2018 Annual American Control Conference (ACC)
June 27–29, 2018. Wisconsin Center, Milwaukee, USA

Learning in Non-Stationary Wireless Control Systems via Newton's Method

Mark Eisen, Konstantinos Gatsis, George J. Pappas and Alejandro Ribeiro

Abstract— **This paper considers wireless control systems over an unknown time-varying non-stationary channel. The goal is to maximize control performance of a set of independent control systems by allocating transmitting power within a fixed budget. Since the channel's time-varying distribution is unknown, samples of the channel are taken at every epoch. By reverting the resulting stochastic optimization problem in its Lagrange dual domain, we demonstrate that it takes the equivalent form of minimizing a certain empirical risk measure, a well-studied problem in machine learning. Newton's method is used to quickly learn approximately optimal power allocation policies over the sampled dual function as the channel evolves over time over windows of epochs. The quadratic convergence rate of Newton is used to establish, under certain conditions on the sampling size and rate of channel variation, an instantaneous learning and tracking of these optimal policies. Numerical simulations demonstrate the effectiveness of the learning algorithm on a low-dimensional wireless control problem.**

Index Terms— wireless autonomous systems, learning, Newton's method, non-stationary channel

I. Introduction

Wireless communication is increasingly used in autonomous applications to connect devices in industrial control environments, teams of robotic vehicles, and the Internet-of-Things. To guarantee safety and control performance it is customary to include a model of the wireless channel, for example an i.i.d. or Markov link quality, alongside the model of the physical system to be controlled. In such modeled-based approaches one can characterize, for example, that it is impossible to estimate or stabilize an unstable plant if its growth rate is larger than the rate at which the link drops packets [1]–[3], or below a certain channel capacity [4], [5]. Models also facilitate the allocation of communication resources to optimize control performance in, e.g., power allocation over fading channels [6], [7], or in event-triggered control [8]–[10].

In practice wireless autonomous systems operate under unpredictable channel conditions following unknown time-varying distributions, which are more often observable via collected channel quality samples. While one may try first to learn a channel model and then employ the above model-based design, common channel effects are not always amenable to modeling, such as mobility in the environment. In this paper we propose an alternative learning-based approach, whereby autonomy relies on collected channel samples to directly optimize control performance in a non-stationary environment, thus bypassing the channel-modeling phase. To achieve this we make a connection between the model-based design approach and an empirical risk minimization (ERM) problem, typical in machine learning.

In this paper we consider a wireless autonomous system where the design goal is to maximize a level of control performance for multiple systems while meeting a desired transmit power budget over the wireless channel (Section II). Power allocation is important in large wireless control systems where battery life is to be maximized, or has to operate under a strict budget of resource availability. The wireless channel is modeled as a fading channel with a time-varying and unknown distribution, and only available through samples taken over time. We show in Section III that the (Lagrange dual of the) power allocation problem can be rewritten using channel samples as an empirical risk minimization problem. To track the time-varying channel distribution in a fast and sample-efficient manner we employ two tools: i) we formulate the ERM problem over a window of channel samples taken from consecutive distributions, and ii) we employ a fast learning algorithm, in particular second order Newton's method (Section IV). More specifically, the quadratic convergence rate of Newton's method is shown to be sufficient to adapt to finding approximate solutions to slowly varying objectives with single steps [11], [12]. We point out that the authors' previous work [13], [14] studied a related multiple-access wireless control problem but under a stationary channel distribution, and employed a first-order stochastic method, which has slow convergence rates and hence not suitable for the present framework. Our methods differ from existing non-stationary learning works, e.g. the two time-scale approach in [15], by optimizing locally at all time epochs, rather than optimizing a global average performance.

Our contribution is an algorithm that uses channel samples to approximate the solution of the power allocation control problem over a non-stationary channel using second order information. Moreover, we prove that, under specific conditions on the sampling size and rate of channel variation, the algorithm reaches an approximately optimal point in a single iteration of Newton's method (Section V). This is further demonstrated in a numerical demonstration of learning over a time-varying channel (Section VI). Proofs for results are included in [16].

Supported by ARL DCIST CRA W911NF-17-2-0181 and Intel Science and Technology Center for Wireless Autonomous Systems. The authors are with the Department of Electrical and Systems Engineering, University of Pennsylvania. Email at: {maeisen, kgatsis, pappasg, aribeiro}@seas.upenn.edu.

II. Power Allocation in Control Systems

The wireless control architecture we consider includes m independent control systems labeled $i = 1, \ldots, m$. Each control system/agent i communicates over a wireless channel in order to close a loop and maximize a level of control performance. Due to propagation effects the channel fading conditions that each system i experiences, denoted by $h^i \in \mathbb{R}_+$, change unpredictably over time [17, Ch. 3], and are thus considered as a random variable. Fading is assumed constant during transmission slots and independently distributed over time slots (block fading), with distribution $\mathcal{H}$ that may change over time epochs in a non-stationary fashion. We consider the distribution $\mathcal{H}$ to be stationary over the course of a single time epoch.

When system i attempts to close its loop over the wireless channel a transmit power level p^i taking values in $\mathbb{R}_+$ is selected. Then channel fading and transmit power determine the signal-to-noise ratio (SNR) at the receiver, which affects the probability of successful decoding of the transmitted packet at the receiver. Overall we express the probability of successful packet delivery by a given relationship of the form $q(h^i, p^i)$ – see [6], [7] for more details on this model.

We proceed by deriving a problem formulation for optimal power allocation during a single epoch, and later in Section IV generalize to the non-stationary setting. Given a random channel state $h^i \in \mathbb{R}$ drawn from the distribution $\mathcal{H}$, assumed to be available at the transmitter at each slot, we wish to determine the amount of transmit power $p^i(h^i) : \mathbb{R}_+ \to \mathbb{R}_+$ to be used when attempting to close its loop. Then the probability of closing the loop is given by the value

$$y^i = \mathbb{E}_h \left\{ q(h^i, p^i(h^i)) \right\} \tag{1}$$

that is, the integral of the transmission success function over the channel distribution.

For each agent i, we also assume a monotonically increasing concave function $J_i : [0, 1] \to \mathbb{R}$ that returns a measure of control system performance as a function of the probability of successful transmission y^i given in (1). A concrete example is presented next.

