Online Optimization of Product-Form Networks

Jaron Sanders
Dept. of Math. & Comp. Science
Eindhoven University of Technology
5612 AZ, Den Dolech 2
Eindhoven, The Netherlands
jaron.sanders@tue.nl

Sem C. Borst
Dept. of Math. & Comp. Science
Eindhoven University of Technology
5612 AZ, Den Dolech 2
Eindhoven, The Netherlands
s.c.borst@tue.nl

Johan S.H. van Leeuwaarden
Dept. of Math. & Comp. Science
Eindhoven University of Technology
5612 AZ, Den Dolech 2
Eindhoven, The Netherlands
j.s.h.v.leeuwaarden@tue.nl

***Abstract*—We develop an online gradient algorithm for optimizing the performance of product-form networks through online adjustment of control parameters. The use of standard algorithms for finding optimal parameter settings is hampered by the prohibitive computational burden of calculating the gradient in terms of the stationary probabilities. The proposed approach instead relies on measuring empirical frequencies of the various states through simulation or online operation so as to obtain estimates for the gradient. Besides the reduction in computational effort, a further benefit of the online operation lies in the natural adaptation to slow variations in ambient parameters as commonly occurring in dynamic environments. On the downside, the measurements result in inherently noisy and biased estimates. We exploit mixing time results in order to overcome the impact of the bias and establish sufficient conditions for convergence to a globally optimal solution.**

***Index Terms*—Gradient algorithm, Markov processes, mixing times, online performance optimization, product-form networks, stochastic approximation, dynamic control.**

I. Introduction

Markov processes provide a versatile framework for modelling a wide variety of stochastic systems, ranging from communication networks and data center applications to content dissemination systems and physical or social interaction processes [1], [2], [3]. In particular, key performance measures of the system under consideration, e.g. buffer occupancies, response times, loss probabilities or user throughputs, can typically be expressed in terms of the stationary distribution $\boldsymbol{\pi}$ of the Markov process.

In many applications, the stationary distribution $\boldsymbol{\pi}$, and hence the performance measures or statistical properties, crucially depend on system parameters $\boldsymbol{r}$ that can be controlled, e.g. admission thresholds, service rates, link weights or resource capacities. In those cases, the interest is often not so much in evaluating the performance of the system for given parameter values, but rather in finding parameter settings $\boldsymbol{r}^{\text{opt}}$ that optimize the performance or achieve an optimal trade-off between service level and costs.

Specifically, let $\bar{u}(\boldsymbol{\pi}(\boldsymbol{r}))$ be a function expressing the performance objective (to be minimized) in terms of the stationary distribution $\boldsymbol{\pi}(\boldsymbol{r})$ as function of the system parameters $\boldsymbol{r}$ and let $c(\boldsymbol{r})$ be a function representing possible cost associated with $\boldsymbol{r}$, e.g. capital expense or power consumption. Introducing $u(\boldsymbol{r}) = \bar{u}(\boldsymbol{\pi}(\boldsymbol{r})) + c(\boldsymbol{r})$, the problem of interest may then be mathematically formulated as finding

$$\boldsymbol{r}^{\text{opt}} = \arg\min_{\boldsymbol{r}} u(\boldsymbol{r}). \tag{1}$$

It is worth observing here that the problem formulation differs from the typical Markov decision processes [4], [5], which focus on selecting optimal actions in various states rather than identifying optimal parameter values.

Optimization problem (1) could in principle be solved using mathematical programming approaches such as gradient-based schemes. In addition to the usual convexity issues, however, a further difficulty arises from the fact that the stationary distribution $\boldsymbol{\pi}(\boldsymbol{r})$ is only implicitly determined as a function of $\boldsymbol{r}$ by the balance equations and is rarely available in explicit form, which severely complicates both the evaluation of the objective function $u(\boldsymbol{r})$ and calculation of its gradient $\nabla_{\boldsymbol{r}} u(\boldsymbol{r})$.

In the present paper we develop a gradient approach to solve the optimization problem (1) for a class of Markov processes with product-form distributions. This class of processes arises in a rich family of stochastic models, such as loss networks [6], [7], open and closed queueing networks [8], [9], wireless random-access networks [10], [11] and various types of interacting-particle systems [1], [3].

As we will show, the partial derivatives $\partial\boldsymbol{\pi}(\boldsymbol{r})/\partial\boldsymbol{r}$ for this class of processes can be written as linear combinations of products of stationary probabilities $\boldsymbol{\pi}(\boldsymbol{r})$, thus reducing the computation of the gradient to the evaluation of the equilibrium distribution. The problem that yet remains in many situations is that the stationary probabilities involve a normalization constant whose calculation is computationally intensive and potentially NP-hard [12]. This issue is particularly pertinent in the context of iterative optimization algorithms such as gradient-based schemes, where partial derivatives need to be calculated repeatedly.

In order to circumvent the computational burden of calculating the stationary probabilities, we adopt a gradient approach which relies on measuring the empirical frequencies of the various states so as to estimate the partial derivatives. Specifically, in each iteration we observe the stochastic process for some time period through simulation or online operation, and we then calculate estimates for the gradient based on the measured time fractions of the various states. Although the number of states may be extremely large, it turns out that in many situations one only needs to track the time

VALUETOOLS 2012, October 09-12, Cargèse, France

DOI 10.4108/valuetools.2012.250261

fractions of aggregate states rather than all individual states, and that these aggregate states can be observed in an entirely distributed fashion. Besides the reduction in computational effort, a further benefit of the online operation lies in the fact that the algorithm will automatically adapt to slow variations in ambient parameters which are fairly common in dynamic environments.

While the measurements bypass the computational effort of calculating the stationary probabilities, they result in inherently *noisy* and *biased* estimates for the gradient. The issue of noisy estimates is paramount in the field of stochastic approximation, where years of research have resulted in many robust stochastic approximation schemes which can cope with various stochastic processes and forms of random noise [13], [14]. In contrast, biased estimates present a much trickier issue, which is usually not accounted for in stochastic approximation schemes. In order to neutralize the impact of the bias, we focus the attention on the family of reversible processes within the above-mentioned class of Markov processes with product-form distributions [9]. For reversible processes, powerful results are known for mixing times [15], [16], which allow us to derive sufficient conditions guaranteeing convergence to the optimal solution of (1). Intuitively, the mixing times provide an indication for the period of time that we need to observe the stochastic process in order to overcome the impact of the bias.

As a further condition to ensure convergence to the globally optimal solution of (1) rather than a possible local optimum, we assume the optimization objective $u(\boldsymbol{r})$ to be convex in $\boldsymbol{r}$. While convexity is generally non-trivial to establish, this can be easily verified for the broad class of so-called log-likelihood functions

$$u(\boldsymbol{r}) = \bar{u}(\boldsymbol{\pi}(\boldsymbol{r})) = -\boldsymbol{\alpha}^{\mathrm{T}} \ln \boldsymbol{\pi}(\boldsymbol{r}) = -\sum_{x\in\Omega} \alpha_x \ln \pi_x(\boldsymbol{r}), \quad (2)$$

where Ω denotes the state space of the process, α_x are fixed coefficients and $\pi_x(\boldsymbol{r})$ is the stationary probability of state x. Taking partial derivatives of (2), we find that the first-order conditions reduce to linear constraints in terms of the stationary probabilities. In other words, the problem of attaining target values for expectations of functionals of the stationary distribution can be cast as an optimization objective of the form (2). A special case of (2) was recently investigated by Jiang and Walrand [17], [18]. Their goal was to achieve target throughput values in CSMA networks by using an algorithm that adjusts the access or backoff parameters (represented by the vector $\boldsymbol{r}$ in (2)) using empirical arrival and service rates. This in fact provided valuable inspiration for the work presented here, where we extend the scope of such algorithms to general product-form Markov processes and a larger class of objective functions. These generalizations require a different approach to deal with the impact of bias, as discussed in §IV-B1.

Further important related work is done by Marbach and Tsitsiklis [19], [20], see also [21] for further background. In [19], [20], an algorithm similar in spirit to ours is considered - an algorithm that aims to tackle a parameter optimization problem by relying on measurement-based evaluation of a gradient. Their convergence proof also involves analysis of noisy and biased estimates and the generic use of Lyapunov functions and martingale arguments. However, their expression for the gradient is fundamentally different and hence the specific proof arguments substantially differ as well. Although [19], [20] can be applied to more general Markov processes and furnishes greater versatility in use, it does not take advantage of simplifications that arise from the specific structure of product-form distributions as in this paper. Most importantly, however, the algorithm in [19], [20] differs in its updating method, because it updates parameters whenever the process visits recurrent states. Knowing whether the entire system is in a recurrent state (and thus when to update) requires information about all components of the system, making the algorithm in [19], [20] global in nature. This differs from our algorithm and that presented in [17], [18], which can be implemented in a distributed manner.

The remainder of the paper is organized as follows. In §II, we present a detailed problem formulation, develop our measurement-based optimization algorithm and state our main results. Some illustrative application scenarios are described next in §III. In §IV, we first identify conditions in terms of the measurement noise and bias which ensure the convergence of the algorithm, and we then prove that these conditions are satisfied.

II. Algorithm description

Throughout this paper, we denote by b_i the i-th component of vector $\boldsymbol{b}$. When taking a scalar function of an n-dimensional vector $\boldsymbol{b}$, we do this component-wise, i.e. $\exp \boldsymbol{b} = (\exp b_1, ..., \exp b_n)^{\mathrm{T}}$. If we have a $|\Omega|$-dimensional vector $\boldsymbol{b}$ in which each component corresponds to some state $x \in \Omega$, we write b_x for that component of $\boldsymbol{b}$ that corresponds to state x. Similarly, we denote by $A_{i,j}$ the element in row i, column j of matrix A. If rows and/or columns correspond to states in Ω, we write $A_{x,y}$ instead. Finally, we denote by $\mathbf{1}_n$ the n-dimensional vector of which all components equal one.

A. Gradient scheme

Consider a Markov process $\{X(t)\}_{t\geq 0}$ that is irreducible, reversible and has a finite state space Ω. Let $\boldsymbol{\pi}(\boldsymbol{r})$ denote its steady-state probability vector as a function of d parameters $\boldsymbol{r} = (r_1, ..., r_d)^{\mathrm{T}}$, which arises naturally if one has a closed-form expression for the stationary distribution. The most prominent examples are the product-form distributions

$$\boldsymbol{\pi}(\boldsymbol{r}) = \frac{1}{Z(\boldsymbol{r})} \exp\left(A\boldsymbol{r} + \boldsymbol{b}\right), \quad (3)$$

where $A \in \mathbb{R}^{|\Omega|\times d}$ is a matrix, $\boldsymbol{b} \in \mathbb{R}^{|\Omega|}$ is a vector and $Z(\boldsymbol{r})$ is the normalization constant.

