Learning the Statistics of Wireless Links
Q-learning approach

Sina Firouzabadi

December 11, 2009

1 Introduction

Here we will present the problem formulation for our network optimization problem in wireless context. The idea is to solve this problem using a technique called Q-learning. This technique is based on stochastic approximation theory and deals with stochastic optimization problems without prior knowledge on the underlying distributions. The algorithm tries to learn the statistics by sampling from the channel states and updating the policies accordingly. In our formulation we deal with an optimization problem where we have some constraints on the expected value of the optimization variables. We tackle this problem with a tool from reinforcement learning literature called Full Recurse Optimization with Expected Constraints, FROEC. This approach is an online discrete time approach for optimization. It takes as input a sequence of samples from the randomness of the network and produces as its output estimates of the optimal policy values. The heart of FROEC is to use stochastic approximation to learn the statistics of the problem and update the control variables in order to get closer to the optimal solution in time.

The main premise of this research is to find optimal adaptive policies to manage self-interfering networks under power, rate and reliability constraints, in a distributed fashion. To be realistic, the inherent decentralized nature of wireless networks mandates that distributed network algorithms be developed to implement our joint optimization. Seeking to find distributed algorithms for specific optimization problems is challenging in general. The main idea that we have now, is based on the decomposition of the problem in both the primal and the dual domain into simpler subproblems and trying to combine the results based on the primal-dual principle. The decomposition of optimization problems via primal-dual methods has a long history. Using such a decomposition, many large-scale optimization problems can be solved via iterative algorithms that can be implemented in a distributed fashion, such as in [2]. In other cases, primal-dual methods allow the decomposition of the original problem into independent and simpler subproblems that otherwise would have to be solved jointly. This rationale has been widely adopted to obtain viable design methods and distributed implementation of optimal network paradigms, such as, for instance, in [4]. The paper [3] provides a recent overview, including a rich collection of references on the subject of decomposition via primal-dual principles.

In the remainder of this report, I will describe the problem formally for a simple case where we only have one link and show how one can use Q-leaning techniques to solve the aforementioned stochastic optimization problem in a distributed fashion.

2 Problem formulation

Suppose that we have M logical source/destination pairs and L links in the network. Each source and destination is associated with an upper layer protocol stack. The flow of information over the network from a logical source to a logical destination, possibly over multiple links, is termed an information flow. Flows from different sources m may traverse the same link l. The routing of information flows over links is described by the routing matrix A, where $A(l,m) = 1$ if information on flow m traverses link l and is otherwise zero.

For the m'th data session, assume r_m denotes the rate of information sent into the link encoder. The encoder uses block convolutional codes, which add additional bits to the information flow to enhance error detection and recovery. The ratio of the total number of useful information bits to the total number of bits exiting the encoder per unit time is termed the code rate $0 \leq \theta_l \leq 1$. Encoded bits are removed from the link buffer and transmitted by the wireless link at rate R_l . Therefore, he rate at which useful information is transmitted across the link is $\theta_l R_l$.

The channel is modeled by a channel state (gain) matrix $G \in \mathbf{R}^{L\times L}$, where G_{ij} is the power gain from the transmitter on link j to the receiver on link i. We assume that this channel state Matrix G is an stationary and ergodic random process with some distribution which is unknown to the network. We also assume the channel state is estimated without error and is known at the set of transmitters. The vector of transmitter powers is given by $\underline{S} \in \mathbf{R}^L$.[1] Each transmitter, say at link l, has an average power budget $\bar{S}_l$. For simplicity, the link rate function is assumed to be of the form

$$R_l(\underline{S}, G) = \log\left(\frac{KG_{ll}S_l}{\sum_{j\neq l} G_{lj}S_j + N_l}\right) \qquad l = 1, ..., L \tag{1}$$

Where K is is fixed and scales the received power [6] and N_l is the variance of the noise at the receiver of l'th link. Clearly, this rate formulation has an underlying high SINR assumption. Note that we could use the general rate formula in our problem formulation. In other words, the underlying assumption of high signal to noise ratio can be lifted and the optimization problem would still remain in the realm of convex optimization. This can be done, simply by adopting exactly the same trick proposed in [5] for overcoming the non-convexity of the capacity formula, with no assumption on high SINR. However, Using this trick makes the equations messy. Therefore, for the sake of simplicity, hereon we will use the high SINR assumption which is also valid for many practical scenarios.