Example 1 *Consider for example that a control system i is a scalar switched linear dynamical system of the form*

$$x_{t+1} = \begin{cases} A_c x_t + w_t & \textit{if loop closes} \\ A_o x_t + w_t & \textit{otherwise} \end{cases} \tag{2}$$

where $x_t \in \mathbb{R}$ is the state at transmission time t, A_c is the closed loop stable dynamics, A_o is the open loop unstable dynamics, and w_t is zero-mean i.i.d. disturbance process with variance W. The system attempts to close the loop at a high rate in order to minimize a control cost objective of the form

$$\lim_{N \to \infty} \frac{1}{N} \sum_{t=0}^{N-1} \mathbb{E} x_t^2 \tag{3}$$

Assuming the control loop in (2) *is closed with the success probability y^i in* (1) *at all future time steps, it is possible to express the above cost explicitly as a function of y_i. Using the system dynamics* (2)*, the variance of the system state satisfies the recursive formula*

$$\mathbb{E} x_{t+1}^2 = y^i \, A_c^2 \, \mathbb{E} x_t^2 + (1 - y^i) \, A_o^2 \, \mathbb{E} x_t^2 + W \tag{4}$$

Substituting this in the cost function (3) *yields*

$$J^i(y^i) = \frac{W}{1 - (y^i A_c^2 + (1 - y^i) A_o^2)} \tag{5}$$

This control cost function is a convex decreasing function, so that its negative satisfies the above problem assumptions.

As can be seen from Example 1, control performance in (3) is a long term objective. However, as the future channel distributions that will be encountered are time-varying and unknown, we do not have a way to incorporate their effect in the planned future control performance. Hence in this paper we use as a surrogate objective function the future control performance assuming the channel distribution will not change in the future (this can also be thought as a model predictive control approach)—-see Remark 3.

Given a set of channel states $\mathbf{h} := [h^1; h^2; \ldots; h^m]$, the goal is to determine a set of power allocation policies $\mathbf{p}(\mathbf{h}) = [p^1(h^1); p^2(h^2); \ldots; p^m(h^m)] \in \mathbb{R}^m$ whose expected aggregate value is within a maximum power budget $p_{\max}$ while maximizing the total system performance over m agents. This problem can be formally stated as the following optimization problem. Because J_i is monotonically increasing, using the auxiliary variables $\mathbf{y} = [y^1; y^2; \ldots; y^m]$ as defined in (1) we formulate the problem as

$$\begin{aligned} \mathbf{p}^*(\mathbf{h}) := \underset{\mathbf{p}, \mathbf{y} \in \mathbb{R}^m}{\operatorname{argmax}} & \sum_{i=1}^m J_i(y^i) \\ \text{s. t.} \quad & \sum_{i=1}^m \mathbb{E}_{h^i}(p^i(h^i)) \le p_{\max} \\ & y^i \le \mathbb{E}_{h^i} \left\{ q(h, p^i(h^i)) \right\} \quad i = 1, \ldots, m \end{aligned} \tag{6}$$

The problem in (6) states the optimal power allocation policy $\mathbf{p}^*(\mathbf{h})$ is the one that maximizes the expected aggregate control performance over channel states while guaranteeing that the expected total transmitting power is below an available budget $p_{\max}$. This problem is challenging to solve, both due to the infinite dimensional variable $\mathbf{p}(\mathbf{h})$ and the possible non-convexity coming from the constraint. However, this is naturally solved in the dual domain. To simplify the presentation, we introduce of a couple of new augmented variables. Define the vector $\tilde{\mathbf{q}}(\mathbf{p}(\mathbf{h})) \in \mathbb{R}^{m+1}$ of transmission probabilities augmented with the total power allocation as $\tilde{\mathbf{q}}(\mathbf{p}(\mathbf{h})) := [q(h^1, p^1(h^1)); \ldots; q(h^m, p^m(h^m)); -\sum_{i=1}^m p^i(h^i)]$ and the auxiliary vector $\tilde{\mathbf{y}} \in \mathbb{R}^{m+1}$ augmented with total power budget as $\tilde{\mathbf{y}} := [y^1; \ldots; y^m; -p_{\max}]$. The Lagrangian function is then formed as

$$\mathcal{L}(\mathbf{p}(\mathbf{h}), \mathbf{y}, \boldsymbol{\mu}) := \sum_{i=1}^m J_i(y^i) + \boldsymbol{\mu}^T \left(\mathbb{E}_{\mathbf{h}} \tilde{\mathbf{q}}(\mathbf{p}(\mathbf{h})) - \tilde{\mathbf{y}} \right), \tag{7}$$

where $\boldsymbol{\mu} = [\mu^1; \ldots; \mu^m; \tilde{\mu}] \in \mathbb{R}_+^{m+1}$ contains the dual

variables associated with each constraint. The dual loss function is then defined as $L(\boldsymbol{\mu}) := \max_{\mathbf{p},\mathbf{y}} \mathcal{L}(\mathbf{p}(\mathbf{h}), \mathbf{y}, \boldsymbol{\mu})$ and the corresponding dual problem as

$$\begin{aligned}\boldsymbol{\mu}^* &:= \operatorname*{argmin}_{\boldsymbol{\mu}\geq 0} L(\boldsymbol{\mu}) \\ &= \operatorname*{argmin}_{\boldsymbol{\mu}\geq 0} \max_{\mathbf{p},\mathbf{y}} \sum_{i=1}^{m} J_i(y^i) + \boldsymbol{\mu}^T \left(\mathbb{E}_{\mathbf{h}} \tilde{\mathbf{q}}(\mathbf{p}(\mathbf{h})) - \tilde{\mathbf{y}}\right). \quad (8)\end{aligned}$$

Observe that the problem in (8) is a simply constrained stochastic problem which can be solved with a variety of projected stochastic descent methods [13], [14], [18]–[20]. We stress that, while not necessarily strongly convex, the original problem in (6) can be shown to exhibit zero duality gap under the technical assumption that the primal problem is strictly feasible and that the channel probability distribution is non-atomic [21], implying that the optimal primal variable $\mathbf{p}^*(\mathbf{h})$ in (6) and be recovered from the optimal dual variable $\boldsymbol{\mu}^*$ in (8). Thus, the power allocation policy for each agent i is found indirectly by solving (8) and recovering as

$$p^i(h, \boldsymbol{\mu}) = \operatorname*{argmax}_{p} \mathbb{E}_h \left\{\mu^i q^i(h, p(h))\right\} - \tilde{\mu} p^i(h), \quad (9)$$

$$\mathbf{y}(\boldsymbol{\mu}) = \operatorname*{argmax}_{\mathbf{y}} \sum_{i=1}^{m} [J_i(y^i) - \mu_i y^i], \quad (10)$$

where the optimal policy is recovered using the optimal dual variable as $\mathbf{p}^*(\mathbf{h}) = \mathbf{p}(\mathbf{h}, \boldsymbol{\mu}^*)$.

