{k,i}^{\text{part}}(\boldsymbol{w}_{k,i-1}) \triangleq \widehat{\nabla J}_k^{\,\text{part}}(\boldsymbol{w}_{k,i-1}) - \nabla J_k(\boldsymbol{w}_{k,i-1}) \tag{13.6}$$

and assume:

$$\mathbb{E}\,\{\boldsymbol{s}_{k,i}(\boldsymbol{w}_{k,i-1})|\boldsymbol{w}_{k,i-1}\} = 0 \tag{13.7}$$

$$\mathbb{E}\left\{\|\boldsymbol{s}_{k,i}(\boldsymbol{w}_{k,i-1})\|^2|\boldsymbol{w}_{k,i-1}\right\} \leq \beta_k^2\|w_k^o - \boldsymbol{w}_{k,i-1}\|^2 + \sigma_k^2 \tag{13.8}$$

then:

$$\mathbb{E}\left\{\boldsymbol{s}_{k,i}^{\text{part}}(\boldsymbol{w}_{k,i-1})|\boldsymbol{w}_{k,i-1}\right\} = 0 \tag{13.9}$$

$$\mathbb{E}\left\{\left\|\boldsymbol{s}_{k,i}^{\text{part}}(\boldsymbol{w}_{k,i-1})\right\|^2|\boldsymbol{w}_{k,i-1}\right\} \leq \left(\frac{\beta_k^2}{\pi_k} + \frac{1-\pi_k}{\pi_k}\delta_k^2\right)\|w_k^o - \boldsymbol{w}_{k,i-1}\|^2 + \frac{\sigma_k^2}{\pi_k} \tag{13.10}$$

It follows that we can recover convergence guarantees for decentralized penalty-based and primal-dual algorithms by making proper substitutions to the expressions in Chapters 9 and 10. For example, for the consensus+innovations and diffusion algorithms, we have from (10.2) that:

$$\limsup_{i\to\infty}\mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\|^2 \leq O\left(\frac{\mu\sum_{k=1}^K\frac{\sigma_k^2}{\pi_k}}{K^2\nu}\right) + O(\mu^2) \tag{13.11}$$

If agents are homogenous, this reduces to:

$$\limsup_{i\to\infty} \mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\|^2 \leq O\left(\frac{\mu\sigma^2}{\pi K\nu}\right) + O(\mu^2) \tag{13.12}$$

The quantity πK quantifies the *expected* number of updates performed across the network in any given iteration. If we assume that $\pi = \frac{1}{K}$, which means that in expectation only one agent performs and update at any given iteration, while all others merely participate in the diffusion of information, we have:

$$\limsup_{i\to\infty} \mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\|^2 \leq O\left(\frac{\mu\sigma^2}{\nu}\right) + O(\mu^2) \tag{13.13}$$

and we recover the performance of a single non-cooperative agent.

A second type of partial participation is one where agents may perform local updates at every iteration, but participate in the communication step intermittently, as is the case in federated learning (see Chapter 5). This kind of partial participation cannot be modelled in a straightforward manner using gradient approximations, and will instead involve the use of time-varying or stochastic combination policies, as we illustrate further ahead.

13.1.2 Quantized Communication Links

Whenever messages are sent over a bandwidth-constrained communication channel, this will inevitably be associated with imperfections in the exchanged messages, which can be modelled as noise. We illustrate this on the diffusion algorithm again, though similar constructions apply to other decentralized algorithms we have encountered. Recall the diffusion algorithm takes the form:

$$\boldsymbol{\psi}_{k,i} = \boldsymbol{w}_{k,i-1} - \mu\widehat{\nabla J}_k(\boldsymbol{w}_{k,i-1}) \tag{13.14}$$

$$\boldsymbol{w}_{k,i} = \sum_{\ell\in\mathcal{N}_k} a_{\ell k}\boldsymbol{\psi}_{\ell,i} \tag{13.15}$$

Agents in this framework exchange intermediate estimates $\boldsymbol{\psi}_{k,i}$. In a digital communication setting, this is achieved by quantizing the vector $\boldsymbol{\psi}_{k,i}$ for a given bit-budget, and subsequently communicating the bit-representation of $\boldsymbol{\psi}_{k,i}$. We employ the notation $\boldsymbol{\mathcal{Q}}_k(\cdot)$ for the general quantization scheme employed by agent k, and can then write:

$$\boldsymbol{\psi}_{k,i} = \boldsymbol{w}_{k,i-1} - \mu\widehat{\nabla J}_k(\boldsymbol{w}_{k,i-1}) \tag{13.16}$$

$$\boldsymbol{w}_{k,i} = \sum_{\ell\in\mathcal{N}_k} a_{\ell k}\boldsymbol{\mathcal{Q}}_k\left(\boldsymbol{\psi}_{\ell,i}\right) \tag{13.17}$$