We consider the optimization problem

$$\min_{\boldsymbol{r}\in\mathcal{R}} u(\boldsymbol{r}), \quad (4)$$

where $u(\boldsymbol{r})$ denotes an objective function that we assume to be convex in $\boldsymbol{r}$ on a hypercube $\mathcal{R} \subset \mathbb{R}^d$, representing the feasible range for the parameters $\boldsymbol{r}$. We furthermore require that (4) has a unique minimizer $\boldsymbol{r}^{\text{opt}} = \arg\min_{\boldsymbol{r}\in\mathcal{R}} u(\boldsymbol{r})$, and we assume that the gradient of $u(\boldsymbol{r})$ can be written as a function of $\boldsymbol{\pi}(\boldsymbol{r})$ and $\boldsymbol{r}$, i.e. $\nabla_{\boldsymbol{r}} u(\boldsymbol{r}) = \boldsymbol{g}(\boldsymbol{\pi}(\boldsymbol{r}), \boldsymbol{r})$ where $\nabla_{\boldsymbol{r}} = (\partial/\partial r_1, ..., \partial/\partial r_d)^{\mathrm{T}}$. For example when $c(\boldsymbol{r}) = 0$, the gradient of $u(\boldsymbol{r}) = \bar{u}(\boldsymbol{\pi}(\boldsymbol{r}))$ can be written as

$$\frac{\partial \bar{u}(\boldsymbol{\pi}(\boldsymbol{r}))}{\partial r_i} = \sum_{x\in\Omega} \frac{\partial \bar{u}(\boldsymbol{\pi}(\boldsymbol{r}))}{\partial \pi_x(\boldsymbol{r})} \frac{\partial \pi_x(\boldsymbol{r})}{\partial r_i} \tag{5}$$

for $i = 1, ..., d$. For the important case of product-form distributions in (3), we have

$$\begin{aligned}\frac{\partial \pi_x(\boldsymbol{r})}{\partial r_i} &= \frac{1}{Z(\boldsymbol{r})^2}\Big(Z(\boldsymbol{r}) A_{x,i} \exp\left(A\boldsymbol{r} + \boldsymbol{b}\right)_x \\ &\quad - \exp\left(A\boldsymbol{r} + \boldsymbol{b}\right)_x \sum_{y\in\Omega} A_{y,i} \exp\left(A\boldsymbol{r} + \boldsymbol{b}\right)_y\Big) \\ &= \pi_x(\boldsymbol{r})\Big(A_{x,i} - \sum_{y\in\Omega} A_{y,i}\pi_y(\boldsymbol{r})\Big),\end{aligned} \tag{6}$$

so that $\partial \bar{u}(\boldsymbol{\pi}(\boldsymbol{r}))/\partial r_i = g_i(\boldsymbol{\pi}(\boldsymbol{r}))$ and therefore $\nabla_{\boldsymbol{r}} u(\boldsymbol{r}) = \boldsymbol{g}(\boldsymbol{\pi}(\boldsymbol{r}))$. While for this example the gradient can be written as a function of only $\boldsymbol{\pi}(\boldsymbol{r})$, in §III-A we will encounter an example for which it is more efficient to write the gradient as a function of both $\boldsymbol{\pi}(\boldsymbol{r})$ and $\boldsymbol{r}$. For a calculation of such partial derivatives in a more general case of product-form networks, we refer the reader to [22].

Our goal is to find $\boldsymbol{r}^{\text{opt}}$ and in order to do so, it is natural to consider the gradient algorithm

$$\boldsymbol{r}^{[n+1]} = [\boldsymbol{r}^{[n]} - a^{[n+1]}\boldsymbol{g}^{[n+1]}]^{\mathcal{R}}, \tag{7}$$

where $\boldsymbol{g}^{[n+1]} = \boldsymbol{g}(\boldsymbol{\pi}(\boldsymbol{r}^{[n]}), \boldsymbol{r}^{[n]})$, and $n \in \mathbb{N}$ indexes the iteration. The $a^{[n]} \in (0, \infty)$ denote the step sizes of the algorithm, and we define the truncation operator as follows.

Definition 1. *For $\mathcal{R} \subset \mathbb{R}^d$ of the form*

$$\mathcal{R} = [\mathcal{R}_1^{\min}, \mathcal{R}_1^{\max}] \times ... \times [\mathcal{R}_d^{\min}, \mathcal{R}_d^{\max}], \tag{8}$$

the truncation $[\boldsymbol{r}]^{\mathcal{R}} \in \mathbb{R}^d$ of $\boldsymbol{r} \in \mathbb{R}^d$ is defined component-wise as

$$[\boldsymbol{r}]_i^{\mathcal{R}} = \max\big\{\mathcal{R}_i^{\min}, \min\big\{\mathcal{R}_i^{\max}, r_i\big\}\big\}. \tag{9}$$

B. Online gradient algorithm

It is well known that under suitable assumptions on the objective function and step sizes, the gradient algorithm in (7) generates a sequence $\boldsymbol{r}^{[n]}$ that converges to the optimal solution $\boldsymbol{r}^{\text{opt}}$. We also come back to this at the end of §IV-A. Calculating the gradient, however, may be difficult in practice, because it depends on $\boldsymbol{\pi}(\boldsymbol{r})$, limiting the applicability of (7).

Instead of using (7), we will estimate $\boldsymbol{\pi}(\boldsymbol{r})$ by observing the evolution of the system. These observations will take place during time intervals $[t^{[n]}, t^{[n+1]}]$, where $0 = t^{[0]} < t^{[1]} < ....$ At the end of each interval, say at time $t^{[n+1]}$, our algorithm will change the current system parameters $\boldsymbol{R}^{[n]}$ to new parameters $\boldsymbol{R}^{[n+1]}$ based on its observations.

The stochastic process $\{Y(t)\}_{t\geq 0}$ that describes the system is given by $Y(t) = Z^{[n]}(t)$, where n is such that $t \in [t^{[n]}, t^{[n+1]}]$. The process $\{Z^{[n]}(t)\}_{t^{[n]}\leq t\leq t^{[n+1]}}$ is a time-homogeneous Markov process, which starts in $Z^{[n-1]}(t^{[n]})$ and evolves according to the generator of $\{X(t)\}_{t\geq 0}$ that corresponds to parameters $\boldsymbol{R}^{[n]}$.

Let us now make precise how our algorithm observes the system and makes decisions. At time $t^{[n+1]}$, marking the end of observation period $n+1$, we calculate

$$\hat{\Pi}_x^{[n+1]} = \frac{1}{t^{[n+1]} - t^{[n]}} \int_{t^{[n]}}^{t^{[n+1]}} \mathbb{1}[Z^{[n]}(t) = x] dt \tag{10}$$

for every state $x \in \Omega$. During each interval, one thus keeps track of the fractions of time that the system is in every state. This constitutes an empirical estimate of $\boldsymbol{\pi}(\boldsymbol{R}^{[n]})$. We then estimate the gradient $\boldsymbol{G}^{[n+1]} = \boldsymbol{g}(\boldsymbol{\pi}(\boldsymbol{R}^{[n]}), \boldsymbol{R}^{[n]})$ by $\hat{\boldsymbol{G}}^{[n+1]} = \boldsymbol{g}(\hat{\boldsymbol{\Pi}}^{[n+1]}, \boldsymbol{R}^{[n]})$. If we then apply (7) using the estimated gradient instead of the actual gradient, we are essentially using the stochastic gradient algorithm

$$\boldsymbol{R}^{[n+1]} = [\boldsymbol{R}^{[n]} - a^{[n+1]}\hat{\boldsymbol{G}}^{[n+1]}]^{\mathcal{R}} \tag{11}$$

to update the parameters.

Note that algorithm (7) is deterministic, whereas (11) is stochastic. Also note that because we are estimating the gradient instead of explicitly calculating it, the algorithm in (11) is no longer guaranteed to converge to $\boldsymbol{r}^{\text{opt}}$.

C. Main result

We now present technical assumptions which will guarantee convergence of (11). For this, we need an additional sequence $e^{[n]}$ which we shall refer to as the error. It is related to the maximum allowable error when estimating the steady-state probability vector, which will be made precise in §IV-B1.

We require the sequences $a^{[n]}$, $e^{[n]}$ and $f^{[n]} = 1/(t^{[n]} - t^{[n-1]})$ to be such that

$$\sum_{n=1}^{\infty} a^{[n]} = \infty, \ \sum_{n=1}^{\infty} (a^{[n]})^2 < \infty, \tag{12}$$

and

$$\sum_{n=1}^{\infty} a^{[n]} e^{[n]} < \infty, \ \sum_{n=1}^{\infty} a^{[n]} \exp\Big(-\frac{(e^{[n]})^2}{4|\Omega|^2 \kappa f^{[n]}}\Big) < \infty, \tag{13}$$

for any $\kappa \in (0, \infty)$. We also require boundedness and regularity of $\boldsymbol{g}(\boldsymbol{\pi}(\boldsymbol{r}), \boldsymbol{r})$, in the sense that there exist constants $c_{\mathrm{g}}, c_{\mathrm{l}} \in [0, \infty)$ such that

$$|g_i(\boldsymbol{\mu}, \boldsymbol{r}) - g_i(\boldsymbol{\nu}, \boldsymbol{r})| \leq c_{\mathrm{l}} ||\boldsymbol{\mu} - \boldsymbol{\nu}||_{\mathrm{var}} \text{ for } i = 1, ..., d, \tag{14}$$

$$\|\boldsymbol{g}(\boldsymbol{\mu}, \boldsymbol{r})\|_2 \leq c_{\mathrm{g}}, \tag{15}$$

for all probability vectors $\boldsymbol{\mu}, \boldsymbol{\nu}$ and all $\boldsymbol{r} \in \mathcal{R}$. Here, $||\boldsymbol{\mu} - \boldsymbol{\nu}||_{\mathrm{var}} = \frac{1}{2}\sum_{x\in\Omega} |\mu_x - \nu_x|$ is the total variation distance. Under conditions (12) – (15) and the assumptions in §II-A and §II-B, the following result holds.

Theorem 1. *The sequence* $\boldsymbol{R}^{[n]}$ *generated by the online algorithm* (11) *converges to the optimal solution* $\boldsymbol{r}^{\text{opt}}$ *of the optimization problem* (4) *with probability one.*

Condition (12) is typical in stochastic approximation. It ensures that step sizes become smaller as n increases, while remaining large enough so that the algorithm does not get stuck in a suboptimal solution. Condition (13) then requires that the error $e^{[n]}$ for which we allow when estimating the steady-state probability vector must decrease. In order to guarantee this, the observation frequency $f^{[n]}$ must eventually become smaller than the error, i.e. $(e^{[n]})^2/f^{[n]} \to \infty$ as $n \to \infty$. Condition (14) ensures that when we approximate the gradient of $u(\boldsymbol{r})$ by using empirical distributions that come increasingly closer to the actual $\boldsymbol{\pi}(\boldsymbol{r})$, our approximation of the gradient also comes increasingly closer to the actual gradient. It is the most non-trivial of all conditions and verification can be cumbersome. In §III we discuss two illustrative examples for which (14) holds. Lastly, condition (15) guarantees that the gradient does not explode, preventing the algorithm from making extremely large errors.

It is not difficult to define sequences that satisfy (12) and (13). For example, setting $a^{[n]} = n^{-1}$, $f^{[n]} = n^{-2\alpha-\beta}$ and $e^{[n]} = n^{-\alpha}$ with $\alpha, \beta > 0$ suffices. In particular, note that for $\alpha = \beta = 1/3$, we have $a^{[n]} = n^{-1}$ and $t^{[n+1]} - t^{[n]} = n+1$, which expresses that the algorithm should take smaller steps as time increases, while simultaneously lengthening the observation period.

