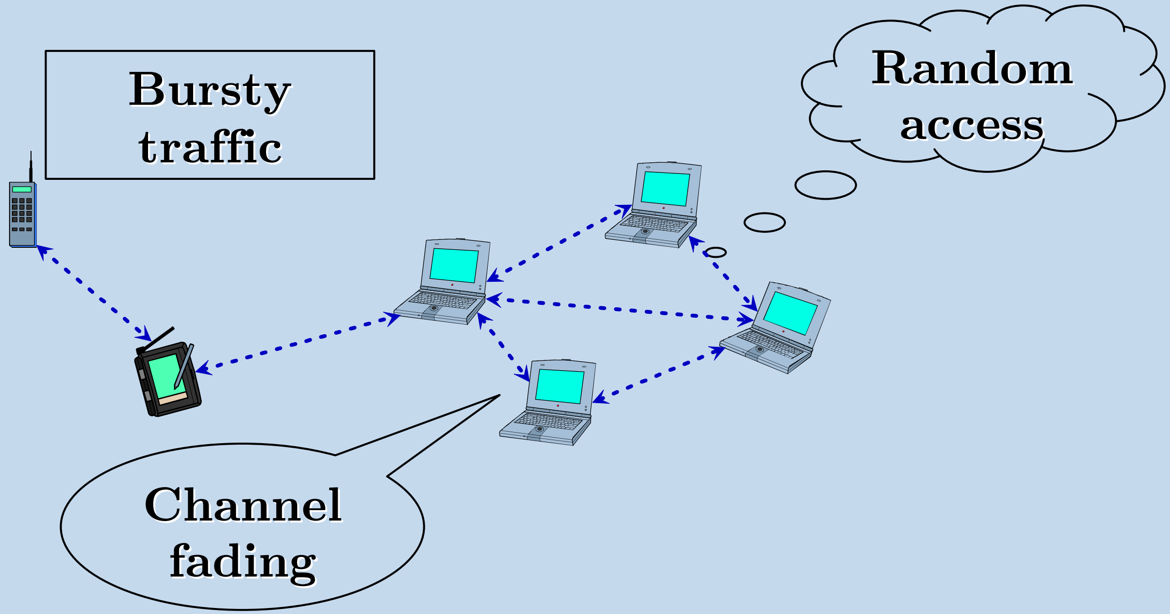
Cross-Layer Rate Control in Multi-hop Networks: Noisy Feedback, Fairness and Stochastic Stability

Junshan Zhang

Department of Electrical Engineering
Arizona State University
May 22, 2006

Multi-hop Wireless Networks



Stochastic attributes of wireless networks:

- time-varying channel conditions and co-channel interference;
- random access;
- bursty traffic flows;
- Mobility, dynamic network topology;
- …

Objective and Approach

•Goal: fair rate control through joint congestion control and MAC design in multi-hop random access networks.

•Approach: network utility maximization

- Treat rate control as a utility maximization problem
- Different layers function “cooperatively” to achieve the optimum point (equilibrium point).

Basic Setting

- Consider a wireless network modelled as a directed graph $G = (N, E)$;
- Let $U_s(x_s)$ denote the utility function of flow s, where x_s is the flow rate;
- Consider persistence scheme with xmission prob. $\{p_{(i,j)}, \forall(i,j) \in E\}$.
- Let $N^I_{to}(i)$ denote the set of nodes whose transmissions interfere with node i's reception; $N_{to}(i) = N^I_{to}(i) \cup i$.
- Use $N_{in}(i)$ to denote the set of the nodes from which node i receives traffic, $N_{out}(i)$ to denote the set of nodes to which node i is sending packets and $N^I_{from}(i)$ to denote the set of nodes whose reception is interfered by node i's transmission.

Problem Formulation

$$
\begin{aligned}
\boldsymbol{\Xi} : \max_{\{x_s\}} \quad & \sum_{s \in \mathcal{S}} U_s(x_s) \\
\text{subject to} \quad & \sum_{s \in \mathcal{S}((i,j))} x_s \le c_{(i,j)}(p_{(i,j)} \prod_{k \in N_{to}(j)} (1 - P_k)), \ \forall \ (i,j) \\
& \sum_{j \in N_{out}(i)} p_{(i,j)} = P_i, \ \forall \ i \\
& 0 \le x_s \le M_s, \ \forall \ s \\
& 0 \le P_i \le 1, \ \forall \ i,
\end{aligned}
$$

where M_s is the maximum for flow data rate of s, $c_{(i,j)}$ is the average transmission rate and the utility function $U_s(\cdot)$ takes the form [Mo, Walrand 00]

$$
U_\kappa(r_i) = \begin{cases} w_i \log r_i, & \text{if } \kappa = 1 \\ w_i (1-\kappa)^{-1} r_i^{1-\kappa}, & \textit{otherwise.} \end{cases}
$$

Observation: Problem $\boldsymbol{\Xi}$ is non-convex and non-separable.

Related work

1. Internet TCP/AQM: [Kelly, Maulloo, Tan 98], [Low, Lapsley 99], [Srikant 05] and many many more (sorry due to limited space ☺);

2. Joint flow control/routing/MAC/PHY design in wireless networks: [Chiang 04], [Lin-Shroff 05], [Wang-Kar 05], [Eryilmaz-Srikant05] [Neely-Modiano-Li 05] [Chen-Low-Chiang-Doyle 06] ….

<u>A common assumption:</u>

Most works above assume deterministic feedback in their distributed algorithms.

Outline

I: Single time-scale algorithm based on dual decomposition ($\kappa \geq 1$):

1. **Deterministic P-D algorithm;**
2. **Stochastic P-D algorithm under noisy feedback:**
   - **Unbiased case: convergence, rate of convergence;**
   - **Biased case: contraction region, stochastic stability.**

II: Two time-scale algorithms via alternative decompositions ($\kappa \geq 0$):

- **Feasible direction method;**
- **Primal decomposition method;**
- **Stability for unbiased case.**

Acknowledgement:
Part I based on CISS'06 paper with Dong Zheng;
Part II based on ongoing collaboration with Mung Chiang.

Deterministic P-D Algorithm: $\kappa \geq 1$

Let $\tilde{x}_s = \log(x_s)$ [Lee, Chiang, Calderbank 06], we have

$$
\begin{array}{ll}
\mathbf{P} : \max\limits_{\{\tilde{x}_s\}} & \sum\limits_{s \in \mathcal{S}} U'_s(\tilde{x}_s) \\
\text{subject to} & \log(\sum_{s \in \mathcal{S}((i,j))} \exp(\tilde{x}_s)) - \log(p_{(i,j)}) \\
& - \sum_{k \in N_{to}(j)} \log(1 - P_k) \leq 0, \ \forall \ (i,j) \\
& \sum_{j \in N_{out}(i)} p_{(i,j)} = P_i, \ \forall \ i \\
& -\infty \leq \tilde{x}_s \leq \tilde{M}_s, \ \forall \ s \\
& 0 \leq P_i \leq 1, \ \forall \ i,
\end{array}
$$

where $U'_s(\tilde{x}_s) = U_s(\exp(\tilde{x}_s))$.

Observation: Problem $\mathbf{P}$ is convex and separable if $\kappa \geq 1$.