Many quantization schemes have been developed and studied in the literature, with different properties, implementations and limitations. A detailed discussion of specific quantization schemes is beyond the scope of this text, and we refer the reader to [Nassif et al., 2022] [1] as well as the references therein for examples of

[1] R. Nassif, S. Vlaski, M. Carpentiero, V. Matta, M. Antonini, A. H. Sayed, "Quantization for decentralized learning under subspace constraints", available as arXiv:2209.07821, 2022.

quantization schemes. Instead, we remark that virtually all relevant quantization techniques satisfy the following condition:

CONDITION 13.1 (**Quantizer conditions**). The random quantization schemes $\boldsymbol{\mathcal{Q}}_k(\cdot)$ are unbiased, i.e.:

$$\mathbb{E}\left\{\boldsymbol{\mathcal{Q}}_k(\boldsymbol{\psi}_{k,i})|\boldsymbol{\psi}_{k,i}\right\} = \boldsymbol{\psi}_{k,i} \tag{13.18}$$

Furthermore, the variance of the quantization error satisfies the bound:

$$\mathbb{E}\left\{\left\|\boldsymbol{\psi}_{k,i} - \boldsymbol{\mathcal{Q}}_k(\boldsymbol{\psi}_{k,i})\right\|^2 |\boldsymbol{\psi}_{k,i}\right\} \leq \beta_{k,q}^2\|\boldsymbol{\psi}_{k,i}\|^2 + \sigma_{q,k}^2 \tag{13.19}$$

□

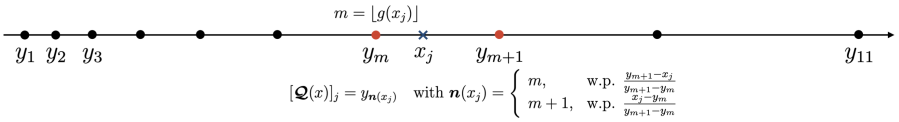

Figure 13.1 Illustration of a stochastic scheme with irregular quantization intervals [Nassif et al., 2022].

Comparing the quantizer conditions (13.18)–(13.19) with our typical gradient noise conditions, we observe the same structure with a relative component proportional to the norm of the quantized quantity, and an absolute component. Indeed, we can bound:

$$\begin{aligned}
&\mathbb{E}\left\{\left\|\boldsymbol{\psi}_{k,i} - \boldsymbol{\mathcal{Q}}_k(\boldsymbol{\psi}_{k,i})\right\|^2 |\boldsymbol{\psi}_{k,i}\right\} \\
&\leq \beta_{k,q}^2\|\boldsymbol{\psi}_{k,i}\|^2 + \sigma_{q,k}^2 \\
&\leq \beta_{k,q}^2\|w_k^o - \boldsymbol{\psi}_{k,i} - w_k^2\|^2 + \sigma_{q,k}^2 \\
&\overset{(a)}{\leq} 2\beta_{k,q}^2\|w_k^o - \boldsymbol{\psi}_{k,i}\|^2 + 2\beta_{k,q}^2\|w_k^o\|^2 + \sigma_{q,k}^2 \\
&\overset{(b)}{\leq} 2\beta_{k,q}^2\|w_k^o - \boldsymbol{w}_{k,i-1}\|^2 + 2\beta_{k,q}^2\|w_k^o\|^2 + \sigma_{q,k}^2
\end{aligned} \tag{13.20}$$

Here, (a) follows from Jensen's inequality, and (b) follows since $\boldsymbol{\psi}_{k,i}$ is obtain from $\boldsymbol{w}_{k,i-1}$ by taking a local stochastic gradient update in the direction of w_k^o. Further, we we defined:

$$\begin{aligned}
\overline{\beta}_{k,q}^2 &= 2\beta_{k,q}^2 \\
\overline{\sigma}_{q,k}^2 &= 2\beta_{k,q}^2\|w_k^o\|^2 + \sigma_{q,k}^2
\end{aligned} \tag{13.21}$$