The choices for $a^{[n]}$, $e^{[n]}$ and $f^{[n]}$ strongly influence the behavior of the algorithm. Consider for instance the following two cases. Setting $a^{[n]} = n^{-1/2-\alpha}$ with $0 < \alpha \ll 1/2$ so that it barely satisfies (12), allows us to let $e^{[n]}$ decrease as slowly as $e^{[n]} = n^{-1/2}$. By (13) we then need that $f^{[n]} < n^{-1}$ or $t^{[n]} - t^{[n-1]} > n$. If we now consider the faster decreasing step size $a^{[n]} = n^{-1}$, which also barely satisfies (12), we find that a much slower decreasing $e^{[n]} = n^{-\alpha}$ with $0 < \alpha \ll 1$ suffices, implying by (13) that $f^{[n]} < n^{-2\alpha}$ or $t^{[n]} - t^{[n-1]} > n^{2\alpha}$ is required. From these two cases, one sees that smaller step sizes allow for shorter observation periods (recall that $0 < \alpha \ll 1$). The search for optimal settings of $a^{[n]}$, $e^{[n]}$ and $f^{[n]}$ is an important topic for future research.

III. Example applications

We now discuss two example scenarios in which Theorem 1 can be applied. The first scenario concerns the optimal trade-off between performance and costs in an Erlang loss system. The second scenario considers a log-likelihood function as an objective function in combination with product-form stationary distributions. We should stress that these two examples, particularly the first one, primarily serve to illuminate the core features of our algorithm in relatively simple settings. These scenarios are not meant to reflect the full scope or unique realm of our algorithm and could conceivably also be tackled via alternative methods.

A. Optimizing service, cost trade-off

Consider the $M/M/s/s$ queue. Customers arrive according to a Poisson process with rate λ and each customer has an exponentially distributed service requirement with unit mean. Each of the s parallel servers works at rate r. The steady-state probability of $x \in \Omega = \{0, 1, ..., s\}$ customers in the system is then given by

$$\pi_x(r) = \frac{\big(\lambda/r\big)^x/x!}{\sum_{y=0}^{s}\big(\lambda/r\big)^y/y!}. \tag{16}$$

The steady-state probability that an arriving customer finds all servers occupied and is blocked is given by the Erlang loss formula $B(s,r) = \pi_s(r)$. The mean stationary queue length is given by $L(s,r) = \sum_{x=1}^{s} x\pi_x(r)$, and by Little's law, $L(s,r) = \lambda(1 - B(s,r))/r$.

Suppose now that we want to minimize $B(s,r)$ by adjusting r and that the costs of operating at service rate r equal $c(r)$. Assume $c(r)$ to be convex in r and its derivative $c'(r)$ to be bounded for all $r \in \mathcal{R}$. We thus aim to minimize $u(r) = B(s,r) + c(r)$. This objective function is convex in r [23]. Furthermore,

$$g(\boldsymbol{\pi}(r), r) = \frac{B(s,r)(L(s,r) - s)}{r} + c'(r), \tag{17}$$

for which we prove the following result in Appendix A.

Lemma 1. *If* $\mathcal{R} = [\mathcal{R}^{\min}, \mathcal{R}^{\max}]$ *with* $0 < \mathcal{R}^{\min} < \mathcal{R}^{\max} < \infty$ *and* $g(\boldsymbol{\mu}, r)$ *is given by* (17)*, then there exists constants* $c_{\text{g}}, c_{\text{l}} \in [0, \infty)$ *such that conditions* (14)*,* (15) *hold for all probability vectors* $\boldsymbol{\mu}, \boldsymbol{\nu}$ *and all* $r \in \mathcal{R}$*.*

Using Lemma 1 we conclude that all conditions of Theorem 1 are met and that the gradient algorithm

$$R^{[n+1]} = \Big[R^{[n]} - a^{[n+1]}\Big(\frac{\hat{B}^{[n+1]}(\hat{L}^{[n+1]} - s)}{R^{[n]}} + c'(R^{[n]})\Big)\Big]^{\mathcal{R}}$$

converges to the optimal solution. Here, $\hat{B}^{[n+1]} = \hat{\Pi}_s^{[n+1]}$ denotes an estimate of the loss probability and $\hat{L}^{[n+1]} = \sum_{x=1}^{s} x\hat{\Pi}_x^{[n+1]}$ denotes an estimate of the mean queue length.

B. Log-likelihood and product forms

Consider the log-likelihood function as defined in (2) as objective function. We prove the following result in Appendix B.

Lemma 2. *If* $\boldsymbol{\pi}(\boldsymbol{r})$ *satisfies the product form* (3)*, then the log-likelihood function* $u(\boldsymbol{r})$ *in* (2) *is convex in* $\boldsymbol{r}$*.*

Using $\partial\bar{u}(\boldsymbol{\pi}(\boldsymbol{r}))/\partial\pi_x = -\alpha_x/\pi_x$ and substituting (6) into (5) yields

$$g_i(\boldsymbol{\pi}(\boldsymbol{r})) = \sum_{x\in\Omega}\alpha_x\Big(\sum_{y\in\Omega}A_{y,i}\pi_y(\boldsymbol{r}) - A_{x,i}\Big). \tag{18}$$

We will only consider $\boldsymbol{\alpha} \in (0,1)^{|\Omega|}$ that are probability vectors, so that $\mathbf{1}_{|\Omega|}{}^{\text{T}}\boldsymbol{\alpha} = 1$. We can then interpret (18) as the difference between the expectation with respect to $\boldsymbol{\pi}(\boldsymbol{r})$,

denoted by $(A^{\mathrm{T}}\boldsymbol{\pi}(\boldsymbol{r}))_i = \sum_{y\in\Omega} A_{y,i}\pi_y(\boldsymbol{r})$, and the expectation with respect to $\boldsymbol{\alpha}$, denoted by $(A^{\mathrm{T}}\boldsymbol{\alpha})_i = \sum_{x\in\Omega} A_{x,i}\alpha_x$, so that

$$\boldsymbol{g}(\boldsymbol{\pi}(\boldsymbol{r})) = A^{\mathrm{T}}\boldsymbol{\pi}(\boldsymbol{r}) - A^{\mathrm{T}}\boldsymbol{\alpha}. \tag{19}$$

We assume that $\boldsymbol{r}^{\mathrm{opt}}$ lies in the interior of $\mathcal{R}$, in which case optimality requires $\boldsymbol{g}(\boldsymbol{\pi}(\boldsymbol{r}^{\mathrm{opt}})) = \mathbf{0}$ and thus $A^{\mathrm{T}}\boldsymbol{\pi}(\boldsymbol{r}^{\mathrm{opt}}) = A^{\mathrm{T}}\boldsymbol{\alpha}$. We call $\boldsymbol{\gamma} = A^{\mathrm{T}}\boldsymbol{\alpha}$ the target vector, a name inspired by the fact that our algorithm seeks $\boldsymbol{r}^{\mathrm{opt}}$ such that $A^{\mathrm{T}}\boldsymbol{\pi}(\boldsymbol{r}^{\mathrm{opt}}) = \boldsymbol{\gamma}$.

Because $u(\boldsymbol{r})$ is convex in $\boldsymbol{r}$ and the target $\boldsymbol{\gamma}$ is achieved by the solution $\boldsymbol{r}^{\mathrm{opt}}$ of (4), we want to use our online gradient algorithm (11) to find $\boldsymbol{r}^{\mathrm{opt}}$. From (19), it follows that $|g_i(\boldsymbol{\mu}) - g_i(\boldsymbol{\nu}) \le 2\max_{x,i}\{|A_{x,i}|\}||\boldsymbol{\mu}-\boldsymbol{\nu}||_{\mathrm{var}}$ for $i = 1, ..., d$, and that $\|\boldsymbol{g}(\boldsymbol{\mu},\boldsymbol{r})\|_2 \le |\Omega| d \max_{x,i}\{|A_{x,i}|\}$, so that (14) and (15) are satisfied. Using Theorem 1, we then arrive at the following result.

Theorem 2. *Given any $\boldsymbol{\gamma} \in \mathbb{R}^d$ for which there exists an $\boldsymbol{r}^{\mathrm{opt}}$ in the interior of $\mathcal{R}$ so that $A^{\mathrm{T}}\boldsymbol{\pi}(\boldsymbol{r}^{\mathrm{opt}}) = \boldsymbol{\gamma}$, the online gradient algorithm*

$$\boldsymbol{R}^{[n+1]} = [\boldsymbol{R}^{[n]} - a^{[n+1]}\big(A^{\mathrm{T}}\hat{\boldsymbol{\Pi}}^{[n+1]} - \boldsymbol{\gamma}\big)]^{\mathcal{R}} \tag{20}$$

converges to $\boldsymbol{r}^{\mathrm{opt}}$ with probability one.

As an illustrative example, consider a loss network consisting of L links with capacities $\boldsymbol{c} = (c_1, ..., c_L)^{\mathrm{T}}$ shared by K customer classes. Class-k customers arrive according to a Poisson process with rate λ_k and require exponentially distributed holding times with mean $1/\mu_k$. Each class-k customer requires capacity $B_{k,l}$ on link l for the duration of its holding time, i.e. $B_{k,l} = b_k J_{k,l}$, where b_k is the nominal capacity requirement of a class-k customer and $J_{k,l}$ has the value 0 or 1, indicating whether the route of class-k customers contains link l or not. When an arriving class-k customer finds insufficient capacity available, it is blocked and lost. Denote the number of class-k customers in the network at time t by $X_k(t)$ and define $\boldsymbol{X}(t) = (X_1(t), ..., X_K(t))^{\mathrm{T}}$. Under these assumptions, $\{\boldsymbol{X}(t)\}_{t\ge 0}$ is a reversible Markov process with state space $\Omega = \{\boldsymbol{x} \in \mathbb{N}^K | B\boldsymbol{x} \le \boldsymbol{c}\}$ and steady-state probability vector

$$\pi_{\boldsymbol{x}}(\boldsymbol{\rho}) = \frac{1}{Z(\boldsymbol{\rho})}\prod_{k=1}^{K}\frac{(\rho_k)^{x_k}}{x_k!}, \text{ where } Z(\boldsymbol{\rho}) = \sum_{y\in\Omega}\prod_{k=1}^{K}\frac{(\rho_k)^{y_k}}{y_k!}.$$

Here, $\rho_k = \lambda_k/\mu_k$ denotes the offered traffic of class k. Rewriting gives

$$\pi_{\boldsymbol{x}}(\boldsymbol{\rho}) = \frac{1}{Z(\boldsymbol{\rho})}\exp\Big(\sum_{k=1}^{K} x_k \ln\rho_k - \ln(x_k!)\Big), \tag{21}$$

which matches (3) with $d = K$, $r_k = \ln\rho_k$, $A_{x,k} = x_k$ and $b_x = -\sum_{k=1}^{K}\ln(x_k!)$. Note that $(A^{\mathrm{T}}\boldsymbol{\pi}(\boldsymbol{r}))_k = \sum_{y\in\Omega} y_k\pi_y$ is the carried traffic of class k, i.e. the steady-state average number of class-k customers in the system, which we can empirically estimate by observing the system. We apply our algorithm by setting

$$\boldsymbol{\rho}^{[n+1]} = \exp\Big([\ln\boldsymbol{\rho}^{[n]} - a^{[n+1]}\big(A^{\mathrm{T}}\hat{\boldsymbol{\Pi}}^{[n+1]} - \boldsymbol{\gamma}\big)]^{\mathcal{R}}\Big), \tag{22}$$

in order to adjust the amount of offered traffic $\boldsymbol{\rho}$ so as to achieve target carried traffic levels $\boldsymbol{\gamma}$. In practice, network operators usually have limited control over the amount of offered traffic, but they can typically adjust route selections fairly easily so as to achieve target blocking levels for a given offered traffic volume. Variations of the above algorithm can be used in such scenarios but go beyond the scope of the present paper.