The error probability of bits flowing over the link is defined as $X(\theta)$ and is assumed to be an increasing function of the code rate θ . Explicit expressions for $X(\theta)$ are difficult to find and we use the upper bound

$$X(\theta) = \frac{1}{2}2^{-N(R_0-\theta)} \tag{2}$$

where N is the code block length used by the encoder and R_0 is the cutoff rate [8]. Then, the reliability of an information flow m is defined [1] by ϕ_m as follows :

$$\underline{\phi} \;\; \leq \;\; 1 - A^T \underline{X}(\theta) \tag{3}$$

where $A^T \underline{X}(\theta)$ is the sum of the error rates on the links traversed by the flow.

The performance of upper layer protocols are modeled as utility functions. Each source m has a utility function $U(r_m, \phi_m)$. Utility functions are strictly concave increasing functions of the information rate and information reliability. In this work, we use the following parameterized family of utility functions:

$$U(r, \phi) = \beta \log r + (1 - \beta) \log \phi \tag{4}$$

where $0 \leq \beta \leq 1$ weights the relative importance of information rate and reliability.

The system can adapt to changing channel conditions by estimating G and adapting parameters such as transmit power $S = S(G)$, transmitter link rate $R = R(S(G), G)$, the information rate $r = r(G)$, code rate $\theta(G)$ and information reliability $\phi(G)$.

The following is the specification of an instance of WNUM problem that we will address hereafter:

Problem 2.1. *(**W-NUM Problem**) Given the above definitions, find adaptive*[2] *rate vector $\underline{r}(G)$, reliability vector $\underline{\phi}(G)$ and power policies $\underline{S}(G)$ that maximize the average utility of the network, under constraints on information rates, link rates, reliability and average power transmitted, in the following sense*

$$\begin{aligned}
\textit{Maximize :} \quad & \boldsymbol{E}\left[\sum_m U_m(r_m(G), \phi_m(G))\right] && (5)\\
\textit{Subject to :} \quad & \boldsymbol{E}\left[S_l(G)\right] \leq \bar{S}_l \quad l \in \{1, 2, ...L\} && (6)\\
& \boldsymbol{E}\left[A\, \underline{r}\,\right] \leq \boldsymbol{E}\left[\, \textit{Diag}(\underline{\theta}(G)) \;\; \underline{R}(S(G), G)\,\right] && (7)\\
& \boldsymbol{E}\left[\, \underline{\phi}(G)\,\right] \leq 1 - \boldsymbol{E}\left[A^T \;\underline{X}(\theta(G))\,\right] && (8)\\
& 0 < \underline{\theta}(G) < 1 && (9)\\
& 0 < \underline{\phi}(G) < 1 && (10)
\end{aligned}$$

[1] in this work we reserved underlined letters for representing variables which are in vector form

[2] with respect to changes in the channel state

Where **E** *is the expectation operator and optimization variables are* $\underline{S}(G), \underline{r}(G), \underline{\theta}(G), \underline{\phi}(G)$.