Having reformulated the conditions on the quantization error, we can absorb it

into the gradient noise as above. Specifically, we have:

$$\boldsymbol{\phi}_{k,i} = \boldsymbol{w}_{k,i-1} - \mu\widehat{\nabla J}_k^{\,\text{quant}}(\boldsymbol{w}_{k,i-1}) \tag{13.22}$$

$$\boldsymbol{w}_{k,i} = \sum_{\ell\in\mathcal{N}_k} a_{\ell k}\boldsymbol{\phi}_{\ell,i} \tag{13.23}$$

where we defined $\boldsymbol{\phi}_{k,i} \triangleq \boldsymbol{\Omega}_k(\boldsymbol{\psi}_{k,i})$ and:

$$\widehat{\nabla J}_k^{\,\text{quant}}(\boldsymbol{w}_{k,i-1}) \triangleq \widehat{\nabla J}_k(\boldsymbol{w}_{k,i-1}) + \frac{1}{\mu}\left(\boldsymbol{\psi}_{k,i} - \boldsymbol{\Omega}_k(\boldsymbol{\psi}_{k,i})\right) \tag{13.24}$$

If we again assume that:

$$\mathbb{E}\left\{\boldsymbol{s}_{k,i}(\boldsymbol{w}_{k,i-1})|\boldsymbol{w}_{k,i-1}\right\} = 0 \tag{13.25}$$

$$\mathbb{E}\left\{\|\boldsymbol{s}_{k,i}(\boldsymbol{w}_{k,i-1})\|^2|\boldsymbol{w}_{k,i-1}\right\} \le \beta_k^2\|w_k^o - \boldsymbol{w}_{k,i-1}\|^2 + \sigma_k^2 \tag{13.26}$$

it follows that:

$$\mathbb{E}\left\{\boldsymbol{s}_{k,i}^{\text{quant}}(\boldsymbol{w}_{k,i-1})|\boldsymbol{w}_{k,i-1}\right\} = 0 \tag{13.27}$$

$$\mathbb{E}\left\{\left\|\boldsymbol{s}_{k,i}^{\text{quant}}(\boldsymbol{w}_{k,i-1})\right\|^2|\boldsymbol{w}_{k,i-1}\right\} \le \left(\beta_k^2 + \frac{\overline{\beta}_{q,k}^2}{\mu^2}\right)\|w_k^o - \boldsymbol{w}_{k,i-1}\|^2 + \left(\sigma_k^2 + \frac{\overline{\sigma}_{q,k}^2}{\mu^2}\right) \tag{13.28}$$

It then follows again from (10.2) that:

$$\limsup_{i\to\infty}\mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\|^2 \le O\left(\frac{\mu\sum_{k=1}^K\sigma_k^2}{K^2\nu}\right) + O\left(\frac{\mu^{-1}\sum_{k=1}^K\overline{\sigma}_{q,k}^2}{K^2\nu}\right) + O(\mu^2) \tag{13.29}$$

In a homogenous setting we recover:

$$\limsup_{i\to\infty}\mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\|^2 \le O\left(\frac{\mu\sigma^2}{K\nu}\right) + O\left(\frac{\mu^{-1}\overline{\sigma}_q^2}{K\nu}\right) + O(\mu^2) \tag{13.30}$$

We conclude that the limiting performance of a quantized decentralized algorithm is given by the superposition of two error terms, one arising from the classical gradient noise component σ^2, and the other arising from quantization noise $\overline{\sigma}_q^2$. It is important to note that the quantization noise component is multiplied by μ^{-1}, rather than μ, causing it to be amplified for small step-sizes. A second limitation of the current implementation is that the absolute quantization noise component $\overline{\sigma}_q^2$, in light of (13.20), grows with $\|w_k^o\|$. This is because the intermediate estimates $\boldsymbol{\psi}_{k,i}$ in the limit will be of the same order as the true models w_k^o, and large vectors, when quantized, result in large quantization error for a fixed bit-budget. We will show how these limitations can be addressed using *differential quantizaiton* further ahead.

13.2 COMPLEX COMMUNICATION CONSTRAINTS

13.2.1 Partial Communication

We saw in Section 13.1.1 that partial, or intermittent, local updates can be modelled readily within the framework of stocahstic gradient approximation. A second type of asynchrony encountered frequently in practical multi-agent systems is partial or intermittent *communication*. Agents may perform multiple local updates between every communication exchange. We already encountered these types of dynamics in the context of federated learning in Chapter 5. More generally, any given agent may only interact with a subset of its neighbors in any given iteration. We can model this kind of behavior using time-varying or random combination policies $\boldsymbol{\mathcal{A}}_i$. We will again illustrate these concepts for the diffusion algorithm, though similar conclusions hold for other types of decentralized strategy we studied so far.