In related work, Jiang and Walrand [17], [18] present an algorithm for achieving target throughputs in wireless CSMA networks. Their model can be interpreted as a special case of a loss network with unit link capacities. Their algorithm and convergence proof are therefore special cases of Theorem 2.

IV. Convergence proof

We will now prove Theorem 1. In §IV-A, we first explain our notion of convergence and then derive conditions on the error bias and zero-mean noise so that convergence is guaranteed. In §IV-B, we show that under the assumptions of Theorem 1, the error bias and zero-mean noise indeed satisfy the conditions derived in §IV-A.

A. Conditions for convergence

Theorem 1 states that $\boldsymbol{R}^{[n]}$ converges to $\boldsymbol{r}^{\mathrm{opt}}$ with probability one. In order to prove that, we will establish that the following two properties hold for arbitrary $\delta, \varepsilon > 0$. As our first property, we want that $\boldsymbol{R}^{[n]}$ comes close to $\boldsymbol{r}^{\mathrm{opt}}$ infinitely often. We make this precise by requiring that for any $\delta > 0$, the set $\mathcal{H}_\delta = \{\boldsymbol{r} \in \mathbb{R}^d | u(\boldsymbol{r}) \le u(\boldsymbol{r}^{\mathrm{opt}}) + \delta/2\}$ is recurrent for $\{\boldsymbol{R}^{[n]}\}_{n\in\mathbb{N}}$. As our second property, we want that once $\boldsymbol{R}^{[n]}$ comes close to $\boldsymbol{r}^{\mathrm{opt}}$, it stays close to $\boldsymbol{r}^{\mathrm{opt}}$ for all future iterations. Mathematically, we require that there exists an $m \in \mathbb{N}$ large enough so that $\|\boldsymbol{R}^{[n]} - \boldsymbol{r}^{\mathrm{opt}}\|_2^2 \le \|\boldsymbol{R}^{[m]} - \boldsymbol{r}^{\mathrm{opt}}\|_2^2 + \varepsilon$ for all $n \ge m$, which we will call capture of $\boldsymbol{R}^{[n]}$.

We shall relate both recurrence and capture to the error bias and zero-mean noise, defined as $\boldsymbol{B}^{[n]} = \mathbb{E}[\hat{\boldsymbol{G}}^{[n]}|\mathcal{F}^{[n-1]}] - \boldsymbol{G}^{[n]}$ and $\boldsymbol{E}^{[n]} = \hat{\boldsymbol{G}}^{[n]} - \mathbb{E}[\hat{\boldsymbol{G}}^{[n]}|\mathcal{F}^{[n-1]}]$, respectively. Here, $\mathcal{F}^{[n-1]}$ denotes the σ-field generated by the random vectors $\boldsymbol{Z}^{[0]}, \boldsymbol{Z}^{[1]}, ..., \boldsymbol{Z}^{[n-1]}$, where $\boldsymbol{Z}^{[0]} = (\boldsymbol{R}^{[0]}, X(0))^{\mathrm{T}}$ and $\boldsymbol{Z}^{[n]} = (\hat{\boldsymbol{G}}^{[n]}, \boldsymbol{R}^{[n]}, X(t^{[n]}))^{\mathrm{T}}$ for $n \ge 1$.

1) Recurrence: We begin with deriving conditions under which the set $\mathcal{H}_\delta$ is recurrent for $\{\boldsymbol{R}^{[n]}\}_{n\in\mathbb{N}}$, using the following result.

Lemma 3 ([14], p. 115)**.** *Let $\{\boldsymbol{R}^{[n]}\}_n$ be an $\mathbb{R}^d$-valued stochastic process, not necessarily a Markov process. Let $\{\mathcal{F}^{[n]}\}$ be a sequence of nondecreasing σ-algebras, with $\mathcal{F}^{[n]}$ measuring at least $\{\boldsymbol{R}^{[i]} | i \le n\}$. Assume that $a^{[n+1]}$ are positive $\mathcal{F}^{[n]}$-measurable random variables tending to zero with probability one and $\sum_n a^{[n]} = \infty$ with probability one. Let $V(\boldsymbol{r}) \ge 0$ and suppose that there are $\delta > 0$ and compact $\mathcal{H}_\delta \subset \mathbb{R}^d$ such that for all large n and all $\boldsymbol{r} \notin \mathcal{H}_\delta$,*

$$\mathbb{E}[V(\boldsymbol{R}^{[n+1]})|\mathcal{F}^{[n]}] - V(\boldsymbol{R}^{[n]}) \le -a^{[n+1]}\delta < 0. \tag{23}$$

Then the set $\mathcal{H}_\delta$ is recurrent for $\{\boldsymbol{R}^{[n]}\}_{n\ge 0}$ in the sense that $\boldsymbol{R}^{[n]} \in \mathcal{H}_\delta$ for infinitely many n with probability one.

Before we can apply Lemma 3, we need to identify a suitable function $V(\boldsymbol{R}^{[n+1]})$. The choice $D(\boldsymbol{R}^{[n+1]}) = \|\boldsymbol{R}^{[n+1]} - \boldsymbol{r}^{\text{opt}}\|_2^2$ comes to mind as a candidate, and we will therefore investigate (23) for $D(\boldsymbol{R}^{[n+1]})$. We will need the following result, the proof of which is relegated to §C.

Lemma 4. *For $x, y \in \mathbb{R}$ and $\mathcal{R} = [\mathcal{R}^{\min}, \mathcal{R}^{\max}] \subset \mathbb{R}$, $|[x]_{\mathcal{R}} - [y]_{\mathcal{R}}| \leq |x - y|$.*

Combining (11) and Lemma 4 gives

$$\begin{aligned} D(\boldsymbol{R}^{[n+1]}) &\leq \sum_{i=1}^{d} \big|R_i^{[n]} - a^{[n+1]}\hat{G}_i^{[n+1]} - r_i^{\text{opt}}\big|^2 \\ &= \sum_{i=1}^{d} \big|R_i^{[n]} - r_i^{\text{opt}}\big|^2 + (a^{[n+1]})^2 \sum_{i=1}^{d} \big|\hat{G}_i^{[n+1]}\big|^2 \\ &\quad - 2a^{[n+1]} \sum_{i=1}^{d} \hat{G}_i^{[n+1]}(R_i^{[n]} - r_i^{\text{opt}}). \end{aligned} \tag{24}$$

Substituting $\hat{\boldsymbol{G}}^{[n]} = \boldsymbol{G}^{[n]} + \boldsymbol{B}^{[n]} + \boldsymbol{E}^{[n]}$ into the last term, we conclude that

$$\begin{aligned} D(\boldsymbol{R}^{[n+1]}) &\leq D(\boldsymbol{R}^{[n]}) + (a^{[n+1]})^2 \|\hat{\boldsymbol{G}}^{[n+1]}\|_2^2 \\ &\quad - 2a^{[n+1]}(\boldsymbol{G}^{[n+1]} + \boldsymbol{B}^{[n+1]} + \boldsymbol{E}^{[n+1]})^{\mathrm{T}}(\boldsymbol{R}^{[n]} - \boldsymbol{r}^{\text{opt}}). \end{aligned} \tag{25}$$

Before we take the conditional expectation that results in a form similar to (23), recall that $u(\boldsymbol{r})$ is convex in $\boldsymbol{r}$. We therefore have that ([24], p. 69)

$$\begin{aligned} &\boldsymbol{G}^{[n+1]\mathrm{T}}(\boldsymbol{r}^{\text{opt}} - \boldsymbol{R}^{[n]}) = \boldsymbol{g}(\boldsymbol{\pi}(\boldsymbol{R}^{[n]}), \boldsymbol{R}^{[n]})^{\mathrm{T}}(\boldsymbol{r}^{\text{opt}} - \boldsymbol{R}^{[n]}) \\ &= \nabla_{\boldsymbol{r}} u(\boldsymbol{R}^{[n]})^{\mathrm{T}}(\boldsymbol{r}^{\text{opt}} - \boldsymbol{R}^{[n]}) \leq u(\boldsymbol{r}^{\text{opt}}) - u(\boldsymbol{R}^{[n]}). \end{aligned} \tag{26}$$

It follows that if $\boldsymbol{R}^{[n]} \notin \mathcal{H}_\delta$, then $\boldsymbol{G}^{[n+1]\mathrm{T}}(\boldsymbol{r}^{\text{opt}} - \boldsymbol{R}^{[n]}) < -\delta/2$. This gives in combination with (25) a term $-\delta a^{[n+1]}$, which we need for (23). We now note that $\mathbb{E}[\boldsymbol{E}^{[n+1]\mathrm{T}}(\boldsymbol{R}^{[n]} - \boldsymbol{r}^{\text{opt}})|\mathcal{F}^{[n]}] = 0$, so that for $\boldsymbol{R}^{[n]} \notin \mathcal{H}_\delta$,

$$\mathbb{E}[D(\boldsymbol{R}^{[n+1]})|\mathcal{F}^{[n]}] - D(\boldsymbol{R}^{[n]}) < -\delta a^{[n+1]} + Y^{[n+1]}, \tag{27}$$

where

$$\begin{aligned} Y^{[n+1]} =& (a^{[n+1]})^2 \mathbb{E}[\|\hat{\boldsymbol{G}}^{[n+1]}\|_2^2 | \mathcal{F}^{[n]}] \\ &+ 2a^{[n+1]} \big|\mathbb{E}[\boldsymbol{B}^{[n+1]\mathrm{T}}(\boldsymbol{R}^{[n]} - \boldsymbol{r}^{\text{opt}})|\mathcal{F}^{[n]}]\big|. \end{aligned} \tag{28}$$

The upper bound in (27) is not yet of the form of the right-hand side in (23). This implies that $D(\boldsymbol{R}^{[n+1]})$ by itself is not an appropriate candidate for $V(\boldsymbol{R}^{[n+1]})$. However, we can modify it slightly so that it does satisfy (23). For this, define $\Delta^{[n]} = \mathbb{E}[\sum_{i=n+1}^{\infty} Y^{[i]}|\mathcal{F}^{[n]}]$ and consider $V(\boldsymbol{R}^{[n+1]}) = D(\boldsymbol{R}^{[n+1]}) + \Delta^{[n+1]}$ instead. The difference $\mathbb{E}[\Delta^{[n+1]}|\mathcal{F}^{[n]}] - \Delta^{[n]}$ is well-defined if $\sum_{i=1}^{\infty} Y^{[i]} < \infty$ with probability one and is then equal to

$$\mathbb{E}[\mathbb{E}[\sum_{i=n+2}^{\infty} Y^{[i]}|\mathcal{F}^{[n+1]}] - \sum_{i=n+1}^{\infty} Y^{[i]}|\mathcal{F}^{[n]}] = -Y^{[n+1]}. \tag{29}$$

We conclude that

$$\begin{aligned} &\mathbb{E}[V(\boldsymbol{R}^{[n+1]})|\mathcal{F}^{[n]}] - V(\boldsymbol{R}^{[n]}) \\ &= \mathbb{E}[D(\boldsymbol{R}^{[n+1]})|\mathcal{F}^{[n]}] - D(\boldsymbol{R}^{[n]}) + \mathbb{E}[\Delta^{[n+1]}|\mathcal{F}^{[n]}] - \Delta^{[n]} \\ &= \mathbb{E}[D(\boldsymbol{R}^{[n+1]})|\mathcal{F}^{[n]}] - D(\boldsymbol{R}^{[n]}) - Y^{[n+1]} \leq -\delta a^{[n+1]}. \end{aligned} \tag{30}$$

The upper bound in (30) is of the form of (23), meaning that we are almost ready to apply Lemma 3. What remains is to check whether

$$\begin{aligned} &\sum_{n=1}^{\infty} Y^{[n]} = \sum_{n=1}^{\infty} (a^{[n]})^2 \mathbb{E}[\|\hat{\boldsymbol{G}}^{[n]}\|_2^2 | \mathcal{F}^{[n-1]}] \\ &+ 2\sum_{n=1}^{\infty} a^{[n]} \big|\mathbb{E}[\boldsymbol{B}^{[n]\mathrm{T}}(\boldsymbol{R}^{[n-1]} - \boldsymbol{r}^{\text{opt}})|\mathcal{F}^{[n-1]}]\big| < \infty \end{aligned} \tag{31}$$

with probability one. Since $\sum_{n=1}^{\infty}(a^{[n]})^2 < \infty$ and $\|\hat{\boldsymbol{G}}^{[n]}\|_2 \leq c_g$ by assumption, the first term is finite. Verifying that the second term is finite with probability one is much harder because it involves regularity conditions on $\boldsymbol{g}(\boldsymbol{\pi}(\boldsymbol{r}), \boldsymbol{r})$ and finiteness of mixing times. This can in fact be shown as stated in the next lemma, proved in §IV-B1.

