Convex Optimization II

Lecture 15: Optimization of Infinite Horizon Time Averages in Stochastic Systems

Hamed Shah-Mansouri

Department of Electrical Engineering

Sharif University of Technology

1400-2

MOTIVATIONS

- In the previous lecture, we studied the optimization problems with *finite horizon* time averages.
- How do we solve the decision problems involving optimizing time averages in infinite horizon?
- Time average constraints can be modeled with stochastic queuing systems with either actual or virtual queues.

[1] M. Neely, Stochastic Network Optimization with Application to Communication and Queueing Systems. Morgan & Claypool Publishers 2010.

OUTLINE

- Stochastic Optimization Problems in Infinite Horizon
- Queue & Rate Stability
- Scheduling for Rate Stability
- Lyapunov Drift and Lyapunov Optimization

Stochastic Optimization Problems in Infinite Horizon

- Consider a stochastic system that operates in discrete time with unit time slots $t \in \{0, 1, 2, \ldots\}$.

- Every slot, a control action $\alpha(t)$ is taken in order to make optimal or sub-optimal decisions for the stochastic problem.

- Each action affects some attributes (e.g., power, delay, profit, ...)

$$\boldsymbol{y}(t) = (y_0(t), y_1(t), \ldots, y_L(t))$$
$$\boldsymbol{e}(t) = (e_1(t), \ldots, e_J(t))$$

- These attributes are given by general functions

$$y_l(t) = \hat{y}_l(\alpha(t), \omega(t)), \text{ for all } l \in \{0, 1, \ldots, L\}$$
$$e_j(t) = \hat{e}_j(\alpha(t), \omega(t)), \text{ for all } j \in \{1, \ldots, J\}$$

where $\omega(t)$ is a random event observed on slot t.

Stochastic Optimization Problems in Infinite Horizon

- The time averages under a particular control algorithm

$$\bar{y}_l \triangleq \lim_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} y_l(\tau), \qquad \bar{e}_j \triangleq \lim_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} e_j(\tau)$$

- Objective: Design a control algorithm that solves

$$\begin{array}{lll} \text{minimize} & \bar{y}_0 & \\ \text{subject to} & \bar{y}_l \leq 0, & \text{for all } l \in \{1, \dots, L\} \\ & \bar{e}_j = 0, & \text{for all } j \in \{1, \dots, J\} \\ & \alpha(t) \in \mathcal{A}_{\omega(t)}, & \text{for all slots } t \\ & \textit{Stability of All Queues}. & \end{array}$$

- Solution: An algorithm for choosing control actions over time in reaction to the existing state, such that all of the constraints are satisfied and the quantity to be minimized is as small as possible.

Queuing Approach

Queueing theory plays a central role in this type of stochastic optimization even if there are no underlying queues in the original problem.

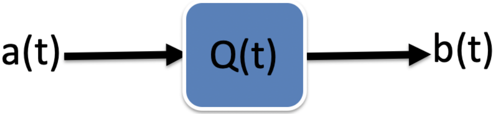
We can introduce virtual queues as a strong method for ensuring that the required time average constraints are satisfied.

The system is described by a collection of queue backlogs, written in vector form

$$\boldsymbol{Q}(t) = (Q_1(t), ..., Q_K(t)),$$

where $Q_k(t)$ represents the contents of a single-server discrete time queuing system defined over integer time slots.

QUEUE DYNAMICS



Backlog : $Q(t)$

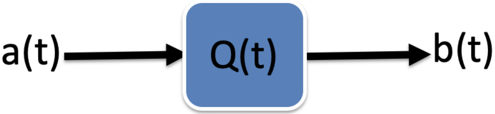
Arrival : $a(t)$

Departure : $\min\{Q(t), b(t)\}$

Queue Dynamic:

$$Q(t+1) = \max\{Q(t) + a(t) - b(t), 0\}, \quad \forall \; t \in \{0, 1, 2, \ldots\}$$

RATE STABILITY



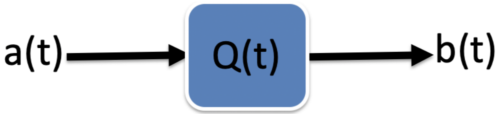
Definition: Q(t) is *rate stable* if:

$$\lim_{t\to\infty} \frac{Q(t)}{t} = 0 \quad \text{with probability 1}$$

Definition: Q(t) is *mean rate stable* if:

$$\lim_{t\to\infty} \frac{\mathbb{E}[|Q(t)|]}{t} = 0 \quad \text{with probability 1}$$

Rate Stability Theorem



Suppose

$$\lim_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} a(t) = a^{av}, \quad \text{with probability 1}$$

$$\lim_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} b(t) = b^{av}, \quad \text{with probability 1}$$

