STABILITY IN QUEUING SYSTEMS

Report for EE 451: Supervised Research Exposition

In this paper we discuss the subtleties regarding stability of dynamical systems (mainly stochastic). For our analysis we will look at discrete time systems i.e. Discrete Time Markov Chain (DTMC) analysis. We will also investigate into some other famous results and techniques like martingales and their connection to Lyapunov Stability. We will throughout also discuss the proofs of some of these results.

ABHINAV SINHA

07D07012

Abhinav Sinha

Stability in Queuing Systems

Report for EE 451: Supervised Research Exposition

Keywords: MARKOV CHAINS, ERGODICITY, POSITIVE RECURRENCE, LYAPUNOV FUNCTION, LYAPUNOV STABILITY, STATIONARY PROCESS, FLUID LIMITS, MARTINGALES and G/G/1 QUEUE.

Table of Contents

I. Introduction

Existence and uniqueness (also convergence) of a stationary solution to a stochastic dynamical system has been at the centre of study for many years. Although there exist many results for stability in specific cases we would like to present here a general result known as Lyapunov's Theorem which will be a sufficient condition. For this we will also introduce the concept of drift. We will prove it using a method similar to martingales difference sequence. Because of large number of systems that may be stable finding a necessary condition maybe more difficult. Even before we begin we must note that although we have a sufficient condition in our hands, concluding stability is still not an easy problem mainly because a sufficient condition still requires us to find appropriate Lyapunov function which barring a few basic cases is still quite difficult.

We will also visit some other famous results used in queueing theory including results for G/G/1 queue. We take a look at other stability techniques like Fluid models. At the end we discuss about martingales themselves.

II. FORMALISM

In here we will first look at the relevant definitions

Preliminaries

Whenever we talk of equations for stochastic processes we generally talk about both sides being equal along most sample paths or being equal in probability or distribution, in this regard we state the following definition,

Almost Sure Convergence:

$$X_n \overset{a.s}{\rightarrow} X \quad or \quad X_n = X \;\; wp\, 1 \quad means \quad \Pr\left(\left\{\omega : \lim_{n\to\infty} X_n(\omega) = X\right\}\right) = 1$$

Similarly we can also have convergence in probability or simply distribution. Note that such convergences are weaker than almost sure convergence.

The concept of stationarity is one of the most helpful while analyzing stochastic processes because if the system is proved to converge towards a stationary system then analysis over large times can simply be performed on the stationary equivalent, which would be easier. Definition is as follows,

A process $\{X_n\}_{n\geq 1}$ is ***stationary*** iff joint distribution of any collection of Random Variables from the process is invariant of uniform shift in its argument, i.e.

$$F_{X_{i_1},\dots,X_{i_l}}\left(t_{1,\dots,}t_l\right) = F_{X_{i_1},\dots,X_{i_l}}\left(s + t_{1,\dots,}s + t_l\right) \quad \forall\, s, t_i$$

In our case a stationary solution would mean the existence of a stationary distribution π on S such that

$$\pi = \pi.\overline{P}$$

where $\overline{P}$ is the state transition matrix.

Basic setup and some more definitions

We would be studying a DTMC $\{X_n\}$ with values in a general state space S (which is assumed to be a completely separable metric space), defined by $\bar{P}$, the state transition matrix. Also worth noting is that whenever the state space is finite we can represent the DTMC as a recurrence relation.

Drift, is defined as the expected change in the state of the Markov chain as enumerated by an appropriate increasing function, let $g: S \to N$ be map numbering all possible states and let $V: S \to R_+$ be a function (the Lyapunov Function) on the state space, then the drift can be defined as

$$DV(x, g) := \boldsymbol{E}_x[V(X_{g(x)}) - V(x)]$$

And for a simple function like $g(x) = n$, we get an n-step drift which is

$$DV(x, n) = \boldsymbol{E}_x[V(X_n) - V(X_0)|X_0 = x]$$

Positive recurrence, a measurable set $B \subseteq S$ is said to be recurrent if expected time of return to the set is finite w.p. 1.

$\tau_B = \inf\{n \geq 1 : X_n \in B\}$ is minimum time of return and for the set B to be ***positive recurrent*** we must have

$$\sup_{x \in B} E_x(\tau_B) < \infty$$

III. Statement and proof

In this section we will state two versions (basic and generalized) of Lyapunov theorem for stability of DTMC. We will be using definitions from previous section. We will prove them and also discuss a few things about Lyapunov Theorem. The method of the proof will tell us a lot about the techniques one can use for proving stability.

