Lectures 4 & 5

Telephone network: resource provisioning and performance evaluation

Model: static

Given a network with fixed topology and routing: there are \( L \) links: \( L = \{1, \ldots, L\} \), link \( l \in L \) has capacity to accommodate \( c \geq 0 \), phone-line connections simultaneously. Plausible routes are \( R = \{1, \ldots, R\} \), each described by a set of connected links.

Let \( A = [A_{lr}] \) be routing adjacency matrix of dimension \( L \times R \). Here,

\[
A_{lr} = \begin{cases} 
1 & \text{if route } r \text{ passes through link } l \\
0 & \text{otherwise}
\end{cases}
\]

Finally a (phone) call along route \( r \) utilizes \( A_{lr} \) amount of resource on link \( l \).
Model: dynamics

Arrival: New calls on route \( r \) arrive as per independent Poisson process of rate \( \lambda r \).

Each arriving call is accepted if there is availability of unit amount of unused capacity on each link \( l \) on router \( r \). Else, the call is dropped.

Departure: A dropped call departs immediately. An accepted last for time distributed as per exponential distribution of mean 1; during this time it holds the unit capacity on all links of its router. After this, it departs.
Network Markov Chain:

Let $X_r(t)$ be number of active calls on router $r$ at time $t$. By definition, capacity conservation must happen; i.e.

$$\sum_{r=1}^{R} X_r(t) A_{lr} \leq C_l \text{ for all } l \in \mathbb{L}.$$  

Equivalently, $X(t) = \lceil X_r(t) \rceil$ is such that

$$X(t) \in X = \left\{ x \in \mathbb{N}^R : A x \leq C \right\}$$

with $C = \lceil C_l \rceil$.

Fact: under the above described dynamics, $X(t)$ is Markov.
State transitions:

“Common Poisson Process” construction:

Consider arrival processes of all $R$ routes. If we merge them we obtain a Poisson process with rate $\lambda = \sum_{r=1}^{R} \lambda_r$. Therefore, one way to construct individual arrival process is to first generate an arrival as per a Poisson process of rate $\lambda$, and each such arrival is assigned to route $r$ with probability $\frac{\lambda_r}{\lambda}$ for $r=1, \ldots, R$. 
For departures, consider state of the network, \( x(t) = [x_r(t)] \) at time \( t \). Now a departure will happen when service of one of the active call is completed. For this, first consider the following: with each active call associate a Poisson process of unit rate. When the ‘next’ event happens for the associated Poisson process, the corresponding active call departs. Therefore, by the same argument as above, we can imagine a “common” departure Poisson process of net. rate \( \sum_{r=1}^{N} x_r(t) \) at time \( t \). When an event happens as per this common process, a call from route \( r \) departs with
probability $\frac{X_r(t)}{\left( \sum_{s=1}^{R} X_s(t) \right)}$. 

Therefore, by combining the net arrival and net departure process we obtain the "system" Poisson process of net rate $(\lambda + \sum_{r=1}^{R} X_r(t))$ at time $t$.

When an event happens as per this combined Poisson process, it results in

departure of a route $r$ call with probability

$\frac{X_r(t)}{\sum_{s=1}^{R} \lambda_s + \sum_{s=1}^{R} X_s(t)}$ \hspace{1cm} (1)$

arrival of a route $r$ call with probability

$\frac{\lambda_r}{\sum_{s=1}^{R} \lambda_s + \sum_{s=1}^{R} X_s(t)}$ \hspace{1cm} (2)$
Equivalent discrete-time Markov chain:

At time $t$, state is $X(t) = [x_r(t)]$.

In time $[t, t+\delta]$, for very very small $\delta > 0$, for the common "1 system" Poisson process, an event will happen with probability

\[
\left( \sum_{s=1}^{R} x_s + \sum_{s=1}^{R} x_s(t) \right) \cdot \delta + o(\delta),
\]

more than one event happens with probability $o(\delta)$ and no event happens with probability $1 - \left( \sum_{s=1}^{R} x_s + \sum_{s=1}^{R} x_s(t) \delta + o(\delta) \right)$.