III. ERM Formulation of Power Allocation

The stochastic program in (8) used to find the optimal resource allocation over a channel can be considered as specific formulation of empirical risk minimization (ERM) problem. In the general ERM problem, we consider a convex loss function $f(\boldsymbol{\mu}, \mathbf{h})$ of a decision variable $\boldsymbol{\mu} \in \mathbb{R}^{m+1}$ and random variable $\mathbf{h}$ drawn from a particular channel distribution $\mathcal{H}$ and seek the optimal variable $\boldsymbol{\mu}^*$ that minimizes the associated expected loss $L(\boldsymbol{\mu}) := \mathbb{E}_{\mathbf{h}}[f(\boldsymbol{\mu}, \mathbf{h})]$. For the optimal power allocation problem in (6), we define the loss function L and associated ERM problem using its dual as

$$\boldsymbol{\mu}^* := \operatorname*{argmin}_{\boldsymbol{\mu}\geq \mathbf{0}} L(\boldsymbol{\mu}) := \operatorname*{argmin}_{\boldsymbol{\mu}\geq \mathbf{0}} \mathbb{E}_{\mathbf{h}} f(\boldsymbol{\mu}, h), \quad (11)$$

$$\text{where } f(\boldsymbol{\mu}, h) := \sum_{i=1}^{m} J_i(y^i(\boldsymbol{\mu})) + \boldsymbol{\mu}^T \left(\tilde{\mathbf{q}}(\mathbf{p}(h, \boldsymbol{\mu})) - \tilde{\mathbf{y}}(\boldsymbol{\mu})\right).$$

Because the channel distribution $\mathcal{H}$ is not known, we cannot minimize $L(\boldsymbol{\mu})$ directly. In ERM problems, we replace the expected loss by an empirical risk by taking n m-dimensional samples from the channel $\mathcal{H}$, labeled $\mathbf{h}^1, \mathbf{h}^2, \ldots, \mathbf{h}^n$, (where $\mathbf{h}^l := [h^{1,l}; \ldots; h^{m,l}]$) and consider a statistical average, i.e.

$$\hat{L}(\boldsymbol{\mu}) := \frac{1}{n} \sum_{l=1}^{n} f(\boldsymbol{\mu}, \mathbf{h}^l) := \frac{1}{n} \sum_{l=1}^{n} f^l(\boldsymbol{\mu}). \quad (12)$$

Note that we define $f^l(\boldsymbol{\mu}) := f(\boldsymbol{\mu}, \mathbf{h}^l)$ to remove dependencies on random realization $\mathbf{h}^l$ for notational simplicity. Further note that the empirical sampling additionally modifies our optimal power allocation policy, previously defined in (9) to the sampled version, i.e.

$$\hat{p}^i(h, \boldsymbol{\mu}) := \operatorname*{argmax}_{p} \frac{1}{n} \sum_{l=1}^{n} \left\{\mu^i q(h^{i,l}, p(h)) - \tilde{\mu} p(h)\right\}. \quad (13)$$

While the empirical risk $\hat{L}(\boldsymbol{\mu})$ is indeed only an approximation of the expected loss $L(\boldsymbol{\mu})$, we may define a constant V called the *statistical accuracy* of $\hat{L}$. The statistical accuracy V provides a bound of the difference in the empirical and expected loss for all $\boldsymbol{\mu}$ with high probability (i.e. at least $1 - \delta$ for some small δ). In other words, we define V to be the constant that satisfies

$$\sup_{\boldsymbol{\mu}} \|\hat{L}(\boldsymbol{\mu}) - L(\boldsymbol{\mu})\| \leq V \quad \text{w.h.p.} \quad (14)$$

The upper bounds on V are well studied in the learning literature and in general may involve a number of parameters of the loss function f as well as, perhaps most importantly, the number of samples n. In the simple sampling of $L(\boldsymbol{\mu})$ defined in (12), a bound for the statistical accuracy V can be obtained in the order of $\mathcal{O}(1/\sqrt{n})$ or, in some cases, $\mathcal{O}(1/n)$ [22], [23]. The inherent error accrued by replacing the expected loss with a statistical average motivates the use of further regularization to impose more desirable properties on the empirical risk $\hat{L}(\boldsymbol{\mu})$, such as strong convexity. Because the optimal empirical value $\hat{L}^* := \min \hat{L}(\boldsymbol{\mu})$ will have a difference of order V from the optimal expected value L^*, an additional regularization whose bias is of order V will produce a solution that is any more inaccurate. We can therefore add the regularization term $\alpha V/2 \|\boldsymbol{\mu}\|^2$ to the empirical risk in (12). In addition, we can remove the non-negativity constraint on the dual variables in (11) through the use of a logarithmic barrier. To preserve smoothness for small $\boldsymbol{\mu}$, however, we specifically use an ϵ-thresholded log function, defined as

$$\log_\epsilon(\boldsymbol{\mu}) := \begin{cases} \log(\boldsymbol{\mu}) & \boldsymbol{\mu} \geq \epsilon \\ \ell_{2,\epsilon}(\boldsymbol{\mu} - \epsilon) & \boldsymbol{\mu} < \epsilon, \end{cases} \quad (15)$$

where $\ell_{2,\epsilon}(\boldsymbol{\mu})$ is a second order Taylor series expansion of $\log(\boldsymbol{\mu})$ centered at ϵ for some small $0 < \epsilon < 1$. We then use $-\beta V \mathbf{1}^T \log_\epsilon \boldsymbol{\mu}$ as a second regularization term, and obtain a regularized empirical risk function

$$R(\boldsymbol{\mu}) := \frac{1}{n} \sum_{l=1}^{n} f^l(\boldsymbol{\mu}) + \frac{\alpha V}{2} \|\boldsymbol{\mu}\|^2 - \beta V \mathbf{1}^T \log_\epsilon \boldsymbol{\mu}. \quad (16)$$

The regularized $R(\boldsymbol{\mu})$ provides a strongly convex approximation to L whose minimizer R^* will satisfy $\boldsymbol{\mu} \geq \mathbf{0}$ and of order V from the true optimal L^*, given appropriate selection of constants α and β. This is demonstrated formally in the following proposition.