Lagrange Dual Approach

The Lagrangian function is

$$L(\tilde{\mathbf{x}}, \mathbf{p}, \lambda) = \left\{ \sum_s U'_s(\tilde{x}_s) - \sum_{(i,j)} \lambda_{(i,j)} \log \left(\sum_{s \in \mathcal{S}((i,j))} \exp(\tilde{x}_s) \right) \right\} + \sum_{(i,j)} \lambda_{(i,j)} \log \left(p_{(i,j)} \prod_{k \in N^I_{to}(j)} (1 - P_k) \right)$$

Then, the Lagrange dual function is

$$Q(\lambda) = \max_{\substack{\sum_{j \in N_{out}(i)} p_{(i,j)} = P_i \\ \mathbf{0} \leq \mathbf{P} \leq \mathbf{1} \\ -\infty \leq \tilde{\mathbf{x}} \leq \tilde{\mathbf{M}}}} L(\tilde{\mathbf{x}}, \mathbf{p}, \lambda),$$

and the dual problem is given by

$$\mathbf{D} : \min_{\lambda \geq \mathbf{0}} \quad Q(\lambda)$$

Strong Duality

Proposition:

a) *There is no duality gap, i.e., the minimum value of the dual problem* $\mathbf{D}$ *is equal to the maximal value of the primal problem* $\mathbf{P}$.

b) *Let* Φ *be the set of* λ *that minimizes* $Q(\lambda)$. *Then* Φ *is non-empty and compact; and* $\forall \lambda \in \Phi$, *there exists a unique vector* $(\tilde{\mathbf{x}}^*, \mathbf{p}^*)$ *that maximizes the Lagrangian function* $L(\cdot, \cdot, \cdot)$.

Distributed Primal-dual Alg.

- *The source rates are updated by*

$$\tilde{x}_s(n{+}1) = \left[\tilde{x}_s(n) + \epsilon_n \underbrace{\left(\dot{U}_s'(\tilde{x}_s(n)) - \exp(\tilde{x}_s(n)) \sum_{(i,j)\in\mathcal{L}(s)} \frac{\lambda_{(i,j)(n)}}{\sum_{s\in\mathcal{S}((i,j))} \exp(\tilde{x}_s(n))} \right)}_{\triangleq L_{\tilde{x}_s}(\tilde{\mathbf{x}}(n),\mathbf{p}(n),\lambda(n))} \right]_{-\infty}^{\tilde{M}_s} .$$

- *The shadow prices are updated by*

$$\lambda_{(i,j)}(n{+}1) = \left[\lambda_{(i,j)}(n) - \epsilon_n \underbrace{\left(\log(p_{(i,j)}(n)) + \sum_{k\in N_{to}^I(j)} \log(1 - P_k(n)) - \log\left(\sum_{s\in\mathcal{S}((i,j))} \exp(\tilde{x}_s(n)) \right) \right)}_{L_{\lambda_{(i,j)}}(\tilde{\mathbf{x}}(n),\mathbf{p}(n),\lambda(n))} \right]_0^{\infty} .$$

- *The persistence probabilities are updated by*

$$p_{(i,j)}(n+1) = \frac{\lambda_{(i,j)}(n)}{\sum_{k\in N_{out}(i)} \lambda_{(i,k)}(n) + \sum_{(l,m):m\in N_{from}^I(i),l\in N_{in}(m)(n)} \lambda_{(l,m)}}.$$

Stochastic Stability

Feedback information is needed to compute the gradients; Unfortunately, the feedback is based on error-prone measurement mechanisms and is noisy in practical systems!

A fundamental open question:

Q) What's the impact of noisy feedback on network utility maximization?

Stochastic stability is pertinent to following issues:
1. **number of users and/or queuing length remain finite;**
2. **algorithms converge in some stochastic sense.**

Related work on stability:

- **Connection-level randomness: [Bonald-Massoulie 01], and [Lin-Shroff 05] [Eryilmaz-Srikant05] [Neely-Modiano-Li 05] [Stolyar 05] and more ;**
- **Rate of convergence around the equilibrium points: [Kelly-Malloo-Tan 98], [Kelly-03], … ;**
- **Deterministic feedback error: [Mehyar-Spanos-Low 04]**

Stochastic Primal-Dual Alg.

Our motivation is to bring back packet-level dynamics and understand convergence of the following stochastic alg.

- *SA algorithm for source rate updating*:

$$\tilde{x}_s(n+1) = [\tilde{x}_s(n) + \epsilon_n \left(\hat{L}_{\tilde{x}_s}(\tilde{\mathbf{x}}(n), \mathbf{p}(n), \lambda(n)) \right)]_{-\infty}^{\tilde{M}_s}.$$

- *SA algorithm for shadow price updating*:

$$\lambda_{(i,j)}(n+1) = \left[\lambda_{(i,j)}(n) - \epsilon_n (\hat{L}_{\lambda_{(i,j)}}(\tilde{\mathbf{x}}(n), \mathbf{p}(n), \lambda(n))) \right]_0^{\infty}.$$

- *The persistence probability updating rule remains the same.*

Structure of Stochastic Gradients

Stochastic gradient $\hat{L}_{\tilde{x}_s}(\cdot,\cdot,\cdot)$: Observe that

$$\hat{L}_{\tilde{x}_s}\left(\tilde{\mathbf{x}}(n), \mathbf{p}(n), \lambda(n)\right) = L_{\tilde{x}_s}\left(\tilde{\mathbf{x}}(n), \mathbf{p}(n), \lambda(n)\right) + \alpha_s(n) + \zeta_s(n),$$

where $\alpha_s(n)$ is the biased random error in $L_{\tilde{x}_s}\left(\tilde{\mathbf{x}}(n), \mathbf{p}(n), \lambda(n)\right)$, given by

$$\alpha_s(n) \triangleq E_n\left[\hat{L}_{\tilde{x}_s}\left(\tilde{\mathbf{x}}(n), \mathbf{p}(n), \lambda(n)\right)\right] - L_{\tilde{x}_s}\left(\tilde{\mathbf{x}}(n), \mathbf{p}(n), \lambda(n)\right),$$

and $\zeta_s(n)$ is a martingale difference noise

$$\zeta_s(n) \triangleq \hat{L}_{\tilde{x}_s}\left(\tilde{\mathbf{x}}(n), \mathbf{p}(n), \lambda(n)\right) - E_n\left[\hat{L}_{\tilde{x}_s}\left(\tilde{\mathbf{x}}(n), \mathbf{p}(n), \lambda(n)\right)\right].$$

Structure of Stochastic Gradients (Cont'd)

Stochastic gradient $\hat{L}_{\lambda_{(i,j)}}(\cdot,\cdot,\cdot)$: Observe that

$$\hat{L}_{\lambda_{(i,j)}}\left(\tilde{\mathbf{x}}(n),\mathbf{p}(n),\lambda(n)\right) = L_{\lambda_{(i,j)}}\left(\tilde{\mathbf{x}}(n),\mathbf{p}(n),\lambda(n)\right)) + \beta_{(i,j)}(n) + \xi_{(i,j)}(n),$$

where $\beta_{(i,j)}(n)$ is the biased random error of $L_{\lambda_{(i,j)}}\left(\tilde{\mathbf{x}}(n),\mathbf{p}(n),\lambda(n)\right)$, given by

$$\beta_{(i,j)}(n) \triangleq E_n\left[\hat{L}_{\lambda_{(i,j)}}\left(\tilde{\mathbf{x}}(n),\mathbf{p}(n),\lambda(n)\right)\right] - L_{\lambda_{(i,j)}}\left(\tilde{\mathbf{x}}(n),\mathbf{p}(n),\lambda(n)\right),$$

and $\xi_{(i,j)}(n)$ is a martingale difference noise:

$$\xi_{(i,j)}(n) \triangleq \hat{L}_{\lambda_{(i,j)}}\left(\tilde{\mathbf{x}}(n),\mathbf{p}(n),\lambda(n)\right) - E_n\left[\hat{L}_{\lambda_{(i,j)}}\left(\tilde{\mathbf{x}}(n),\mathbf{p}(n),\lambda(n)\right)\right].$$