As usual, ignoring $o(\delta)$ term, we have that either an event happens or does not happen.
When an event happens, it is departure or arrival for a route $r$ with probability given as per (1) and (2).

Therefore, effectively in $[t, t+\delta]$, for a very very small $\delta > 0$,

$$X(t+\delta) = X(t) \text{ w.p. } 1 - \left( \frac{K}{\sum_{s=1}^{K} X_s(t)} \right) \delta$$

and otherwise one of the $X_r(t)$ changes as per

$$X_r(t+\delta) = \begin{cases} 
X_r(t) + 1 \text{ w.p. } \lambda_r \delta \\
X_r(t) - 1 \text{ w.p. } \lambda_r(t) \delta.
\end{cases}$$
Stationary distribution:

It is easy to see that thus resolting Markov chain is finite state with state space \( \mathcal{X} \), is irreducible and aperiodic. Therefore, it has a unique stationary distribution. It can be checked that the stationary distribution is

\[
\Pi = \left[ \Pi_x \right]_{x \in \mathcal{X}} \text{ such that } \\
\Pi_x \propto \prod_{r=1}^{R} \left( \frac{x_r}{x_r!} \right).
\]

\[
= \frac{1}{Z} \prod_{r=1}^{R} \frac{x_r^{x_r}}{x_r!},
\]

where

\[
Z = \sum_{x \in \mathcal{X}} \prod_{r=1}^{R} \frac{x_r^{x_r}}{x_r!}.
\]
This is because they satisfy detailed balance equation. That is, for any possible pair of transitions, say $X \rightarrow X + e(r)$, $X + e(r) \rightarrow X$ and $X, X + e(r) \in \mathcal{X}$. 

Here $e(r) \in \mathbb{E}_{0,1}^R$, its $r$th component is 1, others are 0. That is, $X + e(r)$ is obtained when an arrival to route $r$ happens.

\[
\prod_{X_r} P_{x,r,x + e(r)} = \frac{1}{Z} \left( \prod_{r=1}^{R} \frac{\lambda_r x_r}{x_r!} \right) \cdot \lambda_r \cdot \delta \\
= \frac{1}{Z} \left( \prod_{r=1}^{R} \frac{\lambda_r x_r+1}{(x_r+1)!} \right) \cdot \delta
\]
\[ \Pi_{x+e(r)} = P_{x+e(r), x}. \]

Now, from HW1 it follows that \( \Pi \) is indeed the stationary distribution of the Markov chain.

Thus far, we have described a model, dynamics and equilibrium behavior of the telephone network. Now, we are ready to evaluate its performance.
Performance metric

Basic question: what is performance metric?
In case of telephone network, primary metric is the "loss" or "drop" rate.

Let $L_r$ be the fraction of calls arriving on route $r$ that are dropped. We shall use Little's Law to "evaluate" it.

Now a call arriving on route $r$ is either accepted, $1-L_r$ fraction of time, or dropped, $L_r$ fraction of the time.

An accepted call leaves network after unit amount of time on average, while dropped call departs instantly.
Thus, the average delay experienced by route $r$ calls is

$$(1-L_r) \cdot 1 + L_r \cdot 0 = 1-L_r$$

Net arrival rate to route $r$ calls is $\lambda_r$. Therefore, by Little's Law,

$$E[X_r] = \lambda_r \cdot (1-L_r),$$

where

$$E[X_r] = \sum_{x \in X} x_r \cdot \Pi_{x}$$

$$= \sum_{x \in X_i: x_r \geq 1} x_r \cdot \frac{\lambda r}{x_r!} \cdot \left( \prod_{s \neq r} \frac{\lambda s}{x_s!} \right)$$
\[ = \lambda_r \sum_{x \in X : x_r \geq 1} \frac{\lambda_r^{x_r-1}}{(x_r-1)!} \left( \prod_{s \neq r} \frac{\lambda_s^{x_s}}{x_s!} \right) \]

\[ = \lambda_r \sum_{y \in X : y \geq \epsilon(r) \in X} \frac{\lambda_s}{\prod_{s=1}^{R} (\frac{y_s}{y_s!})} \]

Therefore, effectively

\[ 1 - L_r = \sum_{y \in X : y \geq \epsilon(r) \in X} \frac{R}{\prod_{s=1}^{R} (\frac{y_s}{y_s!})} \]
Therefore, we wish to evaluate, in a nutshell, $\mathbb{E}[x_r]$ to obtain $L_r$. Clearly, as per (4), this will require 'summing up' a lot of terms. And this is computationally challenging (or, $\#P$-complete problem).