Lemma 5. *Under the assumptions of Theorem 1, the sum $\sum_{n=1}^{\infty} a^{[n]} \big|\mathbb{E}[\boldsymbol{B}^{[n]\mathrm{T}}(\boldsymbol{R}^{[n-1]} - \boldsymbol{r}^{\text{opt}})|\mathcal{F}^{[n-1]}]\big|$ is finite with probability one.*

2) Capture: Having derived conditions under which $\mathcal{H}_\delta$ is recurrent, we turn our attention to deriving conditions under which capture occurs. Recall that capture means that there must exist an $m \in \mathbb{N}$ large enough so that $D(\boldsymbol{R}^{[n]}) \leq D(\boldsymbol{R}^{[m]}) + \varepsilon$ for all $n \geq m$ with probability one.

After applying (25) repeatedly and using the upper bound $\boldsymbol{G}^{[n]\mathrm{T}}(\boldsymbol{R}^{[n-1]} - \boldsymbol{r}^{\text{opt}}) \geq 0$, which follows by convexity of $u(\boldsymbol{r})$, we find that

$$\begin{aligned} D(\boldsymbol{R}^{[n]}) &\leq D(\boldsymbol{R}^{[m]}) + \sum_{j=m}^{n} (a^{[j+1]})^2 \|\hat{\boldsymbol{G}}^{[j+1]}\|_2^2 \\ &\quad - 2\sum_{j=m}^{n} a^{[j+1]}(\boldsymbol{B}^{[j+1]} + \boldsymbol{E}^{[j+1]})^{\mathrm{T}}(\boldsymbol{R}^{[j]} - \boldsymbol{r}^{\text{opt}}). \end{aligned} \tag{32}$$

We now need to show that each sum in the right-hand side of (32) becomes small for m sufficiently large. Because $\sum_{n=1}^{\infty}(a^{[n]})^2 < \infty$ and $\|\hat{\boldsymbol{G}}^{[n]}\|_2 \leq c_g$, it immediately follows that $\lim_{m\to\infty} \sum_{j=m}^{\infty}(a^{[n]})^2 \|\hat{\boldsymbol{G}}^{[n]}\|_2 = 0$. In turn, this implies that for any ε, there exists an $m_0 \in \mathbb{N}$ so that $\sum_{j=m}^{n}(a^{[n]})^2 \|\hat{\boldsymbol{G}}^{[n]}\|_2 \leq \varepsilon$ for all $n \geq m \geq m_0$. Verifying that the other two sums become small is substantially more difficult. This can be established using martingale arguments, as asserted in Lemma 6, the proof of which is postponed to §IV-B2.

Lemma 6. *Under the assumptions of Theorem 1, for any $\varepsilon > 0$, there exists $m_0 \in \mathbb{N}$ so that for any $n \geq m \geq m_0$*

(i) $\sum_{j=m}^{n} a^{[j]} \boldsymbol{B}^{[j]\mathrm{T}}(\boldsymbol{r}^{\text{opt}} - \boldsymbol{R}^{[j-1]}) \leq \varepsilon$ *and*

(ii) $\sum_{j=m}^{n} a^{[j]} \boldsymbol{E}^{[j]\mathrm{T}}(\boldsymbol{r}^{\text{opt}} - \boldsymbol{R}^{[j-1]}) \leq \varepsilon$

with probability one.

Our work thus far can also be used to prove that the gradient algorithm (7) converges. It is a special case of its stochastic counterpart (11), for which $\boldsymbol{B}^{[n]} = \boldsymbol{0}$, $\boldsymbol{E}^{[n]} = \boldsymbol{0}$, $\boldsymbol{G}^{[n]} = \hat{\boldsymbol{G}}^{[n]} = \boldsymbol{g}^{[n]}$ and $\boldsymbol{R}^{[n]} = \boldsymbol{r}^{[n]}$ for all $n \geq 0$. To prove that (7) converges, we apply (25) repeatedly and use that $\boldsymbol{g}^{[n]\mathrm{T}}(\boldsymbol{r}^{\mathrm{opt}} - \boldsymbol{r}^{[n-1]}) \leq u(\boldsymbol{r}^{\mathrm{opt}}) - u(\boldsymbol{r}^{[n-1]})$ for any $n \in \mathbb{N}$ by convexity of $u(\boldsymbol{r})$, so that

$$D(\boldsymbol{r}^{[n]}) \leq D(\boldsymbol{r}^{[0]}) + \sum_{j=0}^{n} (a^{[j+1]})^2 \|\boldsymbol{g}^{[j+1]}\|_2^2 - 2\sum_{j=0}^{n} a^{[j+1]}(u(\boldsymbol{r}^{[j]}) - u(\boldsymbol{r}^{\mathrm{opt}})). \tag{33}$$

Noting that $D(\boldsymbol{r}^{[n]}) \geq 0$ for all $n \in \mathbb{N}$ and $D(\boldsymbol{r}^{[0]}) \leq c_{\mathrm{r}}$ for some constant $c_{\mathrm{r}} < \infty$ since $\boldsymbol{r}^{[0]} \in \mathcal{R}$, we conclude that

$$2\sum_{j=0}^{n} a^{[j+1]}(u(\boldsymbol{r}^{[j]}) - u(\boldsymbol{r}^{\mathrm{opt}})) \leq c_{\mathrm{r}} + c_{\mathrm{g}}^2 \sum_{j=0}^{n} (a^{[j+1]})^2. \tag{34}$$

Since $\sum_{j=0}^{n} a^{[j+1]}(u(\boldsymbol{r}^{[j]}) - u(\boldsymbol{r}^{\mathrm{opt}})) \geq \min_{i=0,\ldots,n}\{u(\boldsymbol{r}^{[i]}) - u(\boldsymbol{r}^{\mathrm{opt}})\} \sum_{j=0}^{n} a^{[j+1]}$, we have the inequality

$$\min_{i=0,\ldots,n}\{u(\boldsymbol{r}^{[i]}) - u(\boldsymbol{r}^{\mathrm{opt}})\} \leq \frac{c_{\mathrm{r}} + c_{\mathrm{g}}^2 \sum_{j=0}^{n} (a^{[j+1]})^2}{2\sum_{j=0}^{n} a^{[j+1]}}, \tag{35}$$

which converges to 0 as $n \to \infty$.

From this little detour we see that it is much easier to establish convergence for (7) than for its stochastic counterpart (11). It is the error bias and zero-mean noise that make the convergence analysis of (11) so much harder.

B. Evaluating the conditions

We now provide the proofs of Lemma 5 and 6, which together prove Theorem 1. In our proofs, we choose to consider the error bias and zero-mean noise separately, which makes the analysis more tractable.

1) Error bias: We start by showing that the error bias satisfies the property claimed in Lemma 5 under the assumptions of Theorem 1. After substituting the definition of the error bias and using the triangle inequality, one finds that

$$\begin{aligned}
&\sum_{n=1}^{\infty} a^{[n]} \big|\mathbb{E}[\boldsymbol{B}^{[n]\mathrm{T}}(\boldsymbol{R}^{[n-1]} - \boldsymbol{r}^{\mathrm{opt}})|\mathcal{F}^{[n-1]}]\big| \\
&= \sum_{n=1}^{\infty} a^{[n]} \Big|\sum_{i=1}^{d} \mathbb{E}[B_i^{[n]}|\mathcal{F}^{[n-1]}](R_i^{[n-1]} - r_i^{\mathrm{opt}})\Big| \\
&\leq \sum_{n=1}^{\infty} a^{[n]} \sum_{i=1}^{d} (\mathcal{R}_i^{\max} - \mathcal{R}_i^{\min}) \big|\mathbb{E}[B_i^{[n]}|\mathcal{F}^{[n-1]}]\big| \\
&= \sum_{n=1}^{\infty} a^{[n]} \sum_{i=1}^{d} (\mathcal{R}_i^{\max} - \mathcal{R}_i^{\min}) \big|B_i^{[n]}\big|.
\end{aligned} \tag{36}$$

The inequality is a consequence of $\mathcal{R}$ being a hypercube. We have also used the fact that $\mathbb{E}[B_i^{[n]}|\mathcal{F}^{[n-1]}] = B_i^{[n]}$, which follows from the definition $G_i^{[n]} = g_i(\boldsymbol{\pi}(\boldsymbol{R}^{[n-1]}), \boldsymbol{R}^{[n-1]})$.

We now bound $\big|B_i^{[n]}\big|$ from above. After recalling that $B_i^{[n]} = \mathbb{E}[\hat{G}_i^{[n]}|\mathcal{F}^{[n-1]}] - G_i^{[n]}$ and using Jensen's inequality, we find that $\big|B_i^{[n]}\big|$ equals

$$\begin{aligned}
&\big|\mathbb{E}[g_i(\hat{\boldsymbol{\Pi}}^{[n]}, \boldsymbol{R}^{[n-1]})|\mathcal{F}^{[n-1]}] - g_i(\boldsymbol{\pi}(\boldsymbol{R}^{[n-1]}), \boldsymbol{R}^{[n-1]})\big| \\
&= \big|\mathbb{E}[g_i(\hat{\boldsymbol{\Pi}}^{[n]}, \boldsymbol{R}^{[n-1]}) - g_i(\boldsymbol{\pi}(\boldsymbol{R}^{[n-1]}), \boldsymbol{R}^{[n-1]})|\mathcal{F}^{[n-1]}]\big| \\
&\leq \mathbb{E}[\big|g_i(\hat{\boldsymbol{\Pi}}^{[n]}, \boldsymbol{R}^{[n-1]}) - g_i(\boldsymbol{\pi}(\boldsymbol{R}^{[n-1]}), \boldsymbol{R}^{[n-1]})\big||\mathcal{F}^{[n-1]}].
\end{aligned}$$

Recalling condition (14) gives

$$\big|B_i^{[n]}\big| \leq \frac{c_{\mathrm{l}}}{2} \sum_{x \in \Omega} \mathbb{E}[\big|\hat{\Pi}_x^{[n]} - \pi_x(\boldsymbol{R}^{[n-1]})\big||\mathcal{F}^{[n-1]}]. \tag{37}$$

Finiteness of (36) can now be proven by constructing an upper bound for (37). We can obtain such a bound using the following lemma, proved in Appendix D.