Then, $Q(t)$ is *rate stable* if and only if

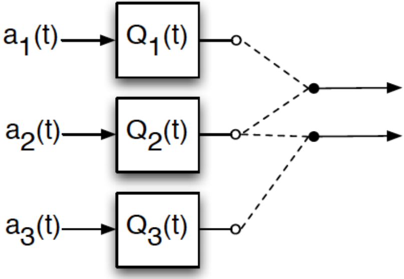
$$a^{av} \leq b^{av}$$

Equivalently,

$$\lim_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} a(t) - b(t) \leq 0, \quad \text{with probability 1}$$

Scheduling for Rate Stability

A 3-queue, 2-server system:



- All jobs have the same size.
- At most a server is allocated to each queue on a given slot and can serve exactly one job on that slot.
- The service is given for $i \in \{1, 2, 3\}$ by

$$b_i(t) = \begin{cases} 1 & \text{if a server is connected to queue } i \text{ on slot } t \\ 0 & \text{otherwise.} \end{cases}$$

Scheduling for Rate Stability

Objective: Design a server allocation algorithm to make all queues rate stable.

Solution: Choose service allocations every slot to ensure that each $b_i(t)$ has time average b_i^{av}

$$b_i^{av} \geq a_i^{av}$$

Then, by the rate stability theorem, all queues are rate stable.

Necessary conditions for the existence of a rate stabilizing algorithm

$$0 \leq a_i^{av} \leq 1, \text{for all } i \in \{1, 2, 3\}$$
$$a_1^{av} + a_2^{av} + a_3^{av} \leq 2$$

Scheduling for Rate Stability

Example 1: $(a_1^{av}, a_2^{av}, a_3^{av}) = (0.5, 0.5, 0.9)$

Choose $(b_1(t), b_2(t), b_3(t)) \subset \{(1,1,0),(1,0,1),(0,1,1)\}$

Two Approaches:

Randomized Scheduling Algorithm
Choose $(0,1,1)$ with probability $1/2$ and $(1,0,1)$ with probability $1/2$ i.i.d. on every slot.

Deterministic Scheduling Algorithm
Alternate between $(0,1,1)$ (on odd slots) and $(1,0,1)$ (on even slots)

Scheduling for Rate Stability

Example 2: $(a_1^{av}, a_2^{av}, a_3^{av}) = (0.7, 0.9, 0.4)$

Choose $(b_1(t), b_2(t), b_3(t)) \subset \{(1,1,0),(1,0,1),(0,1,1)\}$

Linear Programming Approach

Independently choose the service vector $(0,1,1)$ w.p. p_1, $(1,0,1)$ w.p. p_2, and $(1,1,0)$ w.p. p_3.

$$
\begin{array}{ll}
\text{find} & \mathbf{p} \\
\text{subject to} & p_1(0,1,1) + p_2(1,0,1) + p_3(1,1,0) \geq (0.7, 0.9, 0.4) \\
& p_1 + p_2 + p_3 = 1 \\
& p_i \geq 0, \;\; i = \{1,2,3\}
\end{array}
$$

Scheduling for Rate Stability

How would you design rate stable algorithms if you do not have a-priori knowledge of the arrival rates?

We now pursue queue stability via an algorithm that makes decisions based on both the current states of stochastic parameters and the current queue backlogs.

Optimizing Time Averages

Virtual Queues

$$
\begin{array}{lll}
\text{minimize} & \bar{y}_0 & \\
\text{subject to} & \bar{y}_l \leq 0, & \text{for all } l \in \{1, \ldots, L\} \\
& \bar{e}_j = 0, & \text{for all } j \in \{1, \ldots, J\} \\
& \alpha(t) \in \mathcal{A}_{\omega(t)}, & \text{for all slots } t \\
& \textit{Stability of All Queues}. &
\end{array}
$$

We make a virtual queue for each constraint. The queues evolve according to

I $$Z_l(t+1) = \max[Z_l(t) + y_l(t), 0] \text{ for } l \in \{1, \ldots, L\}$$

II $$H_j(t+1) = H_j(t) + e_j(t) \text{ for } j \in \{1, \ldots, J\}$$

The virtual queues are introduced to enforce the constraints.