Note that the aforementioned problem is not in the class of convex optimization problems, directly. Yet, using a set of change of variables and reformulating the rate constraints of the links in (7) we first want to show the above optimization problem can be casted as a convex optimization problem. In particular if we define $\hat{S}(G) = \log(S(G))$, $\hat{\phi}(G) = \log(\phi(G))$ and $\hat{r}(G) = \log(r(G))$, we can re-write our WNUM as :

$$
\begin{aligned}
\text{maximize :} \quad & \mathbf{E}\left[\sum_m \hat{U}_m(\hat{r}_m(G), \hat{\phi}_m(G))\right] && (11)\\
\text{subject to :} \quad & \mathbf{E}\left[exp(\hat{S}_l(G))\right] \le \bar{S}_l \ \ , l \in \{1, 2, ...L\} && (12)\\
& \mathbf{E}\left[\ \underline{\pi}(G)\ \right] \le \mathbf{E}\left[\ \underline{R}(\ \underline{S}(G)\ , G)\right] && (13)\\
& \log(A(l, :)\ \underline{r}\ ) \le \log(\ \theta_l(G)) + \log(\pi_l(G))\ , \quad l \in \{1, 2, ...L\} && (14)\\
& \mathbf{E}\left[\underline{\hat{\phi}}(G)\right] \le 1 - \mathbf{E}\left[A^T\ \underline{X}\,(\underline{\theta}(G))\right] && (15)\\
& 0 < \underline{\theta}(G) < 1 && (16)\\
& 0 < \underline{\phi}(G) < 1 && (17)
\end{aligned}
$$

Where (13),(14) is replaced for rate constraints (7) by introducing auxiliary vector variable $\underline{\pi}(G)$. We can solve this problem by utilizing Full Recurse Optimization with Expected Constraints, FROEC. This approach is an online discrete time approach for optimization. It takes as input the sequence of channel states seen by the network and produces as its output estimates of the optimal policy values. The heart of FROEC is to use stochastic approximation to learn the statistics of the problem and update the control variables in order to get closer to the optimal solution in time. As a byproduct FROEC produces the optimal Lagrange multipliers associated with constraints in (12)-(17). The time index is t, and we indicate the estimates of optimal Lagrange multiplier λ^* by λ^t. Policy values are denoted by $\underline{r}^t = \underline{r}(G^t, \underline{\lambda}^t), \underline{S}^t = \underline{S}(G^t, \underline{\lambda}^t)$, and $\underline{R}^t = \underline{R}(G^t, \underline{\lambda}^t)$, etc. FROEC does not assume knowledge of $p(G)$, the distribution of channel state matrix, and under suitable conditions adjusts to changes in the channels empirical distribution.

FROEC solves the dual problem to (11) Where the dual function is defined as

$$
g(\ \underline{\lambda}\ ) = \max_{\Omega}\ L\ (\ \underline{r}(G)\ ,\ \underline{S}(G)\ ,\ \underline{\pi}(G)\ ,\ \underline{\phi}(G)\ ,\ \underline{\theta}(G)\ ,\ \underline{\lambda}\ ) \qquad (18)
$$

Where $\Omega = \{\ \underline{r}(G)\ ,\ \underline{S}(G)\ ,\ 0 < \underline{\phi}(G) < 1\ ,\ 0 < \underline{\theta}(G) < 1\ \}$ and

$$
\begin{aligned}
L(.) = \mathbf{E}\Bigg[\sum_m \hat{U}_m(\hat{r}_m, \hat{\phi}_m) - \sum_{l=1}^{L} \lambda_{S,l}\left(e^{\hat{S}_l} - \bar{S}_l\right) - \sum_{l=1}^{L} \lambda_{\pi,l}\,(\pi_l - R_l(S, G))\\
- \sum_{l=1}^{L} \lambda_{q,l}\,(\log(A(l, :)\ \underline{r}\ ) - \log(\theta_l) - \log(\pi_l)) - \\
\sum_{m=1}^{M} \lambda_{\phi,m}\left(\hat{\phi}_l - 1 + A(:, m)^T\ \underline{X}\,(\theta)\right)\Bigg] \qquad (19)
\end{aligned}
$$