What we need is a reasonable approximation:

A 'first order' approximation:

$$\mathbb{E}[x_r] = x_r^*$$

where $x^* = (x_r^*)$ is such that it has the maximal probability.
Q1: Is this approximation any good?

Q2: Is this "really" easier to evaluate?

In what follows, we will answer both of these questions (in affirmative).
First question: why $\mathbb{E}[x_r] \approx x_r^*$?

The justification we provide is based on large network consideration. First observe that the computation cost of $\mathbb{E}[x_r]$ scales as $\mathcal{O}(\sqrt{n})$. Now if all link capacities are small, then really there is no need for approximation. Therefore the regime of interest is when link capacities are large. In practice, it makes sense as well since we are dealing with large, complex networks.

Now, what does large mean or more specifically, how to formalize it?
Network Scaling.

Let $N$ be 'scaling' parameter that we use to understand large networks. Consider network with parameter $N$ as:

- the topology: $L, R, A$ as is.
- link capacity $c_e^N = N_c$ for $e \in E$
- arrival rates $\lambda_r^N = N \lambda_r$ for $r \in R$

And interest is in studying behavior of $L_r^N$. Specifically, let

$$L_r^{N, x} = 1 - \frac{1}{\lambda_r^N} \cdot X_r^{N, x}$$
$$= 1 - \frac{1}{\lambda_r} \cdot \left( \frac{X_r^{N, x}}{N} \right)$$
And, actual loss probability

\[ L_r^N = 1 - \frac{1}{X_r^N} \cdot |E[X_r^N]| \]

\[ = 1 - \frac{1}{X_r^N} \cdot \left| E\left[ \frac{1}{N} X_r^N \right] \right| \]

Ideally, we want to show that

\[ |L_r^N - L_r^{N,*}| \to 0 \quad \text{as} \quad N \to \infty. \]

Equivalently,

\[ \left| E\left[ \frac{1}{N} X_r^N \right] - \frac{1}{N} X_r^{N,*} \right| \to 0 \quad \text{as} \quad N \to \infty. \]

For this, we shall first study \( \frac{1}{N} X_r^{N,*} \) as \( N \to \infty \). That will explain the above desired result.
To this end, define for the $N$th system

$$\mathcal{X}^N = \{ x \in \mathbb{N}^R : A x \leq C \}$$

And, its stationary distribution

$$\Pi_N(x) = \frac{1}{Z(N)} \cdot \prod_{r=1}^{R} (N \lambda_r)^{x_r} \frac{x_r!}{Z(N)}$$

Define, $Z = \frac{1}{N} X$.

Then, $Z \in \bar{\mathcal{X}}^N = \{ Z : N \cdot Z \in \mathbb{N}^R, \ A Z \leq C \}

Then, we can think of $\Pi_N$ as defined on $\bar{\mathcal{X}}^N$. 
Specifically,

\[ \prod_N^{(x)} = \prod_N^{(N \cdot 2)} = \frac{1}{Z(N)} \prod_{r=1}^{R} \frac{(N\lambda_r)^{Nz_r}}{Nz_r!} \]

\[ \propto \prod_{r=1}^{R} \exp \left( Nz_r \log N\lambda_r - \log Nz_r! \right) \]

\[ \propto \prod_{r=1}^{R} \exp \left( Nz_r \log N + Nz_r \log \lambda_r - \log Nz_r! \right) \]

By Stirling’s approximation, for \( n \) large enough

\[ \log n! \approx n \log n - n \]

\[ \log Nz_r! = N z_r \log N z_r - N z_r \]

\[ = N z_r \log N + N z_r \log \lambda_r - N z_r \]
Therefore,
$$
\prod_{r=1}^{R} \exp(N z_r \log \lambda_r - N z_r \log z_r + N z_r)
$$

$$
\propto \exp \left( N \left\{ \sum_{r=1}^{R} z_r \log \lambda_r - z_r \log z_r + z_r \right\} \right).
$$