Technical Assumptions

A1. *We assume that the estimators of the gradients are based on the measurements in each iteration only.*

A2. *Condition on the step size:*
$\epsilon_n > 0,\ \epsilon_n \to 0,\ \sum_n \epsilon_n \to \infty$ *and* $\sum_n \epsilon_n^2 < \infty.$

A3. *Condition on the biased error:*
$\sum_n \epsilon_n |\alpha_s(n)| < \infty, \forall\ s$ *and* $\sum_n \epsilon_n |\beta_{(i,j)}(n)| < \infty, \forall\ (i,j).$

A4. *Condition on the martingale difference noise:*
$\sup_n E_n[\zeta_s(n)^2] < \infty, \forall\ s,$ *and* $\sup_n E_n[\xi_{(i,j)}(n)^2] < \infty, \forall\ (i,j).$

Main Result 1: Stability for Unbiased Case

Theorem 1:
***Under Conditions* $\mathbf{A1} - \mathbf{A4}$, *the iterates* $\{(\mathbf{x}(n), \lambda(n), \mathbf{p}(n)), n = 1, 2, \ldots\}$, *generated by the stochastic primal-dual algorithm, converge with probability one to the optimal solutions of Problem* $\mathbf{\Xi}$.**

1. *Good news:* The stochastic algorithm converges to the desired points under conditions A1 – A4.
2. *Caution:*

- When ϵ_n or the biased terms do not go to zero, we cannot hope to get convergence w.p.1.
- Even the expectation of the limiting distribution would not be the equilibrium point if biased terms do not go to zero.
- Nevertheless, we expect that the iterates would converge weakly to some neighborhood "close" to the equilibrium point.

An Example on Exponential Marking

- Assume exponential marking is used to feedback price information, where the overall non-marking probability $q = \exp\left(\sum_{(i,j)\in\mathcal{L}(s)} \frac{\lambda_{(i,j)}}{\sum_{s\in\mathcal{S}(i,j)} x_s}\right)$
- To estimate the overall price, source s sends N_n packets during round n and counts the non-marked packets, say K packets. Then the estimation is $\log(\hat{q})$, where $\hat{q} = K/N_n$.
- By definition, $\alpha_s(n) = \exp(\tilde{x}_s(n))\left(E_n[\log(\hat{q})] - \log(q)\right)$
- From **A2**, to ensure the convergence, it suffices to have that

$$\sum_n \frac{\epsilon_n}{\sqrt{N_n}} < \infty$$

e.g., when $\epsilon_n = 1/n$, $N_n \sim O(\log^4(n))$ is sufficient.

Stochastic Stability: Biased Case

1. When the gradient estimator is biased, we cannot hope for almost sure convergence;
2. Instead, we expect that the iterates return to a neighborhood of the optimal points under certain conditions.
3. Indeed, we can show that if the errors are uniformly bounded, there exists a "*contraction region*" such that the iterates return to this region infinitely often w.p.1.

Main Result 2: Stability for Biased Case

A7. *Condition on the biased error: There exist non-negative constants* $\{\alpha_s^u\}$ *and* $\{\beta_{(i,j)}^u\}$ *such that* $\limsup_n |\alpha_s(n)| \leq \alpha_s$, $\forall$ s *and* $\limsup_n |\beta_{(i,j)}(n)| \leq \beta_{(i,j)}$, $\forall$ (i,j).

Define the "contraction region" A_η as follows:

$$\begin{aligned} A_\eta \quad \triangleq \quad & \{(\mathbf{x}, \lambda) : \alpha_s \geq \eta |L_{\tilde{x}_s}(\tilde{\mathbf{x}}, \mathbf{p}, \lambda)|, \text{ for some } s, \\ & \text{or} \beta_{(i,j)} \geq \eta |L_{\lambda_{(i,j)}}(\tilde{\mathbf{x}}, \mathbf{p}, \lambda)|, \text{ for some } (i,j), 0 \leq \eta < 1\}. \end{aligned}$$

Theorem 2:

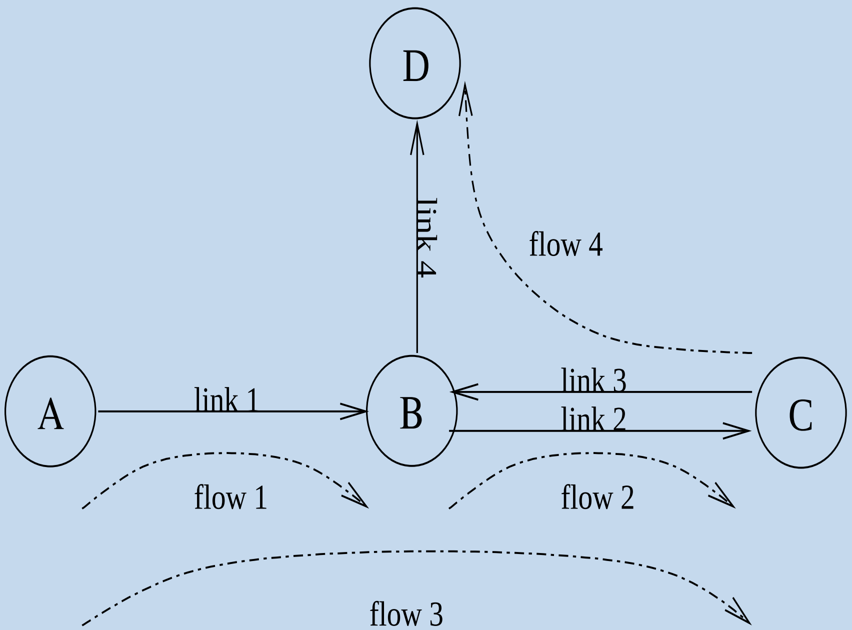
Under Conditions **A1** $-$ **A2**, **A4** and **A7**, the iterates $\{(\mathbf{x}(n), \lambda(n), \mathbf{p}(n)), n = 1, 2, \ldots\}$, generated by stochastic primal-dual algorithm, return to the set A_η infinitely often with probability one.

Remarks

1. Condition **A7** essentially requires that the biased terms are asymptotically bounded. Clearly, this is weaker than **A3**, and in this sense, the stability result is more general.