Proposition 1 *Consider $L^* = \min_{\boldsymbol{\mu}\geq \mathbf{0}} L(\boldsymbol{\mu})$ as the optimal value of the expected loss function with nonnegative $\boldsymbol{\mu}$ and define $R^* := \min_{\boldsymbol{\mu}} R(\boldsymbol{\mu})$ as the optimal value of the regularized empirical risk. The difference $|L^* - R^*|$ is upper*

bounded on the order of statistical accuracy V, i.e.

$$|L^* - R^*| \leq \mathcal{O}(V), \quad w.h.p. \tag{17}$$

The full proof is omitted here for space considerations, however can be obtained from considering the error induced by a logarithmic barrier with respect to the constrained problem is bounded $(m+1)\beta V$—see, e.g., [24, Section 11.2.2]. Likewise, the quadratic regularizer is known to introduce a bias on the order of $\mathcal{O}(V)$ [22].

IV. ERM over non-stationary channel

The regularized problem in (16) can be solved using standard optimization techniques for a stationary channel distribution $\mathcal{H}$. To learn the optimal allocation policies over a non-stationary channel, however, we may consider a modified empirical risk function. Consider an epoch index $k = 0, 1, \ldots$ that specifies a particular channel distribution $\mathcal{H}_k$ with realizations $\mathbf{h}_k$. At each epoch k we draw a new set of n samples from the current distribution $\mathcal{H}_k$ labelled $\mathbf{h}_k^1, \ldots, \mathbf{h}_k^n$ and wish to an approximate solution to the expected loss L_k defined in (11) with $\mathcal{H} = \mathcal{H}_k$. While we may use a simple empirical risk as we did in (12), we instead define a more general statistical loss function for a non-stationary channel using samples from the previous M epochs. Consider that we keep a window of collected samples of total size $N = Mn$ and at epoch k, with some associated statistical accuracy V_N.

The bounds on this constant V_N in general for the collection of M sampling distributions will be dependent on a number of parameters used to define both $\tilde{R}_k(\boldsymbol{\mu})$ and $L_k(\boldsymbol{\mu})$. The sample batch size n, window size M, and correlation between successive distributions $\mathcal{H}_j$ and $\mathcal{H}_{j+1}$ will all play a role. Precise bounds on this statistical accuracy in the case of non-i.i.d. samples would require a sophisticated statistical analysis and is not considered in this work. For the purposes of this work we use a user-defined accuracy $\hat{V}$ that plays the role of the statistical accuracy V_N. Then, using the same regularizations introduced previously, we define the regularized loss function R_k and associated optimal dual variable $\boldsymbol{\mu}_k^*$ as

$$R_k(\boldsymbol{\mu}) := \frac{1}{M} \sum_{j=k-M+1}^{k} \hat{L}_j(\boldsymbol{\mu}) + \frac{\alpha\hat{V}}{2}\|\boldsymbol{\mu}\|^2 - \beta\hat{V}\mathbf{1}^T \log_\epsilon \boldsymbol{\mu} \tag{18}$$

We subsequently define $\boldsymbol{\mu}_k^* := \operatorname{argmin}_{\boldsymbol{\mu}} R_k(\boldsymbol{\mu})$. The definition of the loss function in (18) includes the batches of n samples taken from the previous M channel distributions $\mathcal{H}_{k-M+1}, \ldots, \mathcal{H}_k$. This definition is, in a sense, a generalization of the simpler empirical risk $R(\boldsymbol{\mu})$. Observe that, by using a window size of $M = 1$, we use only samples from the current channel and recover $R(\boldsymbol{\mu})$ for $\mathcal{H} = \mathcal{H}_k$.

The pivotal observation used to establish our learning algorithm is that the exact solution to loss function $\boldsymbol{\mu}_k^*$ will only solve the expected loss L_k to within accuracy V_N. Therefore it is not necessary to minimize (18) exactly, and is indeed sufficient to find an approximately accurate solution. As V_N may not be known in practice, we more select an accuracy $\hat{V}$ that is user-defined and *may*, in the case where the statistical accuracy V_N is known, be equivalent, i.e. $\hat{V} = V_N$. We therefore only try to find $\hat{V}$-accurate solutions to R_k. While many optimization methods can be used to find a minimizer to (18), we demonstrate in the next section that fast second order methods can be used to *instantaneously* learn approximate minimizers—and by extension solve (11)—at each epoch k as the channel distribution $\mathcal{H}_k$ changes. Note that, we here say *instantaneously* in the discrete time sense, or, in other words, with a single iteration over an epoch.

A. Learning via Newton's Method

We propose the use of Newton's method to approximately minimize (18) efficiently as the channel $\mathcal{H}_k$ changes with epoch k. Motivated by recent work in using Newton's method to solve large scale ERM problems through adaptive sampling [11], [12], at each iteration we use the N samples currently stored to find a point $\boldsymbol{\mu}_k$ that approximately solves for $\boldsymbol{\mu}_k^*$ (and thereby a V_N optimal solution to expected loss $L_k(\boldsymbol{\mu})$). The current iterate $\boldsymbol{\mu}_k$ then provides a "soft" start that to finding a point $\boldsymbol{\mu}_{k+1}$ that approximately minimizes the new function $R_{k+1}(\boldsymbol{\mu})$. By quickly finding the near-optimal solutions for each loss function, we efficiently learn the optimal power allocation of the wireless channel as the channel distribution evolves over time.

To discuss the details of such a learning algorithm, we first recall Newton's method. Newton's method at epoch k computes a new iterate $\boldsymbol{\mu}_{k+1}$ by subtracting from the current iterate $\boldsymbol{\mu}_k$ the product of the Hessian inverse and the gradient of the function $R_{k+1}(\boldsymbol{\mu}_k)$. For the empirical dual loss function R_k defined in (18), the gradient $\nabla R_k(\boldsymbol{\mu})$ and Hessian $\nabla^2 R_{k+1}(\boldsymbol{\mu})$ can be computed as

$$\nabla R_k(\boldsymbol{\mu}) = \frac{1}{N} \sum_{j=k-M+1}^{k} \sum_{l=1}^{n} \left(\tilde{\mathbf{q}}(\mathbf{p}(\mathbf{h}_j^l, \boldsymbol{\mu})) - \tilde{\mathbf{y}}(\boldsymbol{\mu})\right) + \alpha\hat{V}\boldsymbol{\mu} - \beta\hat{V}\boldsymbol{\mu}^{-1}. \tag{19}$$

$$\nabla^2 R_k(\boldsymbol{\mu}) = \frac{1}{N} \sum_{j=k-M+1}^{k} \sum_{l=1}^{n} \nabla_{\boldsymbol{\mu}} \left(\tilde{\mathbf{q}}(\mathbf{p}(\mathbf{h}_j^l, \boldsymbol{\mu})) - \tilde{\mathbf{y}}(\boldsymbol{\mu})\right) + \alpha\hat{V}\mathbf{I} + \beta\hat{V}\operatorname{diag}\{\boldsymbol{\mu}^{-2}\}. \tag{20}$$

The new approximate solution $\boldsymbol{\mu}_{k+1}$ is then found from current approximate solution $\boldsymbol{\mu}_k$ using the Newton update

$$\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k - \mathbf{H}_{k+1}^{-1} \nabla R_{k+1}(\boldsymbol{\mu}_k), \tag{21}$$

where we use $\mathbf{H}_{k+1} := \nabla^2 R_{k+1}(\boldsymbol{\mu}_k)$ as simplified notation.