$X^{N,*}_{\lambda}$ is one with maximal probability in $X^N$. That is, for $Z^{N,*}_{\lambda} = \{X^{N,*}_{\lambda} \in \overline{X}^N \}$ has maximal probability. Or, it maximizes
$$
\sum_{r=1}^{R} z_r \log \lambda_r - z_r \log z_r + z_r
$$
$$
= \sum_{r=1}^{R} z_r \log \frac{e^{\lambda_r}}{z_r}.
$$

over $Z \in \overline{X}^N$. 


Note that as $N \to \infty$,

$$\overline{X^N} \to \overline{X} = \{ y \in \mathbb{R}^r, y \geq 0 : Ay \leq c \}.$$

Therefore,

$$Z_{N, *} \to Z^* \text{ where } Z^* \text{ solves}$$

$$\text{maximize } \sum_{r=1}^{R} z_r \log \left( \frac{z_r}{w_r} \right)$$

$$\text{over } Z \in \overline{X}.\$$

\textbf{CLAIM 1:} We shall show that $\text{OPT}$ has a unique solution $Z^*$. Therefore, for any $\hat{Z} \neq z^*$,

$$f(\hat{Z}) < f(z^*)$$
where \( f(z) = \sum_{r=1}^{R} Z_r \log \frac{e^{\lambda_r}}{Z_r} \)

If so, then note that

\[
\prod_{N} \left( \frac{N^Z}{N^{Z^*}} \right) = \exp \left( -N \left( f(Z^*) - f(Z) \right) \right)
\]

\[\to 0 \text{ as } N \to \infty.\]

This suggests that as \( N \to \infty \), the entire distribution \( \prod_N \) “concentrates” around \( NZ^* \). That is, as \( N \to \infty \)

\[
\prod_{N} \left( \{ x : \left| \frac{1}{N} x - Z^* \right| > \varepsilon \} \right) \to 0.
\]
This leads to the fact that

\[ \mathbb{E}\left[ \frac{1}{N} X \right] \rightarrow 2^* \text{ as } N \rightarrow \infty. \]  

\[ (6) \]

From (5) and (6) it follows that

\[ \left| \mathbb{E}\left[ \frac{1}{N} X \right] - \frac{1}{N} X^* \right| \rightarrow 0 \text{ as } N \rightarrow \infty. \]

Next, we justify Claim 1 to complete answer for question one. And in the process we shall learn about "elements of convex optimization" as well as a simple algorithm to solve for $2^*$. 
Convex Minimization or Concave Maximization

Canonical form:

\[
\text{minimize } f(x) \text{ over } x \in \mathbb{R}^n
\]

subject to \( h_i(x) \leq b_i, \ i = 1, \ldots, m \).

In \( \text{OPT} \ 1 \)

\[
f(z) = - \sum_{r=1}^{R} z_r \log \left( \frac{e^{\lambda_r}}{z_r} \right)
\]

\[
h_k(z) = A_k z = \sum_{r=1}^{R} A_{kr} z_r
\]

\[
b_k = c_k
\]

Convex: \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is convex if

\[
f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y)
\]

for \( \theta \in (0,1) \) and any \( x, y \in \mathbb{R}^n \).

It is strictly convex if the inequality is strict, i.e.

\[
f(\theta x + (1-\theta) y) < \theta f(x) + (1-\theta) f(y)
\]
Concave: $f: \mathbb{R}^n \to \mathbb{R}$ is concave if $-f$ is convex.