Basic Lyapunov Theorem

For N_0, c, H positive constants, if we have

$$DV(x,1) \leq -c \qquad if \qquad V(x) > N_0$$

and

$$DV(x,1) \leq H < \infty \qquad if \qquad V(x) \leq N_0$$

then the set $B = \{x : V(x) \leq N_0\}$ is positive recurrent.

Now we discuss the assumptions and why they are necessary for generalized Lyapunov theorem.

Assumptions

A1. V is unbounded from above: $sup_{x \in S} V(x) \to \infty$

A2. h is bounded from below: $inf_{x \in S} h(x) > -\infty$

A3. h is eventually positive: $\underline{lim}_{V(x) \to \infty} h(x) > 0$

A4. g is locally bounded from above: $G(N) = sup_{V(x) \leq N} g(x) < \infty \quad \forall\, N > 0$

A5. g is eventually bounded by h: $\overline{lim}_{V(x) \to \infty} {g(x)}/{h(x)} < \infty$

Generalized Lyapunov Theorem

Let $\tau \stackrel{\text{def}}{=} \tau_N = \inf\{n \geq 1 : V(X_n) \leq N\}$

If $DV(x,g) \leq -h(x)$

Then there exists N_0 such that for all $N \geq N_0$ and any $x \in S$ then the set $B_N = \{x \in S : V(x) \leq N\}$ is positive recurrent.

Proof:

We will do the proof for the generalized case and basic theorem will follow.

The assumptions can all be justified as we go along with the proof (except A5 which will be discussed later).

As $V(X_0)$ will come out of the expectation in the drift condition we can see that $V(x) - h(x) \geq 0 \quad \forall \;\; x$, thus we decide to choose N_0 such that $\inf_{V(x)>N_0} h(x) > 0$. Then for all $N \geq N_0$, we can set

$$d = \sup_{V(x)>N} \frac{g(x)}{h(x)} \qquad\qquad H = -\inf_{x\in S} h(x) \qquad\qquad c = \inf_{V(x)>N} h(x)$$

We will now define stopping times (increasing sequence), recursively,

$$t_0 = 0, \quad t_n = t_{n-1} + g\left(X_{t_{n-1}}\right), \;\; n \geq 1$$

Clearly the sequence $Y_n = X_{t_n}$ forms a *Markov Chain* (Strong Markov Property). From the definition of H, and the drift condition we easily prove by induction on n, $\boldsymbol{E}_x V(Y_{n+1}) \leq \boldsymbol{E}_x V(Y_n) + H$, hence we can see that $\boldsymbol{E}_x V(Y_n) < \infty \;\; \forall \;\; x, n$. We define another stopping time as follows

$$\gamma = inf \;\; \{n \geq 1 : V(Y_n) \leq N\} \;\; \leq \infty$$

Note that γ is defined so that $\mathbf{1}(\gamma \geq i) \in \mathcal{F}_{i-1}$ (where $\mathcal{F}_{i-1}$ is the $(i-1)^{th}$ filtration generated by $Y_0, \ldots, Y_{i-1}$) i.e. we can decide about its *indicator* function just by looking at the past of the process and there is no need for any information from the future.

So now we have $\tau \leq t_\gamma$ (a.s) and hence proving that $\boldsymbol{E}_x t_\gamma < \infty$ is sufficient.

Let's define a *cumulative energy*, between 0 *and* $\gamma \wedge n$ by

$$\mathcal{E}_n = \sum_{i=0}^{\gamma\wedge n} V(Y_i) = \sum_{i=0}^{n} V(Y_i)\, \mathbf{1}(\gamma \geq i)$$

And estimate the change $\boldsymbol{E}_x(\mathcal{E}_n - \mathcal{E}_0)$ as follows

$$\boldsymbol{E}_x(\mathcal{E}_n - \mathcal{E}_0) = \boldsymbol{E}_x \sum_{i=1}^{n} \boldsymbol{E}_x(V(Y_i)\, \mathbf{1}(\gamma \geq i) \mid \mathcal{F}_{i-1})$$

$$= \boldsymbol{E}_x \sum_{i=1}^{n} \mathbf{1}(\gamma \geq i)\, \boldsymbol{E}_x(V(Y_i) \;\mid \mathcal{F}_{i-1})$$

$$\leq \boldsymbol{E}_x \sum_{i=1}^{n} \mathbf{1}(\gamma \geq i)\ (V(Y_{i-1}) - h(Y_{i-1}))$$

$$\leq \boldsymbol{E}_x \sum_{i=1}^{n+1} \mathbf{1}(\gamma \geq i-1)\ V(Y_{i-1}) - \boldsymbol{E}_x \sum_{i=1}^{n} h(Y_{i-1})$$

$$= \boldsymbol{E}_x(\mathcal{E}_n) - \ \boldsymbol{E}_x \sum_{i=0}^{n-1} h(Y_i)\ \mathbf{1}(\gamma \geq i)$$