Sample Path Property

For any discrete time queuing system evolving by

$$Q(t+1) = \max\{Q(t) + a(t) - b(t), 0\}, \quad \forall\ t \in \{0, 1, 2, \ldots\}$$

and for any two slots t_1 and t_2 such that $0 \leq t_1 < t_2$, we have

$$Q(t_2) - Q(t_1) \geq \sum_{\tau=t_1}^{t2-1} a(\tau) - \sum_{\tau=t_1}^{t2-1} b(\tau)$$

Therefore, for any $t > 0$, we have the *sample path property* as follows.

$$\frac{Q(t)}{t} - \frac{Q(0)}{t} \geq \frac{1}{t}\sum_{\tau=0}^{t-1} a(\tau) - \frac{1}{t}\sum_{\tau=0}^{t-1} b(\tau)$$

Virtual Queues

By the sample path property, we have

$$\frac{Z_l(t)}{t} - \frac{Z_l(0)}{t} \geq \frac{1}{t}\sum_{\tau=0}^{t-1} y_l(\tau)$$

Taking expectations of the above and taking $t \to \infty$ shows

$$\lim_{t\to\infty} \sup \frac{\mathbb{E}Z_l(t)}{t} \geq \lim_{t\to\infty} \sup \bar{y}_l(t),$$

where $\bar{y}_l(t)$ is the time average of $y_l(\tau)$ over $\tau \in \{0, 1, \ldots, t\}$.

Thus, if $Z_l(t)$ is mean rate stable, then

$$\lim_{t\to\infty} \sup \bar{y}_l(t) \leq 0$$

Virtual Queues

The virtual queue $H_j(t)$ is designed to turn the time average equality constraint $\bar{e}_j = 0$ into a pure queue stability problem

$$H_j(t) - H_j(0) = \sum_{\tau=0}^{t-1} e_j(\tau)$$

Thus,

$$\frac{\mathbb{E}H_j(t) - \mathbb{E}H_j(0)}{t} = \bar{e}_j(t)$$

Therefore, if $H_j(t)$ is mean rate stable then,

$$\underbrace{\lim_{t\to\infty} \bar{e}_j(t)}_{\bar{e}_j} = 0.$$

Optimizing Time Averages

Virtual queues turn the problem of satisfying time average inequality & equality constraints into a pure queue stability problem.

$$\begin{array}{ll} \text{minimize} & \bar{y}_0 \\ \text{subject to} & \alpha(t) \in \mathcal{A}_{\omega(t)}, \quad \text{for all slots } t \\ & \textit{Stability of All Queues } \boldsymbol{\Theta} = (\boldsymbol{Q}, \boldsymbol{Z}, \boldsymbol{H}). \end{array}$$

Backlog Vector

$$\boldsymbol{\Theta}(t) = (\boldsymbol{Q}(t), \boldsymbol{Z}(t), \boldsymbol{H}(t)) = (\Theta_1(t), \ldots, \Theta_N(t))$$

Lyapunov Drift

Quadratic Lyapunov function, with positive weights w_n (mostly 1)

$$L(\boldsymbol{\Theta}(t)) = \frac{1}{2}\sum_{n=1}^{N} w_n \Theta_n(t)^2$$

- $L(\boldsymbol{\Theta}(t)) \geq 0$ for all backlog vectors, with equality if and only if the system is empty on slot t.
- $L(\boldsymbol{\Theta}(t))$ being small implies that all queue backlogs are small.
- $L(\boldsymbol{\Theta}(t))$ being large implies that at least one queue backlog is large.

If there is a finite constant M such that $L(\boldsymbol{\Theta}(t)) \leq M$ for all t, then clearly all queue backlogs are always bounded.

We aim to design an algorithm that consistently pushes the queue backlog towards $L(\boldsymbol{\Theta}(t)) \leq M$ (a low congestion region). This helps control the congestion and stabilize the queues.

Lyapunov Drift

A bound on the change in the Lyapunov function from one slot to the next

$$
\begin{aligned}
L(\boldsymbol{\Theta}(t+1)) - L(\boldsymbol{\Theta}(t)) &= \frac{1}{2}\sum_{n=1}^{N}\left[\Theta_n(t+1)^2 - \Theta_n(t)^2\right] \\
&= \frac{1}{2}\sum_{i=1}^{N}\left[\left(\max\{\Theta_n(t) - b_n(t) + a_n(t), 0\}\right)^2 - \Theta_n(t)^2\right] \\
&\leq \sum_{i=1}^{N}\frac{a_n(t)^2 + b_n(t)^2}{2} + \sum_{n=1}^{N}\Theta_n(t)\left(a_n(t) - b_n(t)\right)
\end{aligned}
$$

Conditional Lyapunov drift for slot t

$$
\Delta(\boldsymbol{\Theta}(t)) = \mathbb{E}\left\{L(\boldsymbol{\Theta}(t+1) - L(\boldsymbol{\Theta}(t)|\boldsymbol{\Theta}(t)\right\}
$$

Lyapunov Drift

$$\Delta(\boldsymbol{\Theta}(t)) \leq$$

$$\underbrace{\mathbb{E}\left\{\sum_{n=1}^{N}\frac{a_n(t)^2+b_n(t)^2}{2}|\boldsymbol{\Theta}(t)\right\}}_{\leq B}+\sum_{n=1}^{N}\Theta_n(t)a_n^{av}-\mathbb{E}\left\{\sum_{n=1}^{N}\Theta_n(t)b_n(t)|\boldsymbol{\Theta}(t)\right\}$$