Consider that $\boldsymbol{\mu}_k$ is a $\hat{V}$-accurate solution of current loss function R_k, i.e. $R_k(\boldsymbol{\mu}_k) - R_k^* \leq \hat{V}$. The new loss function R_{k+1} differs from R_k only in the discarding of old samples $\hat{L}_{k-M+1}$ and inclusion of samples $\hat{L}_{k+1}$ drawn from current distribution $\mathcal{H}_{k+1}$. If the channel distribution changes sufficiently slowly, i.e. $\mathcal{H}_{k+1}$ is close to $\mathcal{H}_k$, then the respective loss functions R_{k+1} and R_k and their optimal points R_{k+1}^* and R_k^* will also not differ greatly under some

Algorithm 1 Learning via Newton's Method

1: **Parameters:** Sample size increase constants $n_0 > 0$, $M_0 \geq 1$ backtracking params $0 < \delta < 1 < \gamma$, and accuracy $\hat{V}$.
2: **Input:** Initial sample size $n = n_0$ and argument $\boldsymbol{\mu}_n = \boldsymbol{\mu}_{m_0}$ with $\|\nabla R_n(\boldsymbol{\mu}_{k+1})\| < (\sqrt{2\alpha})\hat{V}$
3: **for** $k = 0, 1, 2, \ldots$ **do** {main loop}
4: Reset factor $n = n_0$, $M = M_0$.
5: **repeat** {sample size backtracking loop}
6: Draw n samples from $\mathcal{H}_{k+1}$, discard from $\mathcal{H}_{k-M+1}$.
7: Gradient $\nabla R_{k+1}(\boldsymbol{\mu}_k)$, Hessian $\mathbf{H}_{k+1}$ [cf. (19), (20)]:
8: Newton Update [cf. (21)]:

$$\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k - \mathbf{H}_{k+1}^{-1} \nabla R_{k+1}(\boldsymbol{\mu}_k)$$

9: Determine power allocation, aux. variables [cf. (13), (10)]:

$$\hat{p}^i(h, \boldsymbol{\mu}_{k+1}) = \operatorname*{argmax}_p \frac{1}{N} \sum_j \sum_l \left\{ \mu^i q(h_j^{i,l}, p(h)) - \tilde{\mu} p(h) \right\}.$$

$$\mathbf{y}(\boldsymbol{\mu}_{k+1}) = \operatorname*{argmax}_{\mathbf{y}} \sum_{i=1}^{m} [J_i(y^i) - \mu_{k+1}^i y^i].$$

10: Backtrack sample draw $n = \gamma n$, window size $M = \delta M$.
11: **until** $\|\nabla R_{k+1}(\boldsymbol{\mu}_{k+1})\| < (\sqrt{2\alpha})\hat{V}$
12: **end for**

smoothness assumptions. Therefore, under certain conditions a single step of Newton's method as performed in (21) can in fact be sufficient to reach a $\hat{V}$-accurate solution of the new loss function R_{k+1}. Given then a $\hat{V}$-accurate solution $\boldsymbol{\mu}_0$ of initial loss R_0, the proceeding and all subsequent iterates $\boldsymbol{\mu}_k$ will remain within the statistical accuracy of their respective losses R_k as the channel distribution varies over time. The formal presentation and technical details of this result are discussed in Section V of this paper.

The learning algorithm is presented in Algorithm 1. After preliminaries and initializations in Steps 1-4, the backtracking loop starts in Step 5. Each iteration begins in Step 6 with the drawing of n samples from the new channel distribution $\mathcal{H}_{k+1}$ and discarding the old samples from $\mathcal{H}_{k-M+1}$ to form R_{k+1}. Note that samples will be only be discarded for $k > M$. The gradient ∇R_{k+1} and Hessian $\mathbf{H}_{k+1}$ of the regularized dual loss function are computed in Step 7. The Newton step is taken with respect to R_{k+1} in Step 8. In Step 9, the optimal primal variables are computed with respect to the updated dual variables. This includes both the auxiliary variables $\mathbf{y}(\boldsymbol{\mu}_{k+1})$ and the power allocation policy $\hat{\mathbf{p}}(\mathbf{h}, \boldsymbol{\mu}_{k+1})$ itself. Because there are function and channel system parameters that are not known in practice, we include a backtracking step for the parameters n and M in Step 10 to ensure the new iterate $\boldsymbol{\mu}_{k+1}$ is within the intended accuracy $\hat{V}$ of $\boldsymbol{\mu}_{k+1}^*$. The importance of the true statistical accuracy V_N and other unknown parameters in the convergence result and subsequent need for backtracking is discussed in Remark 2 in Section V.

V. Convergence Analysis

In this section we provide a theoretical analysis of the ERM Newton learning algorithm in (21). Specifically, we analyze the ERM formulation of the power allocation problem in (18) and establish certain conditions that guarantees the instantaneous solving of statistically accurate solutions for each epoch k. Our primary theoretical result gives conditions dependent on statistical accuracy and rate of non-stationarity that allows for a single iteration of the Newton's method in (21) to solve the new loss function R_{k+1} to within its estimated statistical accuracy $\hat{V}$. We begin by presenting a series of assumptions made in our analysis regarding the dual loss functions f.

Assumption 1 *The loss functions $f(\boldsymbol{\mu}, h_k)$ are convex with respect to $\boldsymbol{\mu}$ for all values of $\mathbf{z}$. Moreover, their gradients $\nabla f(\boldsymbol{\mu}, \mathbf{z})$ are Lipschitz continuous with constant Δ.*

Assumption 2 *The loss functions $f(\boldsymbol{\mu}, \mathbf{h})$ are self-concordant with respect to $\boldsymbol{\mu}$ for all $\mathbf{h}$.*

Based on Assumption 1, we obtain that the regularized empirical risk gradients ∇R_k are Lipschitz continuous with constant $\Delta + c\hat{V}$ where $c := \alpha + \beta/\epsilon^2$ and the function R_k is strongly convex with constant $\alpha\hat{V}$. Assumption 2 states the loss functions are additionally self concordant which is a customary assumption in the analysis of second-order methods. It also follows that the functions R_{k+1} are therefore self concordant because both the quadratic and thresholded log regularizers are self-concordant. Before proceeding with our other assumptions, we present a brief remark regarding our power allocation control problem.