Here we have used $V(x) - h(x) \geq 0$, and also that the indicator function increases when we change $i\ to\ i-1$. Now we finally have from above

$$\boldsymbol{E}_x \sum_{i=0}^{n-1} h(Y_i)\ \mathbf{1}(\gamma \geq i)\ \leq\ V(x)$$

For $V(x) > N$ we have $V(Y_i) > N\ for\ i < \gamma$ (definition of γ). So,

$$h(Y_i)\ \geq c > 0, \quad for\ \ i < \gamma$$

(this from definition of c). So we obtain from above two equations that

$$c.\boldsymbol{E}_x \sum_{i=0}^{n} \mathbf{1}(\gamma > i)\ \leq\ V(x) + H + c$$

So we have

$$c.\boldsymbol{E}_x\gamma\ \leq\ V(x) + H + c < \infty$$

Now using the fact that $h(x) \geq d.g(x)$ for $V(x) > N$ we can write

$$\sum_{i=o}^{\gamma-1} h(Y_i) \geq d. \sum_{i=o}^{\gamma-1} g(Y_i) = d.t_\gamma$$

Where $t_\gamma < \infty$ $(a.s)$ and so

$$\boldsymbol{E}_x\tau \leq \boldsymbol{E}_x t_\gamma \leq \frac{V(x) + H + c}{cd}$$

hence proved.

For the other case also we can show by conditioning on Y_1. PROOF IS COMPLETE.

Assumption: A5

We can construct a case of forward transition with only backward transition possible to state 1 (with probability $^1/_k$ if you are in state k) then expected time of return will still be ∞, but we can satisfy conditions in the Lyapunov Theorem by taking $V(k) = \log(1 \vee \mathrm{k}) \quad and \quad g(k) = k^2$ we can see that $h(k) = c_1 V(k) - c_2$ (for appropriate constants) will satisfy the drift condition and yet M.C. is not positive recurrent.

Other words about Lyapunov theorem:

Lyapunov drift for network stability is first used for multi-hop networks, and for opportunistic downlink scheduling. Related quadratic Lyapunov functions are used to make stability and delay claims for $N \times N$ packet switches and multi-hop mobile networks. Non-quadratic Lyapunov functions can sometimes be used to make modified or improved statements about delay. Alternative Lyapunov functions via queue groupings can often lead to improved complexity and/or delay bounds.

IV. Special Cases

(i) Pake's Lemma

Statement: If we have $V(x) \geq 0, \forall\, x$ and if $\exists\, \epsilon > 0$ such that

$$E[V(X_{t+1}) - V\,(X_t) \mid X_t = x] \leq -\epsilon \;\; < \; 0,$$

for all x except on a finite set C, then $\{X_t\}$ is positive recurrent.

To prove this we can take S as $\mathbb{Z}$ and consider one step drift i.e. $g(x) = 1$ and $h(x) = \epsilon - C.\mathbf{1}\{V(x) \leq D\}$ in Generalized Lyapunov Theorem.

(ii) *Foster – Lyapunov Criteria*

For this we will have S as general and $g(x) = 1$ and $h(x) = c - H.\mathbf{1}(V(x) \leq N_0)$, where the constants come from the proof.

(iii) Meyn – Tweedie Criteria – Here we would take

$$h(x) = g(x) - C.\mathbf{1}\{V(x) \leq D\}$$

(iv) Dai's Criteria (Fluid Limits Criteria) – Here we would take

$g(x) = ceiling\,[V(x)]$ and $h(x) = \varepsilon.V(x) - C.\mathbf{1}\{V(x) \leq D\}$.

V. Examples and More Techniques

- *Martingales*

Martingales play a role in stochastic processes similar to that played by *conserved quantities* (sometimes called first integrals) in dynamical systems. Unlike a first integral of a dynamical system, which remains constant in time, a martingale's value can change; however, its expectation remains constant in time. More important, the expectation of a martingale is unaffected by optional sampling. This can be used as a provisional definition: A discrete-time martingale is a sequence $\{X_n\}_{n\geq 0}$ of real (or complex) random variables satisfies

$$EX_\tau = EX_0$$

for every bounded stopping time τ.