We define $\underline{\lambda} = [\ \underline{\lambda_S}\ ,\ \underline{\lambda_\pi}\ ,\ \underline{\lambda_q}\ ,\ \underline{\lambda_\pi}\ ]$ as the vector of Lagrange multipliers. Now our original problem is equivalent to solve the following problem in the dual domain:

$$
\min_{\underline{\lambda}\ \ge 0} g(\ \underline{\lambda}\ ) \qquad (20)
$$

The method proposed in [1] for solving this problem is to solve this dual problem with a gradient decent method by updating the Lagrange multipliers[3] gradually. Utilizing Stochastic Approximation, it has been shown in [1] that

[3] prices

by neglecting the expected value at each iteration and solving (19) for a fixed measured G at each time step, we can still update the prices and converge to the optimal solution. The whole premise of our work here is to observe that in each time step, we can have one step further and solve the Lagrangian (19) within an internal loop via another gradient decent method. The solution we propose is to use the primal-dual algorithm. The primal step consists of an optimization problem that solves (19) iteratively via a gradient decent method, and the dual step is a simple price[4] update. Both of these steps calculate the subgradient from the sampled channel gain and update their variables. In this case it can be seen that in solving (19) iteratively, each variable needs local information to update itself and converge to the optimum and in this way the whole optimization problem can be decomposed. In the next section we will illustrate the performance of this primal-dual approach and the learning rate of the algorithm.

3 Simulations

In this section, we illustrate the performance of the primal-dual approach and its convergence via simulations. In particular, we illustrate the steady state performance of primal-dual optimal policies and finally check the convergence of the algorithm to the global maximum by checking the KKT conditions. Here we consider a wireless network with $L = 10$ interfering links with $m = 5$ data sessions and use the utility function described by (4). In order to illustrate the distributed implementation, we also assume that the first five links are physically apart from the last five links so that these two groups of link do no interfere with each other. The channel state matrix G is drawn iid Rayleigh with the diagonal elements scaled to yield an average SINR of $20dB$ over all links. The transmitter power limit is $\bar{S} = 10$. For the sake of simplicity, we adopt the link rate based on the high SINR regime; Therefore, we use (7) as the link rate function.

Figure 3 show the rates of each data session. As we can see, the proposed primal-dual approach is adapting the link rates to changing channel conditions and finally reaches a steady state value. Validating the final rates in order to check whether they are optimal values or not is also tricky. This is mainly due to the fact that finding the numerical solution of the WNUM problem when we have expectation in the constraints is not easy, even if G is drawn from a set with a finite number of channel states. But, since we already have a candidate for the optimal values from our primal-dual algorithm and we only need to validate it, we can check the KKT conditions for the obtained solution. Figure 3 illustrates one of the KKT conditions associated with one of the rate constraints. This figure shows the difference between the expected value of the capacity with respect to different scenarios for G and the allocated rate, i.e the slack in constraint (7). Figure 3 also show some of the Lagrange Multipliers associated with the rate constraints and their convergence to a fixed value.

4 Acknowledgement

I would like to thank prof. Andrea Goldmith and also Dan Oneil for helpful discussions we had regarding this work.

References

[1] D. ONeill, Boon Sim Thian, A.J. Goldsmith, and S.P. Boyd, *"Wireless NUM: Rate and Reliability Tradeoffs in Random Environments"* IEEE Wireless Communications and Networking Conference, 2009, Budapest Hungary

[2] F. Kelly, A. Mauloo and D. Tan, *"Rate control for communication networks: shadow prices, proportional fairness and stability,"* Journal of Operational Research Society, vol 49, no 3, pp. 237-252, Mar 1998

[3] M. Chiang, S. H. Low, A. R. Calderbank, J. C. Doyle, *"Layering as optimization decomposition: questions and answers,"* Proc. IEEE MILCOM, Washington D.C., October 2006

[4] F. Paganini, W. Wang, J. C. Doyle, S. H. Low, *"Congestion control for high performance, stability, and fairness in general networks,"* IEEE/ACM Transactions on Networking, Volume 13, Issue 1, Feb. 2005 Page(s):43 - 56

[4] We denote the dual variables as prices.