Remark 1 While we state these assumptions in terms of the sampled dual functions f due to their direct use in the proceeding analysis, they indeed have implications on the primal domain problem in (6). Note that the dual function is always convex, while the smoothness condition in Assumption 1 can be obtained from the strong convexity of the control performance $\sum_i J_i$. The self-concordance property on the dual function in Assumption 2 is not easily derived from properties of $J_i(\cdot)$ or $q(\cdot)$, although there has been work that establishing self concordance of the dual for various machine learning problems [25], [26].

Assumption 3 *The difference between the gradients of the empirical risk $\hat{L}_k$ and the statistical average loss L_k is bounded by $V_N^{1/2}$ for all $\boldsymbol{\mu}$ and k with high probability,*

$$\sup_{\boldsymbol{\mu}} \|\nabla L_k(\boldsymbol{\mu}) - \nabla \hat{L}_k(\boldsymbol{\mu})\| \leq V_N^{1/2}, \qquad \textit{w.h.p.} \tag{22}$$

Assumption 4 *The difference between two successive expected loss $L_k(\boldsymbol{\mu}) = \mathbb{E}_{h_k} f(\boldsymbol{\mu}, \mathbf{h}_k)$ and $L_{k+1}(\boldsymbol{\mu}) = \mathbb{E}_{h_{k+1}} f(\boldsymbol{\mu}, \mathbf{h}_{k+1})$ and the difference between gradients are bounded respectively by a bounded sequence of constants $\{D_k\}, \{\bar{D}_k\} \geq 0$ for all $\boldsymbol{\mu}$,*

$$\sup_{\boldsymbol{\mu}} |L_k(\boldsymbol{\mu}) - L_{k+1}(\boldsymbol{\mu})| \leq D_k, \tag{23}$$

$$\sup_{\boldsymbol{\mu}} \|\nabla L_k(\boldsymbol{\mu}) - \nabla L_{k+1}(\boldsymbol{\mu})\| \leq \bar{D}_k. \tag{24}$$

Assumption 3 bounds the difference between gradients of the expected loss and the empirical risk with N samples by $V_N^{1/2}$. This is a reasonable bound for the convergence of gradients to their statistical averages using the law of large numbers. Assumption 4 says that the point-wise difference in the expected loss functions at epochs k and $k+1$ cannot differ more than a constant D_k. This effectively provides a limit on the rate at which the channel evolves between epochs, and is used to establish that optimal dual variables for two consecutive empirical risk functions R_k and R_{k+1} are not very different.

The primary result is arranged as follows. We wish to find conditions on the parameters of the sampling and the non-stationarity that guarantee that, starting from an approximate solution to R_k, a single step of Newton's method generates an approximately accurate solution to R_{k+1}. Using this result, it follows then that, assuming an initial point $\boldsymbol{\mu}_0$ that is within the intended accuracy of R_0, the method will always continue to find a new approximately accurate solution for each epoch k as the channel distribution changes with k. This result is achieved in two primary steps. First, we find a condition that guarantees that an approximate solution of R_k is in the quadratic convergence region of R_{k+1}. Second, we find a condition that guarantees that a point within the quadratic convergence region of R_{k+1} reaches its intended accuracy with a single Newton step.

We begin by deriving a condition for the first property to hold. We wish to show that a $\hat{V}$-accurate solution to R_k, labelled $\boldsymbol{\mu}_k$ is in in the quadratic convergence region of R_{k+1} if certain conditions hold. By quadratic convergence region, we refer to the local region in which Newton's method is known to converge at a fast quadratic rate. The analysis of Newton's method commonly characterizes quadratic convergence in terms of a quantity called the Newton decrement, which is simply a weighted norm of the gradient, and is explicitly defined as $\lambda_{k+1}(\boldsymbol{\mu}) := \|\nabla^2 R_{k+1}(\boldsymbol{\mu})^{-1/2} \nabla R_{k+1}(\boldsymbol{\mu})\|$. We say the dual iterate $\boldsymbol{\mu}$ is in the quadratic convergence region of R_{k+1} when $\lambda_{k+1}(\boldsymbol{\mu}) < 1/4$—see [24, Chapter 9.6.4]. In the following proposition, we give conditions under which a $\hat{V}$-accurate iterate $\boldsymbol{\mu}_k$ is also within the quadratic convergence region of the succeeding loss function R_{k+1}.

Lemma 1 *Define $\boldsymbol{\mu}_k$ as a $\hat{V}$-accurate optimal solution of the loss R_k, i.e., $R_k(\boldsymbol{\mu}_k) - R_k(\boldsymbol{\mu}_k^*) \leq \hat{V}$. In addition, define $\lambda_{k+1}(\boldsymbol{\mu}) := \left(\nabla R_{k+1}(\boldsymbol{\mu})^T \nabla^2 R_{k+1}(\boldsymbol{\mu})^{-1} \nabla R_{k+1}(\boldsymbol{\mu})\right)^{1/2}$ as the Newton decrement of variable $\boldsymbol{\mu}$ associated with the loss R_{k+1}. If Assumptions 1-4 hold, then Newton's method at point $\boldsymbol{\mu}_k$ is in the quadratic convergence phase for the objective function R_{k+1}, i.e., $\lambda_{k+1}(\boldsymbol{\mu}_k) < 1/4$, if we have*

$$\left(\frac{2(\Delta + c\hat{V})\hat{V}}{\alpha \hat{V}}\right)^{1/2} + \frac{2V_N^{1/2} + \bar{D}_k}{(\alpha \hat{V})^{1/2}} < \frac{1}{4}. \quad \textit{w.h.p.} \tag{25}$$

With Lemma 1 we establish the first necessary condition that ensures that a $\hat{V}$-accurate solution to R_k is in the quadratic convergence region of R_{k+1}, as defined by the Newton decrement $\lambda_{k+1}(\boldsymbol{\mu}_k)$. It remains then to show the second step in the analysis, namely that a point in the quadratic convergence region of R_{k+1} will reach its statistical accuracy with a single Newton step as given in (21). To demonstrate this, we first present the following proposition that upper bounds the sub-optimality of the point $\boldsymbol{\mu}_k$ with respect to the optimal solution of R_{k+1}^*.