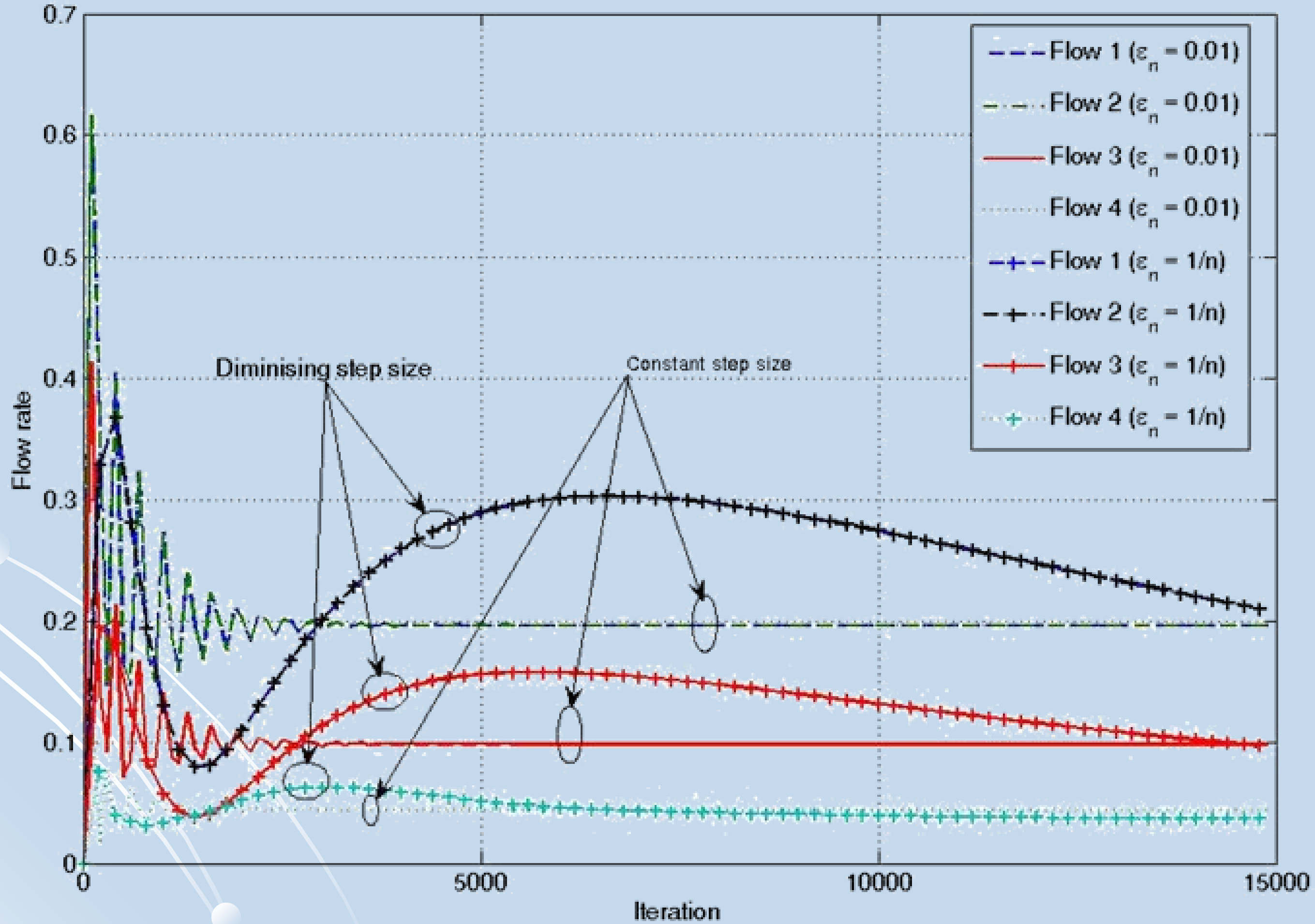
2. The "contraction region" A_η involves a set of nonlinear inequalities, and it is difficult to characterize A_η in closed-form. Numerical method is required to characterize "contraction region".

Numerical Studies

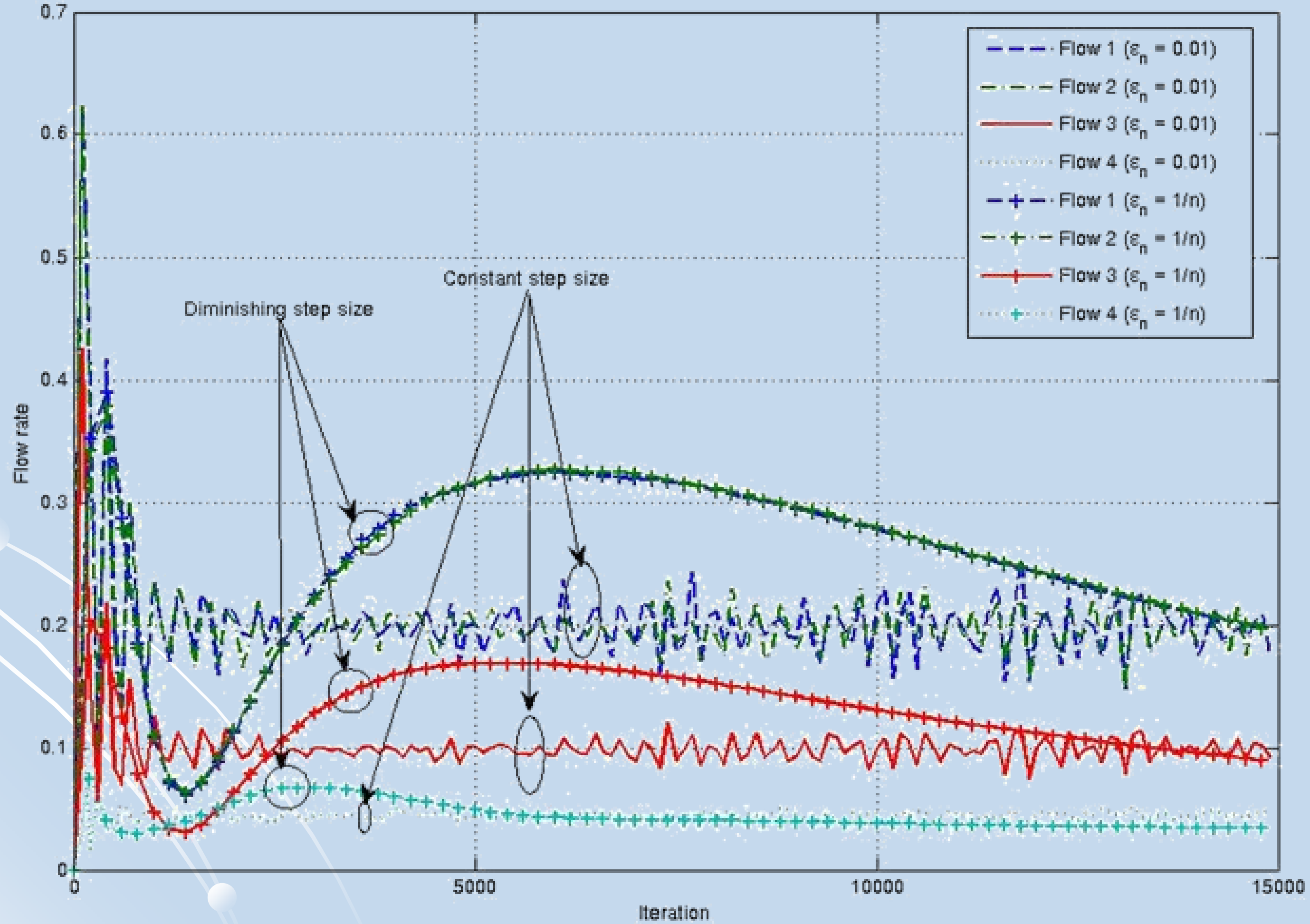


Consider logarithm utility functions where $\kappa = 1$.