Stochastic Combination Policies

We begin with a stochastic model for partial communication. To this end, consider the diffusion recursion in network form:

$$\boldsymbol{w}_i = \boldsymbol{\mathcal{A}}_i^{\mathsf{T}} \left(\boldsymbol{w}_{i-1} - \mu \widehat{\nabla J}(\boldsymbol{w}_{i-1}) \right) \tag{13.31}$$

If we let:

$$\boldsymbol{\mathcal{A}}_i = \begin{cases} \mathcal{A}, & \text{w.p. } \pi, \\ I_{KM}, & \text{otherwise.} \end{cases} \tag{13.32}$$

it follows that:

$$\boldsymbol{w}_i = \begin{cases} \mathcal{A}^{\mathsf{T}} \left(\boldsymbol{w}_{i-1} - \mu \widehat{\nabla J}(\boldsymbol{w}_{i-1}) \right), & \text{w.p. } \pi, \\ \boldsymbol{w}_{i-1} - \mu \widehat{\nabla J}(\boldsymbol{w}_{i-1}), & \text{otherwise.} \end{cases} \tag{13.33}$$

We can interpet this as a probabilistic and decentralized variant of the federated averaging algorithm of Chapter 5, where agents communicate with probability π, while performing only local updates with probability $1-\pi$. If we take expectations, conditioned on the past iterates $\boldsymbol{w}_{i-1}$, we have:

$$\mathbb{E}\left\{ \boldsymbol{w}_i \,|\, \boldsymbol{w}_{i-1} \right\} = \overline{\mathcal{A}}^{\mathsf{T}} \left(\boldsymbol{w}_{i-1} - \mu \nabla J(\boldsymbol{w}_{i-1}) \right) \tag{13.34}$$

where we defined:

$$\overline{\mathcal{A}} = \mathbb{E}\boldsymbol{\mathcal{A}}_i \tag{13.35}$$

It follows that on average the diffusion algorithm with stochastic, time-varying combination policy $\boldsymbol{\mathcal{A}}_i$ will behave as its mean policy $\overline{\mathcal{A}}$. It is then reasonable to expect that, as long as the mean-policy $\overline{\mathcal{A}}$ is strongly-connected, the diffusion algorithm with random combination policy will continue to yield meaningful estimates. We can expect some deterioration in performance resulting from the

variance of $\boldsymbol{\mathcal{A}}_i$ around $\overline{\mathcal{A}}$, which can be quantified. The following result has been established for the diffusion algorithm with random combination policy.

THEOREM 13.1 (Mean-square-behavior of the diffusion algorithm with random combination policy [Zhao and Sayed, 2014). [2]*] Under standard conditions on local gradient approximations and objective functions, and assuming the mean graph described by* $\overline{A}$ *is strongly-connected, there exists a step-size* μ *that is small enough, so that iterates generated by diffusion algorithm with random policy* (13.31) *converge in the mean-square sense and:*

$$\limsup_{i\to\infty} \mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\|^2 = O(\mu) \tag{13.36}$$

Deterministic Time-Varying Combination Policies

The combination polcies in the previous section were random and time-varying, but stationary with constant mean $\overline{A}$. There are also situations where we wish to allow for deterministic, but time-varying combination polcies. For example, the setting where agents perform E local updates inbetween every communication exchange can be modelled as:

$$\boldsymbol{w}_i = \mathcal{A}_i^{\mathsf{T}} \left(\boldsymbol{w}_{i-1} - \mu\widehat{\nabla\mathcal{J}}(\boldsymbol{w}_{i-1})\right) \tag{13.37}$$

where:

$$\mathcal{A}_i = \begin{cases} \mathcal{A}, & \text{if } i \pmod E = 0, \\ I_{KM}, & \text{otherwise.} \end{cases} \tag{13.38}$$

Most decentralized algorithms are also robust to this kind of asynchrony. Convergence guarantees generally rely on the condition that there exists an i^o, such that for every i, the product:

$$\mathcal{A}_{i,i^o} = \prod_{j=i}^{i+i^o} A_j \tag{13.39}$$

is a primitive matrix. This condition essentially ensures that the union of graphs over a time-span of length i^o is always strongly-connected. We refer the reader to [Nedić et al., 2016][3] as well as the references therein for a detailed discussion of the effects of deterministic time-varying combination policies.

13.2.2 Differential Quantization

We observed in 13.1.2 that direct quantization of intermediate estimates $\boldsymbol{\psi}_{k,i}$ results in significant deterioration of performance due to the fact that large

[2] X. Zhao and A. H. Sayed, "Asynchronous Adaptation and Learning Over Networks—Part I: Modeling and Stability Analysis," in IEEE Transactions on Signal Processing, vol. 63, no. 4, pp. 811-826, Feb.15, 2015, doi: 10.1109/TSP.2014.2385046.