Definition

Let $\{\mathfrak{F}_n\}_{n\geq 0}$ be an increasing sequence of $\sigma - algebras$ in a probability space $(\Omega, \mathfrak{F}, P)$. They are called filtrations. For $\{X_n\}_{n\geq 0}$, a sequence of real-valued random variables with the property that for each n the random variable X_n is measurable relative to $\mathfrak{F}_n$. This would mean that each of the X_n are some function of $Y_1, Y_2, \ldots, Y_n$

(some sequence whose cumulative $\sigma - algebra$ is $\mathfrak{F}_n$). Then the sequence X_n is *martingale* if for every n we have,

$$E(X_{n+1} \mid \mathfrak{F}_n) = E(X_n)$$

The most common example of a martingale is the sequence of partial sums of zero mean independent random variables, where we can show that if $X_n = \sum_1^n Y_j$, then the sequence $\{X_n\}$ will be martingale w.r.t $0, Y_1, Y_2, \ldots.$ (filtrations generated by it), where we can see

$$E(X_{n+1} \mid \mathfrak{F}_n) = E(X_n + Y_{n+1} \mid \mathfrak{F}_n)$$

$$= E(X_n \mid \mathfrak{F}_n) + E(Y_{n+1} \mid \mathfrak{F}_n)$$

$$= X_n + E(Y_{n+1} \mid \mathfrak{F}_n)$$

$$= X_n$$

St. Petersburg's Game (Paradox)

This is a famous problem (one may have encountered it in the book *Gambler* by Fyodor Dostoevsky) in which a gambler must maximize his/her profit by choosing to wager an amount W_n in the n^{th} instance, assume the outcome is binary in each instance $\xi_i = 1\ if\ heads\ and\ -1\ if\ tails$ (with equal probability, i.e. fair game). Decisions about the wager each time, W_n ,can be made by looking at all previous results i.e. ξ_i , $1 \le i \le n-1$ but not ξ_n (or later). Then (assuming we bet on heads every time we bet because it's a fair game) what we wish to maximize is

$$S_n = (W.X)_n = \sum_{k=1}^{n} W_k \xi_k$$

the net winnings in n games where $X_n = \xi_1 + \xi_2 + \cdots + \xi_n$.

The result that we can prove from here is that Martingale transforms like this are martingales themselves because ξ_i are bounded and thus this transform is martingale relative to filtrations generated by ξ_i. We can see that

$$E((W.X)_{n+1} \mid \mathfrak{F}_n) = (W.X)_n + E(W_{n+1}\xi_{n+1} \mid \mathfrak{F}_n)$$

$$= (W.X)_n + W_{n+1} E(\xi_{n+1} \mid \mathfrak{F}_n)$$

$$= (W.X)_n$$

Maximal Inequality for Super-Martingales

We consider a simple example of a person A, starting with Rs. 100 and wanting to maximize his/her probability of winning a total of Rs. 200. A simple way is to wager all the money in the first bet and win a total of Rs. 200 with probability 0.5, what we are interested in is finding whether he/she can have a strategy which will give them more than 0.5 probability of winning Rs. 200. The answer turns out to be No and we can use maximal inequality for martingale transforms in the above setting to show this,

$F_n = 100 + \sum_{k=1}^{n} W_k \xi_k$ are the total winnings in n steps, since F_n is a martingale and each of $F_n \geq 0$ we can use the inequality

$$P(\sup_{n\geq 0} F_n \geq \alpha) \leq \frac{EX_0}{\alpha}$$

Here we will EX_0 as Rs. 100 and if we set $\alpha = 200$ then we can see the result straighaway.

- *Fluid – Queue Model & Fluid Limits*

The stochastic-process limits are also called fluid limits because the limit processes are *deterministic* functions of the form $c.t$ for some constant c.

For the fluid model we will consider queueing equation

$Q(t) = Q(0) + A(t) + \sum_{s=0}^{t} R^T \phi(s)$ and write down corresponding fluid equation as

$$Q'(t) = Q'(0) + \lambda t + \int_{s=0}^{t} R^T \phi(s)\, ds$$

where each term represents the usual arrival process and scheduling and users in system. We have changed from discrete to continuous.

Fluid Stability Theorem: If $\exists\, t_0$ *such that* $Q'(t) = 0\ \forall\ t \geq t_0$ then the system is stable.