Lemma 7. *There exist $c_{\mathrm{e}}, \kappa \in [0, \infty)$ such that for $e^{[n]} \in [0, 1]$ and $x \in \Omega$,*

$$\mathbb{P}[\big|\hat{\Pi}_x^{[n]} - \pi_x(\boldsymbol{R}^{[n-1]})\big| \geq e^{[n]}] \leq c_{\mathrm{e}} \exp\Big(-\frac{(e^{[n]})^2}{4|\Omega|^2 \kappa f^{[n]}}\Big).$$

Define $\Phi_x^{[n]} = \big|\hat{\Pi}_x^{[n]} - \pi_x(\boldsymbol{R}^{[n-1]})\big|$ and let $\epsilon^{[n]} \in [0, 1]$. Using (37) and then Lemma 7 yields

$$\begin{aligned}
\big|B_i^{[n]}\big| &\leq \frac{c_{\mathrm{l}}}{2} \sum_{x \in \Omega} \mathbb{E}[\Phi_x^{[n]}|\mathcal{F}^{[n-1]}] \\
&= \frac{c_{\mathrm{l}}}{2} \sum_{x \in \Omega} \Big(\mathbb{P}[\Phi_x^{[n]} < e^{[n]}]\mathbb{E}[\Phi_x^{[n]}|\mathcal{F}^{[n-1]}, \Phi_x^{[n]} < e^{[n]}] \\
&\quad + \mathbb{P}[\Phi_x^{[n]} \geq e^{[n]}]\mathbb{E}[\Phi_x^{[n]}|\mathcal{F}^{[n-1]}, \Phi_x^{[n]} \geq e^{[n]}]\Big) \\
&\leq \frac{c_{\mathrm{l}}}{2} \sum_{x \in \Omega} \Big(e^{[n]} + (1 - e^{[n]})\mathbb{P}[\Phi_x^{[n]} \geq e^{[n]}]\Big) \\
&\leq \frac{c_{\mathrm{l}}|\Omega|}{2} \max\{1, c_{\mathrm{e}}\}\Big(e^{[n]} + \exp\Big(-\frac{(e^{[n]})^2}{4|\Omega|^2 \kappa f^{[n]}}\Big)\Big).
\end{aligned} \tag{38}$$

After bounding (36) from above using (38), it follows from (13) that

$$\sum_{n=1}^{\infty} a^{[n]} \big|\mathbb{E}[\boldsymbol{B}^{[n]\mathrm{T}}(\boldsymbol{R}^{[n-1]} - \boldsymbol{r}^{\mathrm{opt}})|\mathcal{F}^{[n-1]}]\big| < \infty, \tag{39}$$

which completes the proof of Lemma 5.

We now show that the error bias satisfies assertion (i) in Lemma 6 under the assumptions of Theorem 1. Similar to the derivation of (36),

$$\begin{aligned}
&\sum_{n=1}^{\infty} a^{[n]} \big|\boldsymbol{B}^{[n]\mathrm{T}}(\boldsymbol{R}^{[n-1]} - \boldsymbol{r}^{\mathrm{opt}})\big| \\
&\leq \sum_{n=1}^{\infty} a^{[n]} \sum_{i=1}^{d} (\mathcal{R}_i^{\max} - \mathcal{R}_i^{\min}) \big|B_i^{[n]}\big|.
\end{aligned} \tag{40}$$

Combining (40), (38) and (13), we conclude that with probability one,

$$\sum_{n=1}^{\infty} a^{[n]} \big|\boldsymbol{B}^{[n]\mathrm{T}}(\boldsymbol{R}^{[n-1]} - \boldsymbol{r}^{\mathrm{opt}})\big| < \infty, \tag{41}$$

so that $\lim_{m\to\infty}\sum_{j=m}^{\infty} a^{[n]}\boldsymbol{B}^{[n]\mathrm{T}}(\boldsymbol{r}^{\mathrm{opt}}-\boldsymbol{R}^{[n-1]})=0$ with probability one. This implies that there exists an $m_0\in\mathbb{N}$ so that for all $n\geq m\geq m_0$, $\sum_{j=m}^{n} a^{[j]}\boldsymbol{B}^{[j]\mathrm{T}}(\boldsymbol{r}^{\mathrm{opt}}-\boldsymbol{R}^{[j-1]})\leq\varepsilon$ with probability one. The error bias thus satisfies assertion (i) in Lemma 6. All that remains is to show that the zero-mean noise satisfies Lemma 6(ii).

2) Zero-mean noise: We use a martingale argument to show that assertion (ii) in Lemma 6 holds. We start our argument by defining $M^{[n]}=\sum_{j=1}^{n} a^{[j]}\boldsymbol{E}^{[j]\mathrm{T}}(\boldsymbol{R}^{[j-1]}-\boldsymbol{r}^{\mathrm{opt}})$. See Appendix E for a proof of the following result.

Lemma 8. *$M^{[n]}$ is a martingale.*

We will use a martingale convergence theorem [25] to show that for $n\geq m$ both sufficiently large, $M^{[n]}-M^{[m-1]}\leq\varepsilon$ with probability one.

Theorem 3. *If $\{M^{[n]}\}$ is a martingale for which there exists a constant $c_{\mathrm{m}}<\infty$ so that $\mathbb{E}[(M^{[n]})^2]\leq c_{\mathrm{m}}$ for all $n\geq 0$, then there exists a random variable M^{opt} with $\mathbb{E}[(M^{\mathrm{opt}})^2]\leq c_{\mathrm{m}}$ such that $M^{[n]}\to M^{\mathrm{opt}}$ with probability one as $n\to\infty$. Moreover, $\mathbb{E}[|M^{[n]}-M^{\mathrm{opt}}|^2]^{\frac{1}{2}}\to 0$ as $n\to\infty$.*

Before we can apply Theorem 3, we need to show existence of a $c_{\mathrm{m}}\in\mathbb{R}$ such that $\mathbb{E}[(M^{[n]})^2]\leq c_{\mathrm{m}}$ for all $n\in\mathbb{N}$. To show this, expand

$$\sup_n \mathbb{E}[(M^{[n]})^2]=\sup_n\Big\{\sum_{j=1}^{n}(a^{[j]})^2\mathbb{E}[(\boldsymbol{E}^{[j]\mathrm{T}}(\boldsymbol{R}^{[j-1]}-\boldsymbol{r}^{\mathrm{opt}}))^2]$$
$$+\sum_{j\neq k} a^{[j]}a^{[k]}\mathbb{E}[\boldsymbol{E}^{[j]\mathrm{T}}(\boldsymbol{R}^{[j-1]}-\boldsymbol{r}^{\mathrm{opt}})\boldsymbol{E}^{[k]\mathrm{T}}(\boldsymbol{R}^{[k-1]}-\boldsymbol{r}^{\mathrm{opt}})]\Big\},$$

and then consider any one of the cross terms with $k<j$. By the tower property,

$$\mathbb{E}[\boldsymbol{E}^{[j]\mathrm{T}}(\boldsymbol{R}^{[j-1]}-\boldsymbol{r}^{\mathrm{opt}})\boldsymbol{E}^{[k]\mathrm{T}}(\boldsymbol{R}^{[k-1]}-\boldsymbol{r}^{\mathrm{opt}})]$$
$$=\mathbb{E}[\mathbb{E}[\boldsymbol{E}^{[j]\mathrm{T}}(\boldsymbol{R}^{[j-1]}-\boldsymbol{r}^{\mathrm{opt}})\boldsymbol{E}^{[k]\mathrm{T}}(\boldsymbol{R}^{[k-1]}-\boldsymbol{r}^{\mathrm{opt}})|\mathcal{F}^{[j-1]}]]$$
$$=\mathbb{E}[\boldsymbol{E}^{[k]\mathrm{T}}(\boldsymbol{R}^{[k-1]}-\boldsymbol{r}^{\mathrm{opt}})\mathbb{E}[\sum_{i=1}^{d}E_i^{[j]}(R_i^{[j-1]}-r_i^{\mathrm{opt}})|\mathcal{F}^{[j-1]}]]$$
$$=\mathbb{E}[\boldsymbol{E}^{[k]\mathrm{T}}(\boldsymbol{R}^{[k-1]}-\boldsymbol{r}^{\mathrm{opt}})\sum_{i=1}^{d}\mathbb{E}[E_i^{[j]}|\mathcal{F}^{[j-1]}](R_i^{[j-1]}-r_i^{\mathrm{opt}})],$$

and because $\mathbb{E}[E_i^{[j]}|\mathcal{F}^{[j-1]}]=0$, all cross terms are equal to 0. Because the summands are positive, we can give an upper bound by summing over all terms, so that

$$\sup_n \mathbb{E}[(M^{[n]})^2]\leq\sum_{j=1}^{\infty}(a^{[j]})^2\mathbb{E}[(\boldsymbol{E}^{[j]\mathrm{T}}(\boldsymbol{R}^{[j-1]}-\boldsymbol{r}^{\mathrm{opt}}))^2]$$
$$=\sum_{j=1}^{\infty}(a^{[j]})^2\mathbb{E}[\big(\sum_{i=1}^{d}E_i^{[j]}(R_i^{[j-1]}-r_i^{\mathrm{opt}})\big)^2]. \tag{42}$$

Using the triangle inequality, we find that

$$\sup_n \mathbb{E}[(M^{[n]})^2]\leq\sum_{j=1}^{\infty}(a^{[j]})^2\mathbb{E}[\big(\sum_{i=1}^{d}|E_i^{[j]}||R_i^{[j-1]}-r_i^{\mathrm{opt}}|\big)^2].$$

Now note that $\sum_{i=1}^{d}\big|E_i^{[j]}\big|=\|\boldsymbol{E}^{[j]}\|_1$, write

$$\|\boldsymbol{E}^{[j]}\|_1\leq\mathbb{E}[\|\hat{\boldsymbol{G}}^{[j]}\|_1|\mathcal{F}^{[j-1]}]+\|\hat{\boldsymbol{G}}^{[j]}\|_1$$
$$\leq\sqrt{d}\mathbb{E}[\|\hat{\boldsymbol{G}}^{[j]}\|_2|\mathcal{F}^{[j-1]}]+\sqrt{d}\|\hat{\boldsymbol{G}}^{[j]}\|_2\leq 2c_{\mathrm{g}}\sqrt{d} \tag{43}$$

and recall that $\mathcal{R}$ is a hypercube. We conclude that

$$\sup_n \mathbb{E}[(M^{[n]})^2]\leq 4c_{\mathrm{g}}^2 d\max_{i=1,\dots,d}\{(\mathcal{R}_i^{\max}-\mathcal{R}_i^{\min})^2\}\sum_{j=1}^{\infty}(a^{[j]})^2.$$

The right-hand side is finite by condition (12), and we see that there indeed exists a coefficient c_{m} so that $\mathbb{E}[(M^{[n]})^2]\leq c_{\mathrm{m}}$ for all $n\in\mathbb{N}$. We now apply Theorem 3 and conclude that as $n\geq m\to\infty$,

$$\mathbb{E}[|M^{[n]}-M^{[m-1]}|^2]^{\frac{1}{2}}\leq\mathbb{E}[|M^{[n]}-M^{\mathrm{opt}}|^2]^{\frac{1}{2}}$$
$$+\mathbb{E}[|M^{[m-1]}-M^{\mathrm{opt}}|^2]^{\frac{1}{2}}\to 0. \tag{44}$$

This result enables us to use Doob's maximal inequality [25], as reproduced in the lemma below, in order to conclude that Lemma 6(ii) holds.