[3] A. Nedić, A. Olshevsky, and W. Shi, "Achieving Geometric Convergence for Distributed Optimization Over Time-Varying Graphs", available as arXiv:1607.03218, 2016.

quantities need to be quantized, which in turn results in large quantization error for a moderate number of bits. This fact was particularly pronounced for small step-sizes μ, since it is proportional to μ^{-1} (see (13.29)). For small step-sizes, however, models are updated slowly, and hence there is significant correlation between subsequent model estimates. This motivates the introduction of *differential quantization* schemes, which quantize the model updates instead of the models directly. As an example, we consider the algorithm from [Nassif et al., 2022]:

$$\boldsymbol{\psi}_{k,i} = \boldsymbol{w}_{k,i-1} - \mu\widehat{\nabla J}_k(\boldsymbol{w}_{k,i-1}) \tag{13.40}$$

$$\boldsymbol{\phi}_{k,i} = \boldsymbol{\phi}_{k,i-1} + \boldsymbol{\Omega}_k\left(\boldsymbol{\psi}_{k,i} - \boldsymbol{\phi}_{k,i-1}\right) \tag{13.41}$$

$$\boldsymbol{w}_{k,i} = (1-\gamma)\boldsymbol{\phi}_{k,i} + \gamma\sum_{\ell\in\mathcal{N}_k} a_{\ell k}\boldsymbol{\phi}_{\ell,i} \tag{13.42}$$

Step (13.41) employs differential quantization, since we quantize $\boldsymbol{\psi}_{k,i} - \boldsymbol{\phi}_{k,i-1}$ rather than $\boldsymbol{\psi}_{k,i}$ directly. As the learning algorithm approaches convergence, we expect it to take smaller and smaller steps, which reduces the range of values that the quantizer $\boldsymbol{\Omega}_k(\cdot)$ will need to cover. This is in contrast to the absolute quantization scheme of Section 13.1.2, where the quantized values approach a fixed, potentially large value. The combination step (13.42) contains an additional damping parameter γ. Setting $\gamma = 1$ recovers the classical diffusion aggregation step, while choices $0 < \gamma < 1$ add more weight to agents' own models, which has been observed to provide additional stabilizing propeties when employing differential quantization.

THEOREM 13.2 (Mean-square-behavior of the diffusion algorithm with differential quantization [Nassif et al., 2022]). *Under standard conditions on local gradient approximations and objective functions, and condition 13.1 on the quantizers $\boldsymbol{\Omega}_k(\cdot)$, there exists a step-size μ that is small enough, so that iterates generated by diffusion with differential quantization and damping (13.40)–(13.42) converge in the mean-square sense and:*

$$\limsup_{i\to\infty}\mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\|^2 = O(\mu) + O(\mu^{-1}\sum_{k=1}^{K}\sigma_{q,k}^2) \tag{13.43}$$

Examining (13.43), we note that the a component $O(\mu^{-1}\sum_{k=1}^{K}\sigma_{q,k}^2)$ from the quantization process continues to affect the steady-state performance, and continues to be proportional to μ^{-1}. However, this term arises from $\sigma_{q,k}^2$ rather than $\overline{\sigma}_{q,k}^2$ and hence does not exhibit dependence on the potentially large term $\|w_k^o\|^2$. Instead, the quantization term is merely a result of the number of bits allocated to the algorithm, and can be reduced by increasing the number of bits. We show in Fig. 13.2 that very low steady-state error is possible using a small number of bits.

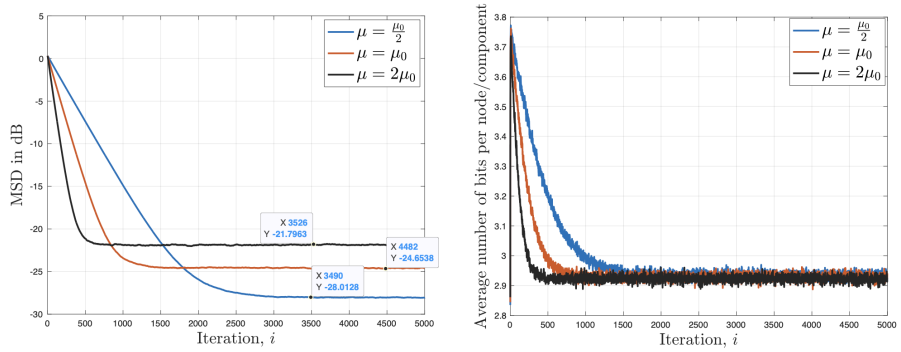

Figure 13.2 *(left)* Performance of the differentially quantized diffusion algorithm (13.40)–(13.41). *(right)* Number of bits needed to achieve this performance. [Nassif et al., 2022].