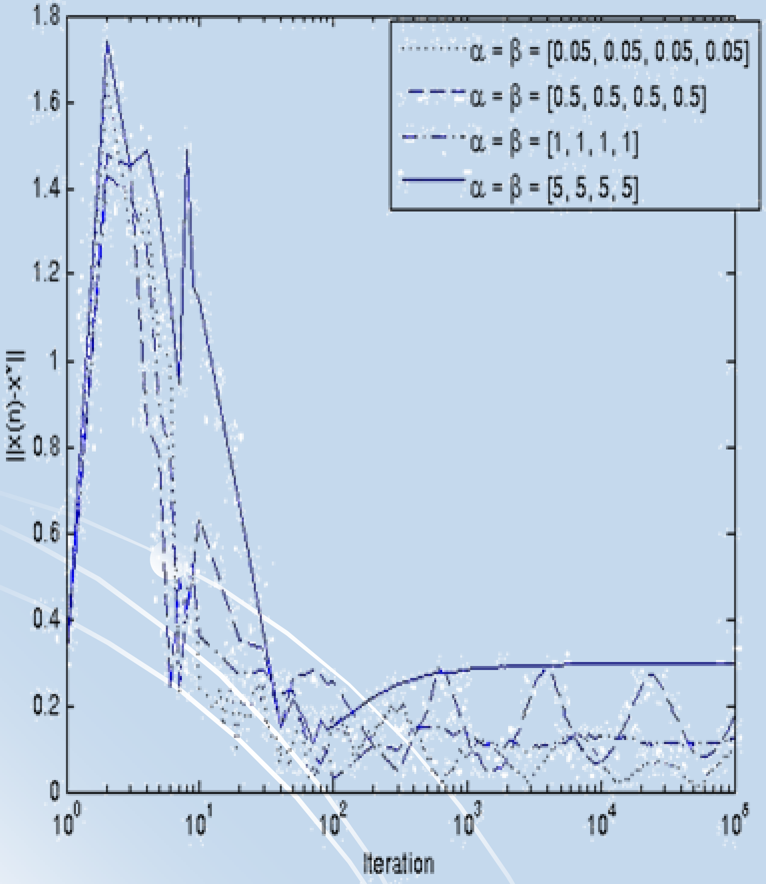
Numerical Examples – Deterministic Case

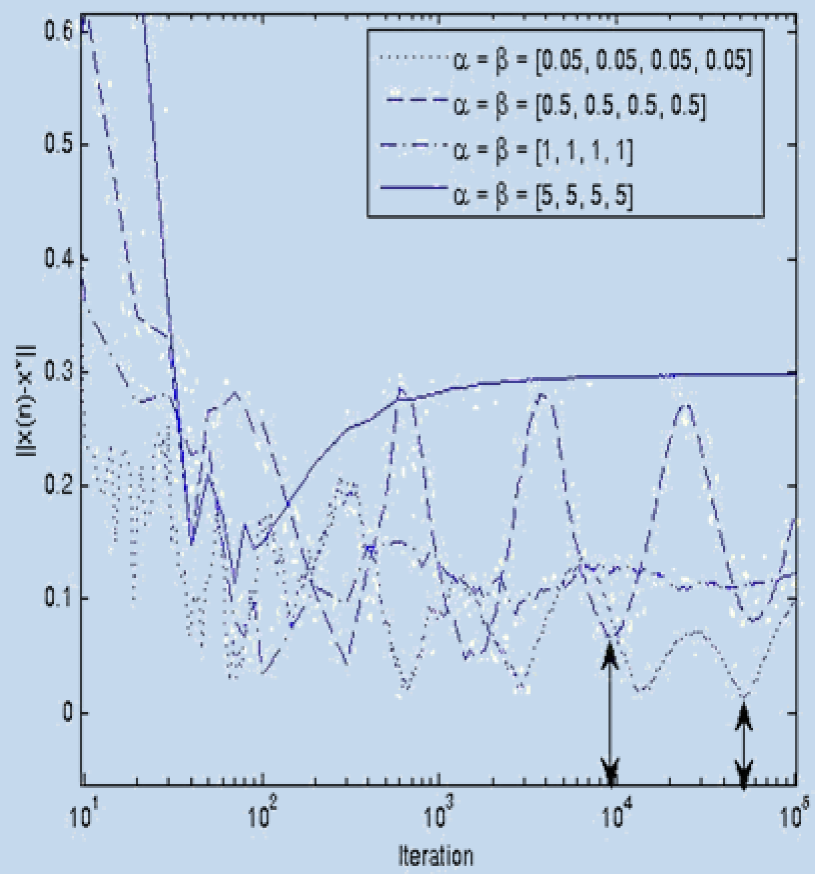


Numerical Examples – Stochastic (unbiased) Case



Numerical Examples: Biased Case





Proof for the Main Result 1

Step I: Using the stochastic Lyapunov Stability Theorem, we establish the recurrence of any arbitrary small neighborhood of the optimal point.

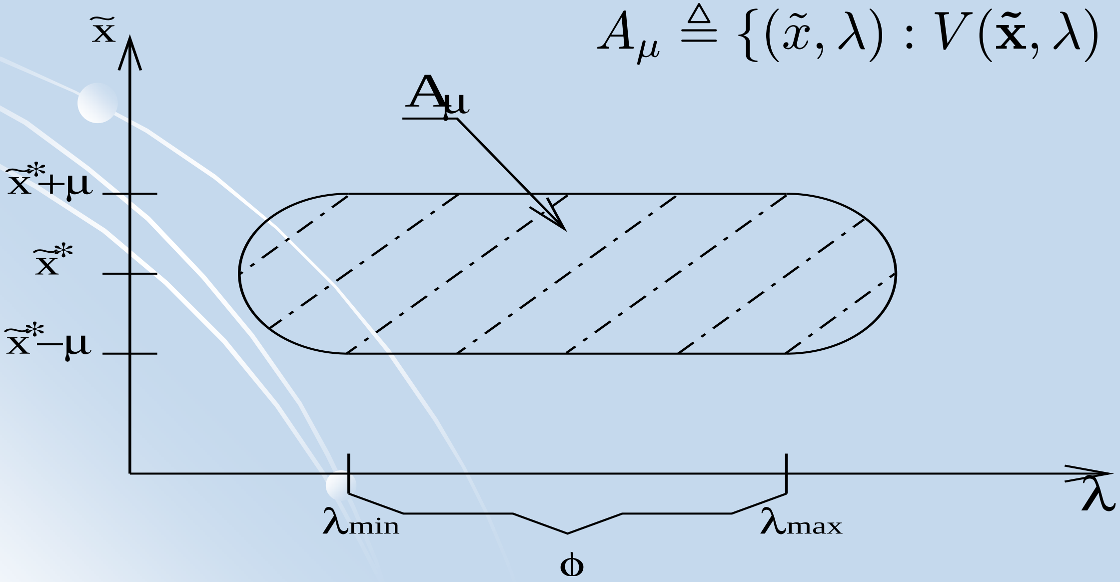
- Let $(\tilde{\mathbf{x}}^*, \mathbf{p}^*, \lambda^*)$ be a saddle point.

- Define the Lyapunov function $V(\cdot)$ as follows:

$$V(\tilde{\mathbf{x}}, \lambda) \triangleq ||\tilde{\mathbf{x}} - \tilde{\mathbf{x}}^*||^2 + \min_{\lambda^* \in \Phi} ||\lambda - \lambda^*||^2,$$

- Define for a given $\mu > 0$, a neighborhood set

$$A_\mu \triangleq \{(\tilde{x}, \lambda) : V(\tilde{\mathbf{x}}, \lambda) \leq \mu\}.$$



Stochastic Lyapunov Function

We show

$$\begin{aligned} E_n[V(\tilde{\mathbf{x}}(n+1), \lambda(n+1))] \quad &\leq \quad V(\tilde{\mathbf{x}}(n), \lambda(n)) + 2\epsilon_n G(\tilde{\mathbf{x}}(n), \lambda(n)) \\ &\quad + O\left(\epsilon_n(||\alpha(n)|| + ||\beta(n)||)\right) + O(\epsilon_n^2), \end{aligned}$$

with the understanding that

$$\begin{aligned} G(\tilde{\mathbf{x}}(n), \lambda(n)) \quad &= \quad (\tilde{\mathbf{x}}(n) - \tilde{\mathbf{x}}^*)^T L_{\tilde{x}}(\tilde{\mathbf{x}}(n), \mathbf{p}(n), \lambda(n)) \\ &\quad -(\lambda(n) - \lambda^*_{\min})^T L_\lambda(\tilde{\mathbf{x}}(n), \mathbf{p}(n), \lambda(n)), \end{aligned}$$

where $\lambda^*_{\min} = \arg\min_{\lambda \in \Phi} ||\lambda(n) - \lambda||^2$.