Checking convexity:

First order condition: $\nabla f(x) = \left[ \frac{\partial f(x)}{\partial x_i} \right]_i$

$f$ is convex if for any $x, y$

$$f(y) \geq f(x) + (y-x)^T \nabla f(x)$$

Strictly convex if for any $x \neq y$,

$$f(y) > f(x) + (y-x)^T \nabla f(x).$$

Second order condition: $\nabla^2 f(x) = \left[ \frac{\partial^2 f(x)}{\partial x_i \partial y_j} \right]_{i,j}$

Convex: $z^T \nabla^2 f(x) z \geq 0$ for any $z \in \mathbb{R}^n, x \in \mathbb{R}^n$

Strictly convex: $z^T \nabla^2 f(x) z > 0$ for any $z \in \mathbb{R}^n, x \in \mathbb{R}^n$. 
Convexity "preservation":

\[ f = f_1 + f_2 \text{ is convex if } f_1, f_2 \text{ are} \]

Our objective:

\[ f(z) = \sum_{r=1}^{R} f_r(z) \]

where \( f_r(z) = -z_r \log \frac{e^{\lambda r}}{z_r} = z_r \log \frac{z_r}{e^{\lambda r}} \)

\[ \nabla f_r(z) = \left[ \frac{\partial f_r(z)}{\partial z_s} \right]_s \]

Now \( \frac{\partial f_r(z)}{\partial z_s} = \begin{cases} 0 & \text{if } s \neq r \\ \frac{2}{z_r} \left( z_r \log \frac{z_r}{e^{\lambda r}} \right) & \text{if } s = r \end{cases} \]

\[ = -\log e^{\lambda r} + 1 + \log z_r \]

\[ = \log \left( \frac{z_r}{e^{\lambda r}} \right) \text{ if } s = r \]
$\nabla^2 f_r(z) = \left[ \frac{\partial^2 f_r(z)}{\partial z_i \partial z_j} \right]_{i,j}$

$$\frac{\partial^2 f_r(z)}{\partial z_i \partial z_j} = \begin{cases} \frac{1}{z_r} & i = j = r \\ 0 & \text{o.w.} \end{cases}$$

Therefore,

$$\nabla^2 f(z) = \sum_r \nabla^2 f_r(z) = \left[ \frac{\partial^2 f(z)}{\partial z_i \partial z_j} \right]_{i,j}$$

where

$$\frac{\partial^2 f(z)}{\partial z_i \partial z_j} = \begin{cases} \frac{1}{z_i} & i = j \\ 0 & \text{o.w.} \end{cases}$$
Therefore, $\nabla^2 f(z)$ is diagonal matrix with strictly positive entries on diagonal if $z > 0$ (componentwise).

Thus, $f$ is strictly convex for $z > 0$ and convex when $z \geq 0$.

Convex minimization: convex objective over linear constraints.

Key property: unique soln that is easy to find, if strictly convex objective.

Primary reasons:

locally optimal = globally optimal

Therefore to find optimal, just move along the direction of ‘improvement’ till you can’t
Analogy: think of a mountain with exactly one “peak”. Then as long as you keep climbing, you will eventually reach the top!

Locally optimal \equiv \text{globally optimal}

OPT I \quad \begin{align*}
\text{minimize} & \quad f(z) \quad \text{over} \quad z \in \mathbb{R}^n, \quad z \geq 0 \\
\text{subject to} & \quad A z \leq c
\end{align*}

Let \( x^L \) be locally optimal solution of the above optimization problem. That is, there exists some \( \delta > 0 \), such that \( x^L \) is an optimal solution of

\begin{align*}
\text{minimize} & \quad f(z) \quad \text{over} \quad z \in \mathbb{R}^n, \quad z \geq 0 \\
\text{subject to} & \quad A z \leq c \quad \text{and} \quad \|z - x^L\| \leq \delta.
\end{align*}
Result: \( x^L \) is a solution of the opt 1.

Proof: Suppose \( x^* \) is an optimal solution such that \( f(x^*) < f(x^L) \). Now consider

\[
x(\theta) = \theta x^* + (1-\theta) x^L, \quad \theta \in [0,1]
\]

Now, \( x(\theta) \) is feasible solution of opt 1 since

\[
x(\theta) \geq 0 \text{ because } x^*, x^L \geq 0.
\]

\[
Ax(\theta) = \theta Ax^* + (1-\theta) Ax^L 
\leq \theta c + (1-\theta) c = c.
\]

\[
f(x(\theta)) \leq \theta f(x^*) + (1-\theta) f(x^L) \text{ because } f \text{ convexity.}
\]
For any $\theta \in (0,1)$,

\[ f(x(\theta)) < f(x^*) \text{ because of } f(x^*) < f(x^\perp). \]

That means, $x^\perp$ cannot be locally optimal since as $\theta \to 0^+$, $x(\theta) \to x^\perp$.