Lemma 9. *If $\{M^{[n]}\}_{n\geq 0}$ is a nonnegative submartingale and $\lambda>0$, then*

$$\lambda\mathbb{P}[\sup_{m\leq n}M^{[m]}\geq\lambda]\leq\mathbb{E}[M^{[n]}\mathbb{1}[\sup_{m\leq n}M^{[m]}\geq\lambda]]\leq\mathbb{E}[M^{[n]}].$$

Fix $m\in\mathbb{N}$ and define $W^{[n]}=M^{[n+m-1]}-M^{[m-1]}$ for $n\in\mathbb{N}$. $|W^{[n]}|$ is a submartingale by Jensen's inequality with respect to the sequence $\mathcal{F}^{[m-1]},\mathcal{F}^{[m]},\mathcal{F}^{[m+1]},\dots$, since $\mathbb{E}[|W^{[n+1]}||\mathcal{F}^{[n+m-1]}]\geq\mathbb{E}[W^{[n+1]}|\mathcal{F}^{[n+m-1]}]=W^{[n]}$. Applying Lemma 9 to $|W^{[n]}|$, we find that

$$\mathbb{P}[\sup_{0\leq t\leq n}|M^{[t+m-1]}-M^{[m-1]}|\geq\lambda]$$
$$\leq\frac{\mathbb{E}[|M^{[n+m-1]}-M^{[m-1]}|]}{\lambda}$$
$$\leq\frac{\mathbb{E}[|M^{[n+m-1]}-M^{\mathrm{opt}}|]+\mathbb{E}[|M^{\mathrm{opt}}-M^{[m-1]}|]}{\lambda}$$
$$\leq\frac{\mathbb{E}[|M^{[n+m-1]}-M^{\mathrm{opt}}|^2]^{\frac{1}{2}}+\mathbb{E}[|M^{\mathrm{opt}}-M^{[m-1]}|^2]^{\frac{1}{2}}}{\lambda}$$

for any $\lambda\in(0,\infty)$ and $m\in\mathbb{N}$. This upper bound converges to 0 as $n,m\to\infty$, implying that there exists an $m_0\in\mathbb{N}$, such that for all $n\geq m\geq m_0$, $M^{[n]}-M^{[m-1]}\leq\varepsilon$ with probability one.

Having established Lemma 5 and 6, the proof of Theorem 1 is now completed.

V. Conclusions

We have developed an online gradient algorithm for finding parameter values that optimize the performance of reversible Markov processes with product-form distributions. As a key feature, the approach avoids the computational complexity of calculating the gradient in terms of the stationary probabilities and instead relies on measuring empirical time fractions of the various states so as to obtain estimates for the gradient. While the impact of the induced measurement noise can be handled without too much trouble, the bias in the estimates

presents a trickier issue. In order to exploit mixing time results to deal with the bias, we focussed on reversible processes. We expect however that convergence can be established under milder conditions.

For fast convergence, the algorithm needs to strike a balance between the step sizes and the lengths of observation periods, which is a consequence of the existence of two time scales - one being the mixing time of the underlying stochastic process and the other being the iteration sequence generated by the algorithm. Intuitively, the step sizes should not have become too small by the time that the observation periods have become larger than the mixing time. The convergence of the algorithm would otherwise slow down drastically. A challenging issue for further research is to gain a more detailed understanding of the effect of step sizes and the role of mixing times in relation to the convergence speed. A related direction is to explore the trade-off between accuracy in static scenarios and responsiveness in dynamic environments, which relates to convergence in distribution for non-vanishing step sizes as opposed to the almost-sure convergence for decreasing step sizes as considered here.

ACKNOWLEDGMENTS

This research was financially supported by The Netherlands Organization for Scientific Research (NWO) in the framework of the TOP-GO program and by an ERC Starting Grant.

REFERENCES

[1] P. Bremaud, *Markov Chains – Gibbs Fields, Monte Carlo Simulation and Queues*. Springer, 1999.
[2] F. Kelly, "Stochastic models of computer communication systems," *Journal of the Royal Statistical Society, Series B*, vol. 47, no. 3, pp. 379–395, 1985.
[3] T. Liggett, *Interacting Particle Systems*. Springer, 1985.
[4] D. Bertsekas, *Dynamic Programming and Optimal Control*. Athena Press, 1996.
[5] M. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley-Interscience, 1994.
[6] K. Jung, Y. Lu, D. Shah, M. Sharma, and M. Squillante, "Revisiting stochastic loss networks: structures and algorithms," *ACM SIGMETRICS*, pp. 407–418, 2008.
[7] F. Kelly, "Loss networks," *Annals of Applied Probability*, vol. 1, no. 3, pp. 319–378, 1991.
[8] F. Baskett, K. Chandy, R. Muntz, and F. Palacios, "Open, closed and mixed networks of queues with different classes of customers," *Journal of the Association for Computing Machinery*, vol. 22, pp. 248–260, 1979.
[9] F. Kelly, *Reversibility and Stochastic Networks*. Wiley, Chichester, 1979.
[10] R. Boorstyn, A. Kershenbaum, B. Maglaris, and V. Sahin, "Throughput analysis in multihop CSMA packet radio networks," *IEEE Transactions on Communications*, vol. 35, pp. 267–274, 1987.
[11] X. Wang and K. Kar, "Throughput modeling and fairness issues in CSMA/CA based ad-hoc networks," *IEEE Infocom*, 2005.
[12] G. Louth, M. Mitzenmacher, and F. Kelly, "Computational complexity of loss networks," *Theoretical Computer Science*, vol. 125, no. 1, pp. 45–59, 1994.
[13] V. Borkar, *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press, 2008.
[14] H. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*. Springer, 2003.
[15] P. Diaconis and D. Stroock, "Geometric bounds for eigenvalues of Markov chains," *Annals of Applied Probability*, vol. 1, no. 1, pp. 36–61, 1991.
[16] D. Levin, Y. Peres, and E. Wilmer, *Markov Chains and Mixing Times*. American Mathematical Society, 2008.
[17] L. Jiang and J. Walrand, "Convergence and stability of a distributed CSMA algorithm for maximal network throughput," University of California, Berkeley, http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-43.html, Tech. Rep., 2009.
[18] ——, "A distributed CSMA algorithm for throughput and utility maximization in wireless networks," *IEEE/ACM Transactions on Networking*, vol. 18, pp. 960–972, 2010.
[19] P. Marbach and J. Tsitsiklis, "Simulation-based optimization of Markov reward processes," *IEEE Transactions on Automatic Control*, vol. 46, no. 2, pp. 191–209, 2001.
[20] ——, "Approximate gradient methods in policy-space optimization of Markov reward processes," *Discrete Event Dynamical Systems*, vol. 13, pp. 111–148, 2003.
[21] X. Cao, *Stochastic Learning and Optimization: A Sensitivity-Based Approach*. Springer, 2007.
[22] Z. Liu and P. Nain, "Sensitivity results in open, closed and mixed product form queueing networks," *Performance Evaluation*, vol. 13, no. 4, pp. 237–251, 1991.
[23] A. Harel, "Convexity properties of the Erlang loss formula," *Operations Research*, vol. 38, no. 3, pp. 499–505, 1990.
[24] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[25] J. Steele, *Stochastic Calculus and Financial Applications*. Springer, 2001.
[26] P. Cattiaux and A. Guillin, "Deviation bounds for additive functionals of Markov processes," *ESAIM: Probability and Statistics*, vol. 12, pp. 12–29, 2006.

APPENDIX

A. Proof of Lemma 1

Define $B_{\boldsymbol{\mu}} = \mu_s$ and $L_{\boldsymbol{\mu}} = \sum_{x=1}^{s} x\mu_x$ for all $\boldsymbol{\mu} \in [0,1]^{|\Omega|}$ for which $\mathbf{1}_{|\Omega|}{}^{\mathrm{T}}\boldsymbol{\mu} = 1$. By definition of $g(\boldsymbol{\mu}, r)$, $\|g(\boldsymbol{\mu}, r)\|_2 \leq |B_{\boldsymbol{\mu}}(L_{\boldsymbol{\mu}} - s)|/r + |c'(r)| < \infty$. The first term is finite because $r \geq \mathcal{R}^{\min} > 0$, $B_{\boldsymbol{\mu}} \leq 1$ and $L_{\boldsymbol{\mu}} \leq s < \infty$. The second term is finite by our assumption that $c'(r)$ is bounded for all $r \in \mathcal{R}$. This proves that condition (15) is met.

We now turn to condition (14). Write $|g(\boldsymbol{\mu}, r) - g(\boldsymbol{\nu}, r)| = |B_{\boldsymbol{\mu}}(L_{\boldsymbol{\mu}} - s) - B_{\boldsymbol{\nu}}(L_{\boldsymbol{\nu}} - s)|/r \leq |B_{\boldsymbol{\mu}}L_{\boldsymbol{\mu}} - sB_{\boldsymbol{\mu}} - B_{\boldsymbol{\nu}}L_{\boldsymbol{\nu}} + sB_{\boldsymbol{\nu}}|/\mathcal{R}^{\min} \leq (|B_{\boldsymbol{\mu}}L_{\boldsymbol{\mu}} - B_{\boldsymbol{\nu}}L_{\boldsymbol{\nu}}| + s|B_{\boldsymbol{\mu}} - B_{\boldsymbol{\nu}}|)/\mathcal{R}^{\min}$. We then conclude that $|B_{\boldsymbol{\mu}}L_{\boldsymbol{\mu}} - B_{\boldsymbol{\nu}}L_{\boldsymbol{\nu}}| = |B_{\boldsymbol{\mu}}L_{\boldsymbol{\mu}} - B_{\boldsymbol{\mu}}L_{\boldsymbol{\nu}} + B_{\boldsymbol{\mu}}L_{\boldsymbol{\nu}} - B_{\boldsymbol{\nu}}L_{\boldsymbol{\nu}}| \leq B_{\boldsymbol{\mu}}|L_{\boldsymbol{\mu}} - L_{\boldsymbol{\nu}}| + L_{\boldsymbol{\nu}}|B_{\boldsymbol{\mu}} - B_{\boldsymbol{\nu}}| \leq |L_{\boldsymbol{\mu}} - L_{\boldsymbol{\nu}}| + s|B_{\boldsymbol{\mu}} - B_{\boldsymbol{\nu}}|$, so that $|g(\boldsymbol{\mu}, r) - g(\boldsymbol{\nu}, r)| \leq (|L_{\boldsymbol{\mu}} - L_{\boldsymbol{\nu}}| + 2s|B_{\boldsymbol{\mu}} - B_{\boldsymbol{\nu}}|)/\mathcal{R}^{\min}$. Finally, by definition of $B_{\boldsymbol{\mu}}$, $|B_{\boldsymbol{\mu}} - B_{\boldsymbol{\nu}}| = |\mu_s - \nu_s| \leq 2||\boldsymbol{\mu} - \boldsymbol{\nu}||_{\mathrm{var}}$. Similarly for $L_{\boldsymbol{\mu}}$, $\left|L_{\boldsymbol{\mu}} - L_{\boldsymbol{\nu}}\right| \leq \sum_{x=1}^{s} x\left|\mu_x - \nu_x\right| \leq 2s||\boldsymbol{\mu} - \boldsymbol{\nu}||_{\mathrm{var}}$. Thus $|g(\boldsymbol{\mu}, r) - g(\boldsymbol{\nu}, r)| \leq 6s||\boldsymbol{\mu} - \boldsymbol{\nu}||_{\mathrm{var}}/\mathcal{R}^{\min}$, which concludes the proof after setting $c_{\mathrm{l}} = 6s/\mathcal{R}^{\min}$. □