Define B as a finite constant, thus,

$$\Delta(\boldsymbol{\Theta}(t)) \leq B+\sum_{n=1}^{N}\Theta_n(t)a_n^{av}-\mathbb{E}\left\{\sum_{n=1}^{N}\Theta_n(t)b_n(t)|\boldsymbol{\Theta}(t)\right\} \qquad (1)$$

Recall that

$$b_n(t)=\hat{b}_n(\alpha(t),\omega(t)),\ \text{for all } n\in\{1,\dots,N\}$$

The Min-Drift or Max-Weight Algorithm:

For the rate stability, we now take an action that minimizes the right hand side of (1). Technically, we seek to maximize a weighted sum of b_n, where the weights are queue backlogs.

Lyapunov Drift for Stable Scheduling

Lyapunov Drift Theorem:

Consider the quadratic Lyapunov function, and assume $\mathbb{E}\{L(\boldsymbol{\Theta}(0))\} < \infty$. Suppose there are constants $B > 0$, $\epsilon \geq 0$ such that the following drift condition holds for all slots $\tau \in \{0, 1, 2, \ldots\}$ and all possible $\boldsymbol{\Theta}(t)$:

$$\Delta(\boldsymbol{\Theta}(t)) \leq B - \epsilon \sum_{n=1}^{N} |\Theta_n(t)|$$

Then, all queues $\Theta_n(t)$ are mean rate stable.

No knowledge of ω's distribution is required.

Drift-Plus-Penalty

Instead of taking a control action to minimize a bound on $\Delta(\mathbf{\Theta}(t))$, we minimize a bound on the following drift-plus-penalty expression

$$\Delta(\mathbf{\Theta}(t)) + V\mathbb{E}\{y_0(t)|\mathbf{\Theta}(t))\} \tag{2}$$

V is a parameter to show how much we emphasize the objective.

We want to make $\mathbf{\Theta}(t)$ small to push queue backlog towards a lower congestion state, but we also want to make $\mathbb{E}\{y_0(t)|\mathbf{\Theta}(t))\}$ small so that we do not incur a large penalty.

This intuitive algorithm leads to a provable cost-backlog tradeoff.

Lyapunov Optimization Theorem

Consider the quadratic Lyapunov function, and assume $\mathbb{E}\{L(\mathbf{\Theta}(0))\} < \infty$ and $\mathbb{E}\{y_0(t)\} \geq y_{\min}$. Suppose there are constants $B > 0$, $\epsilon \geq 0$, $V > 0$, and y^* such that for all slots $\tau \in \{0, 1, 2, \ldots\}$ and all possible $\mathbf{\Theta}(t)$):

$$\Delta(\mathbf{\Theta}(t)) + V\mathbb{E}\{y_0(t)|\mathbf{\Theta}(t))\} \leq \overset{1}{B} + \overset{2}{Vy^*} - \overset{3}{\epsilon \sum_{n=1}^{N}} |\Theta_n(t)|$$

Then, all queues are mean rate stable.
Further, if $V > 0$ and $\epsilon > 0$, then time average expected penalty and queue backlog satisfy

$$\bar{y}_0 = \lim_{t\to\infty} \sup \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}\{y_0(t)\} \leq y^* + \frac{B}{V}$$

$$\lim_{t\to\infty} \sup \frac{1}{t} \sum_{\tau=0}^{t-1} \sum_{n=1}^{N} \mathbb{E}\{|\Theta_n(t)|)\} \leq \frac{B + V(y^* - y_{\min})}{\epsilon}$$

Drift-plus-Penalty Algorithm

The drift-plus-penalty algorithm observers queue backlogs and the realization of random parameter every slot t and then chooses an action $\alpha(t)$ to minimize the right hand-side of the above inequality.

Performance Tradeoff Theorem

For any $V > 0$, if the problem is feasible, then the drift-plus-penalty algorithm stabilizes all virtual queues. Further,

$$\bar{y}_0 \leq y^* + \frac{B}{V}$$

پایداری صفها

$$\bar{y}_l \leq 0, \qquad \text{for all } l \in \{1, \ldots, L\}$$

$$\bar{e}_j = 0, \qquad \text{for all } j \in \{1, \ldots, J\}$$

Stability of All Actaull Queues.

SUMMARY

- How to deal with infinite horizon time averages in stochastic problems?
- Stochastic programming problems can be mapped into queuing systems so that the queues' stability guarantees the inequality & equality constraints are met.
- Lyapunov drift provides queue stability for all actual & virtual queues.
- Lyapunov optimization balances the tradeoff between queue back logs and penalty (i.e., cost).