This is a contradiction to our assumption that $f(x^*) < f(x^\perp)$. That is, $x^\perp$ is indeed an optimal solution to \text{OPT} \, 1.
Gradient algorithm:

First, we consider algorithm for simpler version: unconstrained, i.e.

$$\text{minimize } f(x) \text{ over } x \in \mathbb{R}^n$$

Start with an arbitrary initial value, $x_0 \in \mathbb{R}^n$. Iteratively, update as follows.

For some $\delta > 0$, $t \geq 0$,

$$x_{t+\delta} = x_t - \nabla f(x_t) \delta$$

Equivalently,

$$\frac{x_{t+\delta} - x_t}{\delta} = -\nabla f(x_t)$$

That is,

$$\frac{dx_t}{dt} = -\nabla f(x_t) \iff \frac{dx_{r,t}}{dt} = -\frac{\partial f(x_t)}{\partial x_r}$$
Consider $f(x_t)$:

$$\frac{d}{dt} f(x_t) = \sum_{r=1}^{R} \frac{\delta f(x_t)}{\delta x_r} \cdot \frac{\partial x_r}{\partial t}$$

$$= -\sum_{r=1}^{R} \left( \frac{\partial f(x_t)}{\partial x_r} \right)^2$$

$$\leq 0$$

Further, it is equal to 0 only if

$$\nabla f(x_t) = 0 \iff x_t \text{ is locally optimal}.$$ 

$$\iff x_t \text{ is globally optimal}.$$ 

In summary, gradient algorithm finds local/global solution for unconstrained convex optimization.
Our question: constrained optimization

minimize $f(z)$ over $z \in \mathbb{R}^n$.

subject to $Z \geq 0$

$A_z \leq C_z$.

This has constrains and hence the gradient algorithm, as is, may not work.

A simple idea/fix: penalty and intercept

penalty: for each constraint, add appropriate 

‘barrier’ or ‘penalty’ function in

objective.

In our setup, for each link $l$, constraint is

$A_z z \leq C_z$, add penalty $p_z(\cdot)$ s.t.
\[ p_e(z) = \begin{cases} 
0 & \text{if } A_\epsilon z \leq C \\
\infty & \text{otherwise} 
\end{cases} \]

Then, our setup is equivalent to:

\[
\text{minimize } f(z) + \sum_\epsilon p_e(z) \text{ over } z \in \mathbb{R}^n
\]

subject to \( z \geq 0 \).

To keep objective \( f(z) + \sum_\epsilon p_e(z) \) convex, we consider "convex" approximation like

\[ p^\delta_e(z) = \exp\left(\frac{\delta}{2} (A_\epsilon z - C_\epsilon)\right) \quad \text{as } \delta \to \infty \]
New optimization:

\[
\text{minimize } f(z) + \sum_{\ell} \exp\left(\frac{z}{\lambda_{\ell}} \left( A_{\ell} z - c_{\ell} \right) \right)
\]

over \( z \in \mathbb{R}^n, z \geq 0 \).

Algorithm:

Initially, start with \( x_0 \) that is feasible for original problem: \( x_0 \in \mathbb{R}^n, x_0 \geq 0, A x_0 \leq c \).

Gradient update:

\[
\frac{d x_{r,t}}{dt} = - \log \frac{x_{r,t}}{\lambda_r} - 3 \sum_{\ell} A_{\ell r} \exp\left(3 \left( A_{\ell} x_{r,t} - c_{\ell} \right) \right)
\]

\[
= - \log \frac{x_{r,t}}{\lambda_r} - 3 \sum_{\ell} A_{\ell r} p_{\ell}^3 (x_{t,\ell})
\]

marginal benefit penalty

Convergence \( \iff \) marginal benefit = penalty