Lemma 2 *Consider a point $\boldsymbol{\mu}_k$ that minimizes the loss function R_k to within accuracy $\hat{V}$, i.e. $R_k(\boldsymbol{\mu}_k) - R_k(\boldsymbol{\mu}_k^*) \leq \hat{V}$. Provided that Assumptions 1-4 hold, the sub-optimality $R_{k+1}(\boldsymbol{\mu}_k) - R_{k+1}^*$ is upper bounded as*

$$R_{k+1}(\boldsymbol{\mu}_k) - R_{k+1}^* \leq 4V_N + \hat{V} + 2D_k \quad \textit{w.h.p.} \tag{26}$$

Lemma 2 demonstrates a bound on the suboptimality with respect to R_{k+1} of a point $\boldsymbol{\mu}_k$ that is within the statistical accuracy V_N of the previous risk function R_k. From here, we establish the suboptimality of the iterate $\boldsymbol{\mu}_{k+1}$ with respect to R_{k+1}, found through the Newton update in (21).

Lemma 3 *Consider $\boldsymbol{\mu}_k$ to be in the quadratic neighborhood of the loss R_{k+1}, i.e., $\lambda_{k+1}(\boldsymbol{\mu}_k) \leq 1/4$. Recall the definition of the variable $\boldsymbol{\mu}_{k+1}$ in* (21) *as the updated variable using Newton's method. If Assumptions 1-3 hold, then the difference $R_{k+1}(\boldsymbol{\mu}_{k+1}) - R_{k+1}^*$ is upper bounded by*

$$R_{k+1}(\boldsymbol{\mu}_{k+1}) - R_{k+1}^* \leq 144 (R_{k+1}(\boldsymbol{\mu}_k) - R_{k+1}^*)^2. \tag{27}$$

The proof of this lemma is omitted for space considerations, but follows closely the result in [11, Proposition 4]. With Lemma 3 we establish a quadratic rate of convergence of the suboptimality of the Newton update in (21). Observe that by substituting the upper bound on $R_{k+1}(\boldsymbol{\mu}_k) - R_{k+1}^*$ from Lemma 2, a condition can easily be derived under which the suboptimality of the new iterate is within the accuracy $\hat{V}$ of R_{k+1}. Using the results of Lemmata 1-3, we present our main result in the following theorem.

Theorem 1 *Consider Newton's method defined in* (21) *and the full learning method detailed in Algorithm 1. Define V_N to be the statistical accuracy of $N = Mn$ potentially non-i.i.d. samples, with n samples taken from each of the M most recent channel distributions $\mathcal{H}_k$. Further consider the variable $\boldsymbol{\mu}_k$ as a $\hat{V}$-optimal solution of the loss R_k, and suppose Assumptions 1-4 hold. If the sample size m and window size M are chosen such that the following conditions*

$$\left(\frac{2(\Delta + c\hat{V})\hat{V}}{\alpha \hat{V}}\right)^{1/2} + \frac{2V_N^{1/2} + \bar{D}_k}{(\alpha \hat{V})^{1/2}} < \frac{1}{4} \tag{28}$$

$$144(4V_N + \hat{V} + 2D_k)^2 \leq \hat{V} \tag{29}$$

are satisfied, where V_N depends upon n and M, then the variable $\boldsymbol{\mu}_{k+1}$ computed from (21) *has the suboptimality of $\hat{V}$ with high probability, i.e.,*

$$R_{k+1}(\boldsymbol{\mu}_{k+1}) - R_{k+1}^* \leq \hat{V}, \qquad \textit{w.h.p.} \tag{30}$$

The inequalities specify conditions under which $\boldsymbol{\mu}_{k+1}$ is a $\hat{V}$-optimal solution of R_{k+1}, and come directly from the

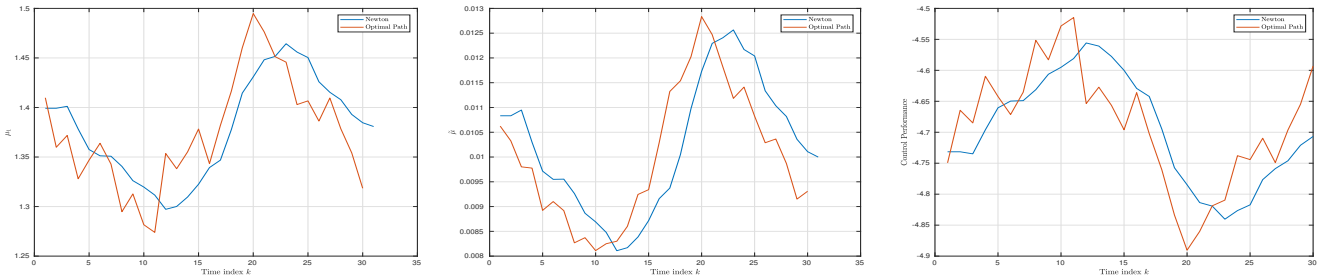

Fig. 1: Convergence paths of optimal values vs. values generated by the Newton learning method for time-varying $\mathcal{H}_k$ for dual variables (left) μ^1, (center) $\tilde{\mu}$, and (right) control performance $\sum J_i(y^i)$. Newton's method is able to find an approximately optimal value for the dual variables and respective control performance at each iteration.

preceding lemmata. We conclude our analysis with series a remarks discussing the applicability of the result in Theorem 1 and its connection to the power allocation problem in (6).

Remark 2 Observe that the conditions in (28) and (29) are functions of four primary terms, $\hat{V}$, V_N, D_k, and $\bar{D}_k$. The latter terms, D_k and $\bar{D}_k$, bound the difference in the neighboring expected loss functions L_k and L_{k+1}, and can be viewed as a bound on the degree of non-stationarity of the channel distribution $\mathcal{H}$ between successive iterations. This limits the degree that the distributions can vary over epochs.

Recall that the second term, V_N, in fact indirectly provides conditions on the sample size n and window size M used to define R_k necessary to learn a $\hat{V}$-optimal solution. Because the specific nature of both V_N and D_k come from the channel distributions that are most likely not known in practice, the satisfaction of these conditions motivates the use of the backtracking step in Step 10 of Algorithm 1. The free parameters n and M can be respectively increased and decreased until the new iterate is within the intended accuracy $\hat{V}$, as verified in Step 11.

Remark 3 Theorem 1 indirectly provides a guarantee that the selected power allocation is tracking the optimal power allocation that would solve our original problem (6) and would maximize the objective function $J_i(y_i)$. Recall that, as explained after Example 1, this objective function is a proxy for control performance but not exactly the realized long term control performance as in, e.g., (3). In that sense, the above theorem does not directly guarantee control performance, or in fact even system stability. Regardless, we expect that if the time-varying channel distribution remains always sufficiently favorable, good control performance will be maintained. We aim to explore this issue in future work.