Example: Consider a 2 queue system with scheduling policy to serve the longer queue then we have

$$Q_1(t) = Q_1(0) + A_1(t-1) - \sum_{s=1}^{t} \mathbf{1}(Q_1(s-1) > Q_2(s-1))$$

$$Q_2(t) = Q_2(0) + A_2(t-1) - \sum_{s=1}^{t} \mathbf{1}(Q_1(s-1) \leq Q_2(s-1), Q_2(s-1) > 0)$$

And now consider the scaled sytem by n, also take $Q_1^n(0) + Q_2^n(0) = n$ so that system's initial condition are scaled analogously. We will have

$$\frac{Q_1^n(nt)}{n} = \frac{Q_1^n(0)}{n} + \frac{A_1(nt-1)}{n} - \frac{1}{n}\sum_{s=1}^{nt} \mathbf{1}(\frac{Q_1^n(s-1)}{n} > \frac{Q_2^n(s-1)}{n})$$

$$\frac{Q_2^n(nt)}{n} = \frac{Q_2^n(0)}{n} + \frac{A_2(nt-1)}{n} - \frac{1}{n}\sum_{s=1}^{nt} \mathbf{1}(\frac{Q_1^n(s-1)}{n} \leq \frac{Q_2^n(s-1)}{n}, \quad \frac{Q_2^n(s-1)}{n} > 0)$$

For ergodic queues take $n \to \infty$ and from SLLN the limits should exist and we would have,

$$Q_1'(t) = Q_1'(0) + \lambda_1 t - \int_0^t \mathbf{1}(Q_1'(s^-) > Q_2'(s^-))\, ds$$

$$Q_2'(t) = Q_2'(0) + \lambda_2 t - \int_0^t \mathbf{1}(Q_1'(s^-) \leq Q_2'(s^-), Q_2'(s^-) > 0\,)\, ds$$

If $\lambda_1 + \lambda_2 < 1$ then $for\ all\ t > t_0 = \frac{1}{1-\lambda_1-\lambda_2}$ we will have $Q_1'(t) = Q_2'(t) = 0$, for we can see

$$\lim_{n\to\infty} \left(\frac{Q_1^n(nt_0) + Q_2^n(nt_0)}{n}\right)^1 = 0$$

So we have $\delta > 0$ *and* n_0 *such that*

$$E\big(Q_1^n(nt_0) + Q_2^n(nt_0)\big) \leq (1-\delta)n \quad \forall \quad n \geq n_0$$

This implies

$$E(Q_1^n(nt_0) + Q_2^n(nt_0) - n) \leq -\delta n$$

This gives a Lyapunov Equation for $Q'(t)$. END OF PROOF.

This technique can also be applied to many other types of networks like 1-station or 2-station Jackson Networks, Kelly Network, Multiclass Queues to deduce basic conditions on stability.

- *G/G/1 Queue*

This is the Single-server queue with first-in-first out discipline and with a general distribution on the sequences of inter-arrival and service times. Customers are numbered $n = 0, 1, \dots\dots$ We assume that customer 0 arrives to a system at time $t = 0$ and finds there an initial amount of work, so has to wait for W_0 units of time for the start of its service. Let t_n be the time between n^{th} and $n + 1^{st}$ arrivals and let s_n be the service time of n^{th} customer and let W_n be the waiting time in the system (time between arrival and start of service) for n^{th} customer. Then the sequence $\{W_n\}$ follows the *Lindley's Equations,*

$$W_{n+1} = (W_n + s_n - t_n)^+$$

We can again use Lyapunov function which is or any other approach (like *Spectral Analysis*) to conclude stability of this recursion whenever expected service time is lesser that expected inter-arrival time.

VI. Results and Discussions

From all of the above we have found out that Lyapunov Theorem is a generalized theorem and can be used to find conditions of stability in many different types of queues like *Lindley's Recursion, Jackson Network, Kelly Network* and many more. We also went through fluid limits. Fluid limits can also handle the stability problem as we saw and they make use of Lyapunov functions for proving optimality as well. Martingales provide yet another method to handle series summations and provide bounds on their expectations.

VII. Acknowledgements

I would like to thank my guide for clarifying my queries in the field as research is new to me and also for pinpointing the best material available to get a hold on techniques of this kind. I would also like to thank my family and friends for helping me adapt to the ways surveying various topics as well as research.

VIII. References

1. Lecture Notes by Sergey Foss, available online at http://www.web.abo.fi/fak/mnf/mate/tammerfors08.

2. Foster, F.G. (1953). On the stochastic matrices associated with certain queuing processes. Ann. Math. Stat. **24**, 355-360.

3. Meyn, S. and Tweedie, R. (1993). *Markov Chains and Stochastic Stability*. Springer, New York.