B. Proof of Lemma 2

Substituting (3) into (2) gives

$$u(\boldsymbol{r}) = \ln \sum_{y \in \Omega} \exp\left(A\boldsymbol{r} + \boldsymbol{b}\right)_y - \sum_{x \in \Omega} \alpha_x (A\boldsymbol{r} + \boldsymbol{b})_x. \quad (45)$$

The function $v(\boldsymbol{s}) = \ln \sum_{y \in \Omega} \exp s_y - \sum_{x \in \Omega} \alpha_x s_x$ is convex on $\mathbb{R}^{|\Omega|}$ [24], p. 72. We see that $u(\boldsymbol{r})$ is a composition of a convex function with an affine mapping, i.e. $u(\boldsymbol{r}) = v(A\boldsymbol{r} + \boldsymbol{b})$, and such functions are convex [24], p. 79. □

C. Proof of Lemma 4

Define $l = \mathcal{R}^{\min}$ and $r = \mathcal{R}^{\max}$. If $x, y \in \mathcal{R}$, equality holds. Consider the case $x \notin \mathcal{R}, y \in \mathcal{R}$. If $x > r$, $|[x]_{\mathcal{R}} - [y]_{\mathcal{R}}| = |r - y| = r - y \leq x - y = |x - y|$. If $x < l$, $|[x]_{\mathcal{R}} - [y]_{\mathcal{R}}| =$

$|l-y| = y-l \le y-x = |x-y|$. Finally, consider the case $x, y \notin \mathcal{R}$. If $x, y > r$ or $x, y < l$, $|[x]_{\mathcal{R}} - [y]_{\mathcal{R}}| = 0 \le |x-y|$. If $x > r, y < l$, $|[x]_{\mathcal{R}} - [y]_{\mathcal{R}}| = |r-l| = r-l \le x-y = |x-y|$. The case $x < l, y > r$ follows from a similar argument. □

D. Proof of Lemma 7

Let $\mathrm{Var}_{\boldsymbol{\mu}}[f] = \frac{1}{2}\sum_{x,y\in\Omega}\big(f(x)-f(y)\big)^2\mu_x\mu_y$, $(f,g)_{\boldsymbol{\mu}} = \sum_{x\in\Omega} f(x)g(x)\mu_x$ and $\|\boldsymbol{\mu}\|_{2,\boldsymbol{\nu}} = (\sum_{x\in\Omega}\mu_x^2\nu_x)^{1/2}$.

Proposition 1 ([26], p. 2)**.** *On some Polish space Ω, let us consider a conservative (continuous-time) Markov process denoted by $\{X(t)\}_{t\ge 0}$ and with infinitesimal generator $\mathcal{L}$. Let $\boldsymbol{\mu}$ be a probability measure on Ω which is invariant and ergodic with respect to P_t.*

Assume that $\boldsymbol{\mu}$ satisfies the Poincaré inequality $\mathrm{Var}_{\boldsymbol{\mu}}[f] \le -\kappa(\mathcal{L}f, f)_{\boldsymbol{\mu}}$. Then for all θ such that $\sup|\theta| = 1$, all $0 < \epsilon \le 1$ and all $t > 0$, assuming that the initial distribution of X_s is ν,

$$\mathbb{P}\Big[\Big|\frac{1}{t}\int_0^t \theta(X(s))ds - \int\theta d\mu\Big| \ge \epsilon\Big] \le \Big\|\frac{d\nu}{d\mu}\Big\|_{2,\mu}\exp\Big(-\frac{t\epsilon^2}{8\kappa\mathrm{Var}_{\boldsymbol{\mu}}[\theta]}\Big). \tag{46}$$

Lemma 7 is a direct consequence of Proposition 1. Before we can use Proposition 1 to prove Lemma 7, however, we need to verify all of its assumptions. We will now verify these assumptions for continuous-time, reversible Markov processes with a product form solution. Our method is based on an approach for discrete-time Markov chains [15].

Define a graph $G = (V, E)$, where V denotes the vertex set in which each vertex corresponds to a state in Ω and E denotes the set of directed edges. An edge $e = (x, y)$ is in E if $\phi(e) = \pi_x Q_{x,y} = \pi_y Q_{y,x} > 0$. Here, Q denotes the generator matrix of $\{X(t)\}_{t\ge 0}$. For every pair of distinct vertices $x, y \in \Omega$, choose a path $\gamma_{x,y}$ (along the edges of G) from x to y. Paths may have repeated vertices but a given edge appears at most once in a given path. Let Γ denote the collection of paths (one for each ordered pair x, y). Irreducibility of $\{X(t)\}_{t\ge 0}$ guarantees that such paths exist. For $\gamma_{x,y} \in \Gamma$ define the path length by $\|\gamma_{x,y}\|_\phi = \sum_{e\in\gamma_{x,y}}(1/\phi(e))$. Also, let

$$\kappa = \max_e \sum_{\{\gamma_{x,y}\in\Gamma | e\in\gamma_{x,y}\}} \|\gamma_{x,y}\|_\phi \pi_x\pi_y \tag{47}$$

and $f(e) = f(y) - f(x)$ for $e = (x, y) \in E$. Then write

$$\mathrm{Var}_{\boldsymbol{\pi}}[f] = \frac{1}{2}\sum_{x,y\in\Omega}\Big(\sum_{e\in\gamma_{x,y}}\Big(\frac{\phi(e)}{\phi(e)}\Big)^{\frac{1}{2}} f(e)\Big)^2\pi_x\pi_y. \tag{48}$$

Use the Cauchy-Schwarz inequality $|\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}|^2 \le \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\cdot\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}$ to obtain

$$\begin{aligned}\mathrm{Var}_{\boldsymbol{\pi}}[f] &\le \frac{1}{2}\sum_{x,y\in\Omega}\pi_x\pi_y\Big(\sum_{e\in\gamma_{x,y}}\frac{1}{\phi(e)}\Big)\Big(\sum_{e\in\gamma_{x,y}}\phi(e)f(e)^2\Big)\\ &= \frac{1}{2}\sum_{x,y\in\Omega}\pi_x\pi_y\|\gamma_{x,y}\|_\phi\Big(\sum_{e\in\gamma_{x,y}}\phi(e)f(e)^2\Big)\\ &= \frac{1}{2}\sum_{e\in E}\phi(e)f(e)^2\sum_{\{\gamma_{x,y}\in\Gamma|e\in\gamma_{x,y}\}}\|\gamma_{x,y}\|_\phi\pi_x\pi_y.\end{aligned}$$

Use the definition of κ and the symmetry of $\phi(e)$ to write

$$\begin{aligned}\mathrm{Var}_{\boldsymbol{\pi}}[f] &\le \frac{\kappa}{2}\sum_{e\in E}\phi(e)f(e)^2\\ &= \frac{\kappa}{2}\sum_{x,y\in\Omega}\pi_y Q_{y,x}(f(y)^2 - f(y)f(x))\\ &\quad + \frac{\kappa}{2}\sum_{x,y\in\Omega}\pi_x Q_{x,y}(f(x)^2 - f(y)f(x))\\ &= \kappa\sum_{x,y\in\Omega}Q_{x,y}(f(x)-f(y))f(x)\pi_x\\ &= \kappa\sum_{x\in\Omega}\Big(\sum_{y\in\Omega}Q_{x,y}(f(x)-f(y))\Big)f(x)\pi_x.\end{aligned} \tag{49}$$

By definition of the infinitesimal generator $\mathcal{L}$, we find that

$$\begin{aligned}(\mathcal{L}f)(x) &= \lim_{t\to 0}\frac{1}{t}\Big(\sum_{y\in\Omega}\big(\mathrm{e}^{tQ}\big)_{x,y}f(y) - f(x)\Big)\\ &= \lim_{t\to 0}\frac{1}{t}\Big(\sum_{y\in\Omega}\big(I + tQ + \mathcal{O}(t^2)\big)_{x,y}f(y) - f(x)\Big)\\ &= \sum_{y\in\Omega}Q_{x,y}f(y) = \sum_{y\in\Omega\setminus\{x\}}Q_{x,y}f(y) + Q_{x,x}f(x)\\ &= \sum_{y\in\Omega\setminus\{x\}}Q_{x,y}f(y) - \sum_{y\in\Omega\setminus\{x\}}Q_{x,y}f(x)\\ &= \sum_{y\in\Omega}Q_{x,y}(f(y)-f(x)),\end{aligned} \tag{50}$$

after which one can conclude that $\mathrm{Var}_{\boldsymbol{\pi}}[f] \le -\kappa(\mathcal{L}f, f)_{\boldsymbol{\pi}}$. We also note that when choosing $\theta(X(t)) = \mathbb{1}[X(t) = z]$, we have that

$$\mathrm{Var}_{\boldsymbol{\pi}}[\theta] = \frac{1}{2}\sum_{x,y\in\Omega}\big(\mathbb{1}[x=z] - \mathbb{1}[y=z]\big)^2\pi_x\pi_y \le \frac{|\Omega|^2}{2}.$$

Now starting from any state y, i.e. the probability distribution with unit mass in state y, we have for the initial distance

$$\Big\|\frac{d\nu}{d\mu}\Big\|_{2,\mu} = \Big(\sum_{x\in\Omega}\Big(\frac{\nu_x}{\mu_x}\Big)^2\mu_x\Big)^{\frac{1}{2}} = \frac{1}{\sqrt{\pi_y}} \le \frac{1}{\sqrt{\min_{x\in\Omega}\pi_x}},$$

since $\boldsymbol{\mu} = \boldsymbol{\pi}$. Because $\mathcal{R}$ is bounded, $\min_{x\in\Omega}\pi_x$ is bounded from below by some constant $1/c_{\mathrm{e}} \in (0, \infty)$. □

E. Proof of Lemma 8

First note that $M^{[n]} \in \mathcal{F}^{[n]}$ and that its expectation is bounded, which can be concluded after writing

$$\begin{aligned}\mathbb{E}[|M^{[n]}|] &\le \sum_{j=1}^n a^{[j]}\mathbb{E}[\Big|\sum_{i=1}^d E_i^{[j]}(R_i^{[j-1]} - r_i^{\mathrm{opt}})\Big|]\\ &\le \sum_{j=1}^n a^{[j]}\max_{i=1,\dots,d}\{\mathcal{R}_i^{\max} - \mathcal{R}_i^{\min}\}\mathbb{E}[\sum_{i=1}^d\big|E_i^{[j]}\big|]\end{aligned} \tag{51}$$

and then substituting (43). Also,

$$\begin{aligned}\mathbb{E}[M^{[n]}|\mathcal{F}^{[n-1]}] &= \mathbb{E}[\sum_{j=1}^n a^{[j]}\boldsymbol{E}^{[j]\mathrm{T}}(\boldsymbol{R}^{[j-1]} - \boldsymbol{r}^{\mathrm{opt}})|\mathcal{F}^{[n-1]}]\\ &= M^{[n-1]} + a^{[n]}\mathbb{E}[\boldsymbol{E}^{[n]\mathrm{T}}(\boldsymbol{R}^{[n-1]} - \boldsymbol{r}^{\mathrm{opt}})|\mathcal{F}^{[n-1]}] = M^{[n-1]},\end{aligned}$$

which concludes the proof. □