It suffices to show that $G(\tilde{\mathbf{x}}(n), \lambda(n)) < 0$ when $(\tilde{\mathbf{x}}(n), \lambda(n)) \in A^c_\mu$.

A Supermartingale Lemma

Let $\{X_n\}$ *be an* $\mathcal{R}^r$*-valued stochastic process, and* $V(\cdot)$ *be a real-valued and non-negative function on* $\mathcal{R}^r$*. Suppose that* $\{Y_n\}$ *is a sequence of random variables satisfying that* $\sum_n |Y_n| < \infty$ *with probability one. Let* $\{\mathcal{F}_n\}$ *be a sequence of* $\sigma-$*algebras, generated by* $\{X_i, Y_i, i \le n\}$*. Suppose that there exists a compact set* $A \subset \mathcal{R}^r$ *such that for all* n

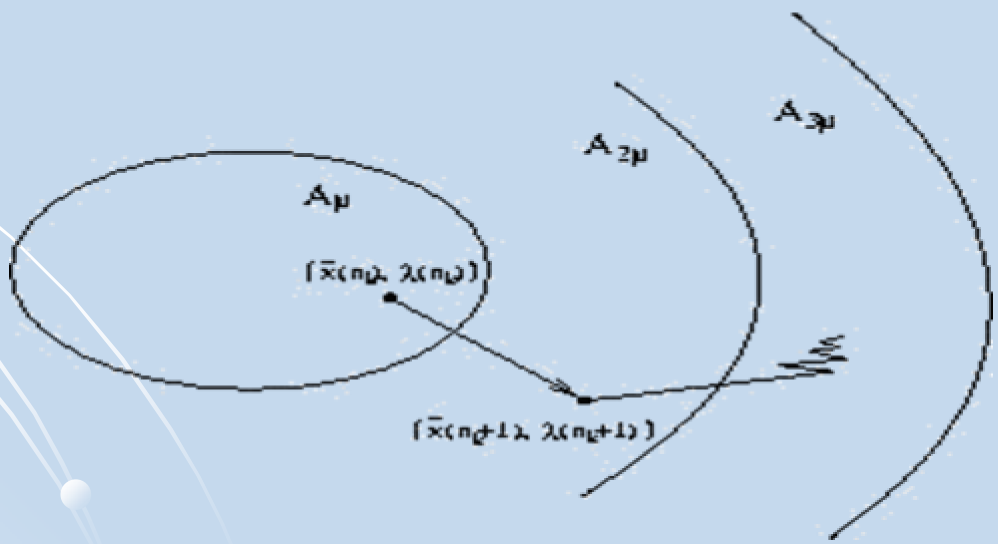
$$E_n[V(X_{n+1})] - V(X_n) \le -\epsilon_n \sigma + Y_n, \text{ for } X_n \notin A,$$

where ϵ_n *satisfies* **A1** *and* σ *is a positive constant. Then the set* A *is recurrent for* $\{X_n\}$*, i.e.,* $X_n \in A$ *for infinitely many* n *with probability one.*

Step II – Local Analysis

Step II: We establish, via "local analysis," that the recurrent iterates eventually reside in an arbitrary small neighborhood of the optimal points, by showing $\{(\tilde{\mathbf{x}}(n), \lambda(n), n = 1, 2, \ldots\}$ leaves $A_{3\mu}$ only finitely often with probability one.

Let $\{n_k, k = 1, 2 \ldots\}$ denote the recurrent times such that $(\tilde{\mathbf{x}}(n_k), \lambda(n_k)) \in A_\mu$. it suffices to show that there exists n_{k_0}, such that for all $n \geq n_{k_0}$, the original iterates $\{(\tilde{\mathbf{x}}(n), \lambda(n)), n = 1, 2, \ldots\}$ reside in $A_{3\mu}$ w.p.1.



Proof for Main Result 2

1) Define $A \triangleq A_\mu \cup A_\eta$. It can be shown that A is compact.

$$A_\eta \triangleq \{(\mathbf{x}, \lambda) : \alpha_s \geq \eta |L_{\tilde{x}_s}(\tilde{\mathbf{x}}, \mathbf{p}, \lambda)|, \text{ for some } s, \\ \text{or } \beta_{(i,j)} \geq \eta |L_{\lambda_{(i,j)}}(\tilde{\mathbf{x}}, \mathbf{p}, \lambda)|, \text{ for some } (i,j), 0 \leq \eta < 1\}.$$

2) Similar to the proof of Main Result 1, it can be shown that

$$E_n[V(\tilde{\mathbf{x}}(n+1), \lambda(n+1))] \leq V(\tilde{\mathbf{x}}(n), \lambda(n)) + 2\epsilon_n(1-\eta)G(\tilde{\mathbf{x}}(n), \lambda(n)) + O(\epsilon_n^2).$$

3) Since $\eta < 1$, using the fact that $G(\tilde{\mathbf{x}}(n), \lambda(n)) < -\delta$ for some positive constant δ when $(\tilde{\mathbf{x}}(n), \lambda(n)) \notin A$, it follows that the iterates return to A infinitely often with probability one by appealing to the stability lemma (where $Y_n = 0, \forall n$).

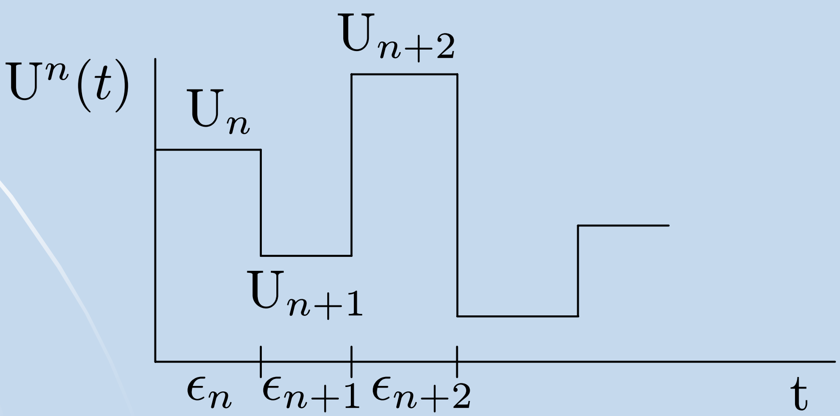
Now let $\mu \to 0$, we have that $A \to A_\eta$.

Rate of Convergence

- The rate of convergence is concerned with the asymptotic behavior of normalized distance of iterates from the optimal points.

- As is standard, assume that the iterates generated by the stochastic primal-dual algorithm have entered in a small neighborhood of an optimal solution $(\tilde{\mathbf{x}}^*, \lambda^*)$.
- Define $U_{\tilde{\mathbf{x}}}(n) \triangleq (\tilde{\mathbf{x}}(n) - \tilde{\mathbf{x}}^*)/\sqrt{\epsilon_n}$ and $U_\lambda(n) \triangleq (\lambda(n) - \lambda^*)/\sqrt{\epsilon_n}$.