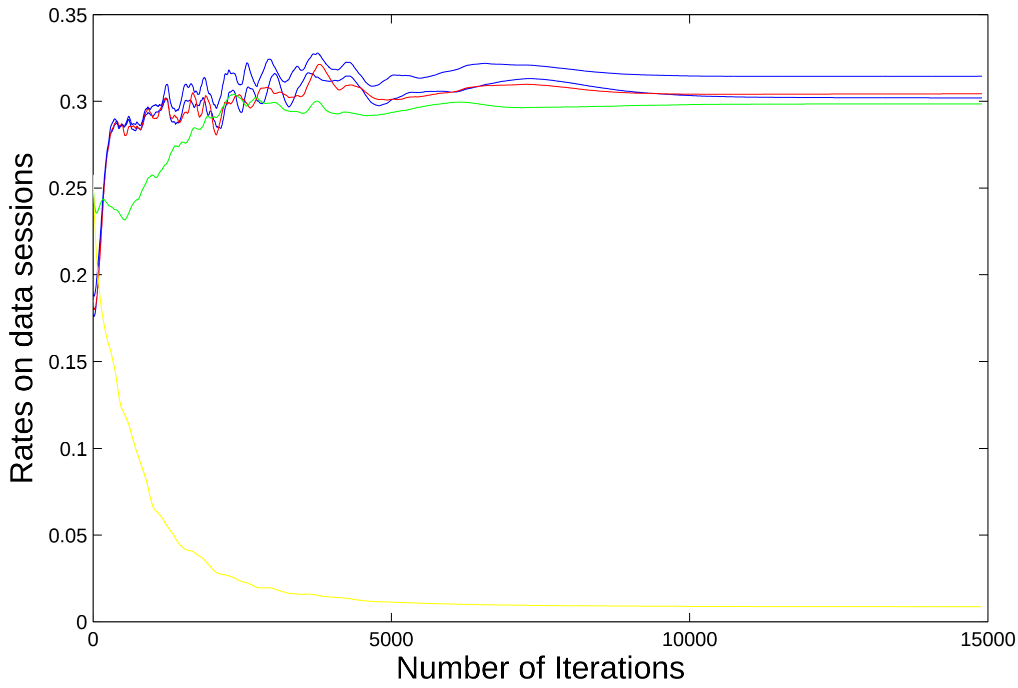

Figure 1: Rates on data sessions

[5] John Papandriopoulos, Subhrakanti Dey and Jamie S. Evans, 'Distributed Cross-Layer Optimization of MANETs in Composite Fading', IEEE International Conference on Communications, Istanbul, Turkey, Jun. 2006.

[6] G. J. Foschini and J. Salz, Digital communications over fading radio channels, Bell Syst. Tech. J., pp. 429456, Feb. 1983.

[7] V. Borker and S. Meyn, The o.d.e. method for convergence of stochastic approximation and reinforement learning, SIAM Journal of Control Optimization, vol. 38,

[8] J. G. Proakis, Ed., Digital Communications. New York: McGraw-Hill, 1989.