VI. Simulation Results

We simulate the performance of our second order learning method on a simple wireless control problem. Consider the dynamical system from Example 1 with $A_c = 0$ and $A_o = a$. Our control performance for the ith agent $J_i(\mathbb{E}q(h, p^i(h^i)))$ measures the mean square error performance and is now given by

$$J_i(y^i) = \frac{1}{1 - a^2(1 - y^i)}. \tag{31}$$

The probability of successful transmission for agent i is modeled as a negative exponential function of both the power and channel state, $q(h^i, p^i(h^i) := 1 - \nu e^{-\nu h^i p^i(h^i)}$ for some $\nu > 0$. The channel states at epoch k are drawn from an exponential distribution with mean u_k. To model a time-varying channel, we slowly vary u_k for different epochs k. We draw $n = 200$ samples and store a window of the previous $M = 5$ channels for a total of $N = 1000$ samples.

To demonstrate the ability of Newton's method to instantaneously learn an approximately optimal power allocation as the channel distribution varies over time, we perform Algorithm 1 over the ERM problem in (11) with the defined control performance $J(\cdot)$, transmission probabilities $q(\cdot)$ and channel distributions $\mathcal{H}_k$. In Figure 1 we show the path of Newton's method at each epoch k for the dual variables $\boldsymbol{\mu}_k$, $\tilde{\mu}_k$, and the control performance $\sum_{i=1}^m J_i(\mathbf{y}_k^i)$. The red line of each figure plots the optimal values for the current distribution parameter u_k as it changes with k. The blue line plots the values generated by Newton's method over epochs. The channel evolves at over epochs by a fixed rate $u_{k+1} = u_k \pm r$ for some rate r. Within some small error, Newton's method is indeed able to quickly and approximately find each new solution as the channel varies over time.

VII. Conclusion

In this paper we develop a novel, second order algorithm to efficiently learn optimal power allocation policies for a network of wireless agents over a time varying channel. We employ previous results of strong duality for a power allocation problem in a control system to solve the problem in a dual domain. Because channel conditions are unknown and time-varying, we take samples and formulate the dual problem as an empirical risk minimization (ERM) problem. We show the time-varying optimal power allocation can be found quickly and, under certain conditions, instantaneously using Newton's method up to statistical approximation, and demonstrate with a numerical simulation.

References

[1] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry, "Kalman filtering with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1453–1464, 2004.

[2] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. Sastry, "Foundations of control and estimation over lossy networks," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 163–187, 2007.

[3] J. P. Hespanha, P. Naghshtabrizi, and Y. Xu, "A survey of recent results in networked control systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 138–162, 2007.

[4] S. Tatikonda and S. Mitter, "Control under communication constraints," *IEEE Transactions on Automatic Control*, vol. 49, no. 7, pp. 1056–1068, 2004.

[5] A. Sahai and S. Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link&# 8212; part i: Scalar systems," *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3369–3395, 2006.

[6] K. Gatsis, A. Ribeiro, and G. J. Pappas, "Optimal power management in wireless control systems," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1495–1510, 2014.

[7] D. E. Quevedo, A. Ahlén, A. S. Leong, and S. Dey, "On Kalman filtering over fading wireless channels with controlled transmission powers," *Automatica*, vol. 48, no. 7, pp. 1306–1316, 2012.

[8] M. Mazo and P. Tabuada, "Decentralized event-triggered control over wireless sensor/actuator networks," *IEEE Transactions on Automatic Control*, vol. 56, no. 10, pp. 2456–2461, 2011.

[9] J. Araújo, M. Mazo, A. Anta, P. Tabuada, and K. H. Johansson, "System architectures, protocols and algorithms for aperiodic wireless control systems," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 1, pp. 175–184, 2014.

[10] M. H. Mamduhi, D. Tolić, A. Molin, and S. Hirche, "Event-triggered scheduling for stochastic multi-loop networked control systems with packet dropouts," in *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*. IEEE, 2014, pp. 2776–2782.

[11] A. Mokhtari, H. Daneshmand, A. Lucchi, T. Hofmann, and A. Ribeiro, "Adaptive newton method for empirical risk minimization to statistical accuracy," in *Advances in Neural Information Processing Systems*, 2016, pp. 4062–4070.

[12] M. Eisen, A. Mokhtari, and A. Ribeiro, "Large scale empirical risk minimization via truncated adaptive newton method," *arXiv preprint arXiv:1705.07957*, 2017.

[13] K. Gatsis, M. Pajic, A. Ribeiro, and G. J. Pappas, "Opportunistic control over shared wireless channels," *IEEE Transactions on Automatic Control*, vol. 60, no. 12, pp. 3140–3155, December 2015.

[14] K. Gatsis, A. Ribeiro, and G. J. Pappas, "Random access design for wireless control systems," *Automatica*, 2018, to appear. Available on Arxiv. [Online]. Available: http://arxiv.org/pdf/1605.00627v1.pdf

[15] A. Mokkadem and M. Pelletier, "A generalization of the averaging procedure: The use of two-time-scale algorithms," *SIAM Journal on Control and Optimization*, vol. 49, no. 4, pp. 1523–1543, 2011.

[16] M. Eisen, K. Gatsis, G. J. Pappas, and A. Ribeiro, "Learning in wireless control systems over non-stationary channels," 2018, available at http://www.seas.upenn.edu/ $\sim$ maeisen/wiki/wirelesscontrol2.pdf.

[17] A. Goldsmith, *Wireless communications*. Cambr. Univ. Press, 2005.

[18] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010*. Springer, 2010, pp. 177–186.

[19] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of Machine Learning Research*, vol. 12, no. Jul, pp. 2121–2159, 2011.

[20] A. Defazio, F. Bach, and S. Lacoste-Julien, "Saga: A fast incremental gradient method with support for non-strongly convex composite objectives," in *Advances in Neural Information Processing Systems*, 2014, pp. 1646–1654.

[21] A. Ribeiro, "Ergodic stochastic optimization algorithms for wireless communication and networking," *IEEE Transactions on Signal Processing*, vol. 58, no. 12, pp. 6369–6386, 2010.

[22] O. Bousquet and L. Bottou, "The tradeoffs of large scale learning," in *Advances in neural information processing systems*, 2008, pp. 161–168.

[23] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013.

[24] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[25] A. B. Owen, "Self-concordance for empirical likelihood," *Canadian Journal of Statistics*, vol. 41, no. 3, pp. 387–397, 2013.

[26] I. Necoara and J. Suykens, "Interior-point lagrangian decomposition method for separable convex optimization," *Journal of Optimization Theory and Applications*, vol. 143, no. 3, p. 567, 2009.