- Construct $U^n(t)$ to be the piecewise constant interpolation of $U(n) = \{U_{\tilde{\mathbf{x}}}(n), U_\lambda(n)\}$, i.e., $U^n(t) = U_{n+i}$, for $t \in [t_{n+i} - t_n, t_{n+i+1} - t_n)$, where $t_n \triangleq \sum_{i=0}^{n-1} \epsilon_n$.



Assumptions

A5. *Let* $\theta(n) \triangleq (\tilde{\mathbf{x}}(n), \lambda(n))$ *and* $\phi_n \triangleq (\zeta(n), \xi(n))$. *Suppose for any given small* $\rho > 0$, *there exists a positive definite symmetric matrix* $\Sigma = \sigma\sigma'$ *such that*

$$E_n[\phi_n \phi_n^T - \Sigma] I \{|\theta(n) - \theta^*| \leq \rho\} \to 0$$

as $n \to \infty$.

Define

$$A \triangleq \begin{bmatrix} L_{\tilde{x}\tilde{x}}(\tilde{\mathbf{x}}^*, \mathbf{p}^*, \lambda^*) & L_{\lambda\tilde{x}}(\tilde{\mathbf{x}}^*, \mathbf{p}^*, \lambda^*) \\ -L_{\lambda\tilde{x}}(\tilde{\mathbf{x}}^*, \mathbf{p}^*, \lambda^*) & 0 \end{bmatrix}. \tag{1}$$

A6. *Let* $\epsilon_n = 1/n$; *and assume* $A + I/2$ *is a Hurwitz matrix.*

Note that it can be shown that the real parts of the eigenvalues of A are all non-positive [Bertsekas, 99].

Main Result on Rate of Convergence

a) Under Conditions **A1** and **A3** $-$ **A6**, $U^n(\cdot)$ converges weakly to the solution (denoted as U) to the Skorohod problem

$$\begin{pmatrix} dU_{\tilde{x}} \\ dU_\lambda \end{pmatrix} = \left(A + \frac{I}{2}\right)\begin{pmatrix} U_{\tilde{x}} \\ U_\lambda \end{pmatrix} dt + \sigma dw(t) + \begin{pmatrix} dz_{\tilde{x}} \\ dz_\lambda \end{pmatrix},$$

b) If $(\tilde{\mathbf{x}}^*, \lambda^*)$ is an interior point in the constraint set, then the limiting process U is a stationary Gaussian diffusion process, and $U(n)$ converges in distribution to a normally distributed random variable with mean zero and covariance Σ.

c) If $(\tilde{\mathbf{x}}^*, \lambda^*)$ is on the boundary of the constraint set, then the limiting process U is a stationary reflected linear diffusion process.

Remarks and Engineering Insights

1) In general, the limit process is a stationary reflected linear diffusion process, not necessarily the standard Gaussian diffusion process.

2) The limit process would be Gaussian if there is no reflection term. For instance, when all the link constraints in Problem **P** are active at the optimal point.

3) The rate of convergence depends heavily on the smallest eigenvalue of $\left(A + \frac{I}{2}\right)$. The more negative the smallest eigenvalue is, the faster the rate of convergence would be.

4) Intuitively speaking, the reflection terms would help increase the speed of convergence, which unfortunately cannot be characterized exactly.

5) The covariance matrix of the limit process gives a measure of the spread at the equilibrium point, and is typically "smaller" than the unconstrained case.

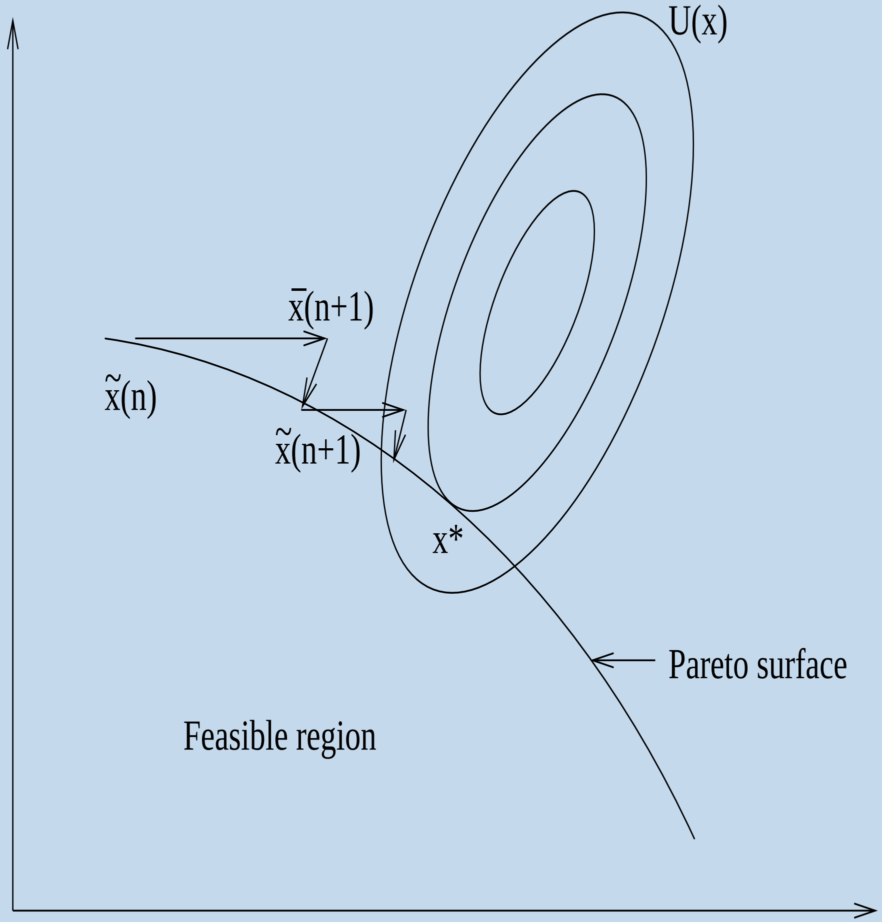
Two Time-Scale Algorithms Using Alternative Decomposition

- In above, we focus on $\kappa \geq 1$. What about $\kappa > 0.$?
- Different decomposition methods lead to different layering algorithms [Palomar-Chiang 06].
- We show that using the feasible direction method and the primal decomposition method, the NUM problem can be solved via two time-scale algorithms
  - Feasible direction method is applicable to non-convex constraint cases.

Feasible Direction Method: $\kappa > 0$.

- On a larger time scale, the source rates are updated using gradient method to maximize the utility functions;

- On a smaller time scale, the updated source rates are projected back to the feasible region imposed by the constraints.



A key advantage: κ does to have to be less than 1. Feasible direction method can be used for non-convex optimization.