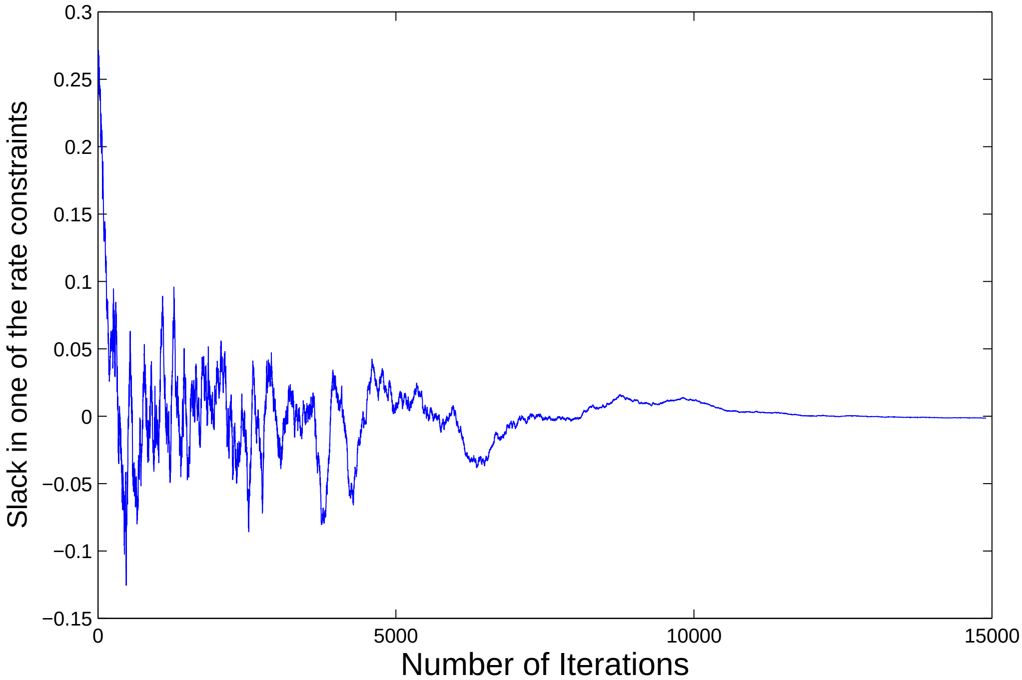

Figure 2: Slack in one of the rate constraints

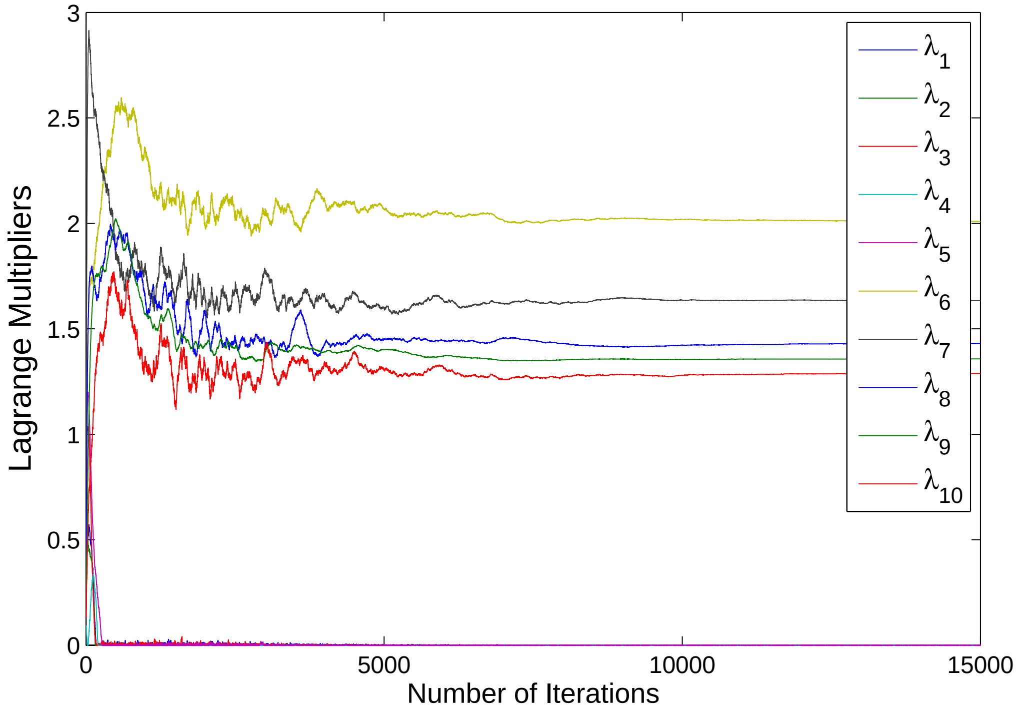

Figure 3: some of the Lagrange Multipliers