Feasible Direction Algorithm

Large (slow) time scale:

$$\bar{x}_s(n+1) = \tilde{x}_s(n) + b_n \frac{dU'_s(\tilde{x}_s)}{d\tilde{x}_s},$$

Small (fast) time scale $(b_n = o(a_n))$:

- *The source rates are updated by*

$$\tilde{x}_s(n+1) = \left[\tilde{x}_s(n) + a_n \left(\bar{x}_s(n+1) - \tilde{x}_s(n) - \exp(\tilde{x}_s(n)) \sum_{(i,j)\in\mathcal{L}(s)} \frac{\lambda_{(i,j)(n)}}{\sum_{k\in\mathcal{S}((i,j))} \exp(\tilde{x}_k(n))}\right)\right]_{-\infty}^{\tilde{M}_s}.$$

- *The shadow prices are updated by*

$$\lambda_{(i,j)}(n+1) = \left[\lambda_{(i,j)}(n) - a_n \left(\log(p_{(i,j)}(n)) + \sum_{k\in N_{to}(j)} \log(1 - P_k(n)) - \log\left(\sum_{s\in\mathcal{S}((i,j))} \exp(\tilde{x}_s(n))\right)\right)\right]_0^{\infty}.$$

- *The persistence probabilities are updated by*

$$p_{(i,j)}(n+1) = \frac{\lambda_{(i,j)}(n)}{\sum_{k\in N_{out}(i)} \lambda_{(i,k)}(n) + \sum_{(l,m):m\in N^I_{from}(i), l\in N_{in}(m)} \lambda_{(l,m)}(n)}.$$

Stochastic Feasible Direction Algorithm

Large (slow) time scale:

$$\bar{x}_s(n+1) = \tilde{x}_s(n) + b_n \frac{dU_s'(\tilde{x}_s)}{d\tilde{x}_s},$$

Small (fast) time scale $(b_n = 0(a_n))$:

- *The source rates are updated by*

$$\tilde{x}_s(n+1) = \left[\tilde{x}_s(n) + a_n \left(\bar{x}_s(n+1) - \tilde{x}_s(n) - \exp(\tilde{x}_s(n)) \sum_{(i,j)\in\mathcal{L}(s)} \frac{\lambda_{(i,j)(n)}}{\sum_{k\in\mathcal{S}((i,j))} \exp(\tilde{x}_k(n))} + M_s(n)\right)\right]_{-\infty}^{\tilde{M}_s}$$

- *The shadow prices are updated by*

$$\lambda_{(i,j)}(n+1) = \left[\lambda_{(i,j)}(n) - a_n \left(\log(p_{(i,j)}(n)) + \sum_{k\in N_{to}(j)} \log(1 - P_k(n)) - \log\left(\sum_{s\in\mathcal{S}((i,j))} \exp(\tilde{x}_s(n))\right) + N_{(i,j)}(n)\right)\right]_0^{\infty}.$$

- *The updating rule for the persistence probability remains the same*

Assumptions

B1. *Condition on the step sizes:*

$$\begin{aligned}&a_n > 0,\ \ b_n > 0,\\ &\textstyle\sum_n a_n = \infty, \sum_n b_n = \infty,\\ &\textstyle\sum_n a_n^2 < \infty,\ \ \sum_n b_n^2 < \infty,\\ &b_n = o(a_n).\end{aligned}$$

B2. *Condition on the estimation error:*

$$\sum_n a_n M(n) < \infty \ \ a.s. \ \ and \ \ \sum_n b_n N(n) < \infty \ \ a.s.$$

Main Result 3: Stability for Two Time-Scale Algorithms

Theorem 3: Under Conditions A1, **B1** and **B2**, the iterates $\{\tilde{\mathbf{x}}(n), n = 1, 2, \ldots\}$, generated by the stochastic feasible direction algorithm, converge with probability one to the optimal solutions of Problem **P**.

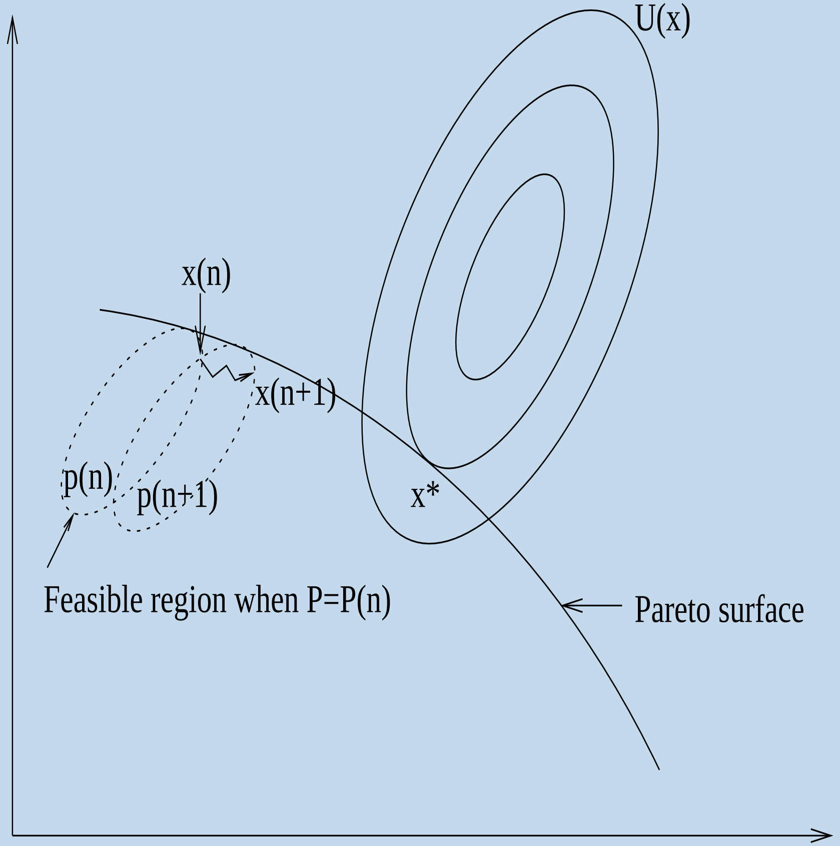
Remarks:

1. Condition **B1** is a standard assumption in two time-scale stochastic approximation algorithms; and intutively speaking, $b_n = o(a_n)$ requires that the updating of the candidate source rates execute at a larger time scale than the projection.

2. Condition **B2** assumes that the average (combined) effects of the biased error and the martingale noise at the two time scales are asymptotically "negligible", i.e., the estimation errors are *asymptotically unbiased*. Again, when the estimators are biased, we cannot hope that the iterates converge to the optimal points with probability one.

Primal Decomposition Method

1. On the smaller time scale, we first fix p, and solve the end-to-end flow control problem with a fixed capacity that is a function of p;
2. On the larger time scale, p is updated using a sub-gradient method.



Primal Decomposition Algorithm

- **Small (fast) time scale:**
  - *The source rates are updated by*

$$x_s(n+1) =_{0 \le x_s \le M_s} \left(U_s(x_s) - x_s \sum_{(i,j) \in \mathcal{L}(s)} \lambda_{(i,j)}(n) \right).$$

  - *The shadow prices are updated by*

$$\lambda_{(i,j)}(n+1) = \left[\lambda_{(i,j)}(n) - a_n \left(c_{(i,j)(n)} - \sum_{s \in \mathcal{S}((i,j))} x_s(n) \right) \right]_0^{\infty}.$$

- **Large (slow) time scale:**
  - *The persistence probabilities are updated by*

$$p_{(i,j)}(n+1) = \left[p_{(i,j)}(n) + b_n \sum_{(s,t) \in L} \lambda_{(s,t)}(n) \frac{\partial \left(p_{(s,t)} \prod_{k \in N_{to}(t)} (1 - P_k) \right)}{\partial p_{(i,j)}} \right]_0^1.$$

Conclusions

1) We have studied joint flow control and MAC design in multi-hop wireless networks with random access; and formulate rate control as a network utility maximization problem.

2) We have shown that the proposed primal-dual algorithm converges almost surely to the optimal solutions provided that the gradient estimator is asymptotically unbiased. For the biased case, we show that the iterates return to a contraction region infinite often w.p.1 provided that the biased errors are uniformly bounded.

3) Our findings on rate of convergence reveal that in general the limit process of the interpolated process, corresponding to the normalized iterate sequence generated from the primal-dual algorithm, is a reflected linear diffusion process, not necessarily the Gaussian diffusion process.

4) Using alternative decomposition methods, we show that the NUM problem can be solved via two time-scale algorithms.