To appear in SIAM Journal on Applied Dynamical Systems (2019)

Non-linear Dynamics in Queueing Theory: Determining Size of Oscillations in Queues with Delayed Information*

Sophia Novitzky†, Jamol Pender‡, Richard H. Rand§, and Elizabeth Wesson†

Abstract. Internet and mobile services often provide waiting time or queue length information to customers. This information allows a customer to determine whether to remain in line or, in the case of multiple lines, better decide which line to join. Unfortunately, there is usually a delay associated with waiting time information. Either the information itself is stale, or it takes time for the customers to travel to the service location after having received the information. Recent empirical and theoretical work uses functional dynamical systems as limiting models for stochastic queueing systems. This work has shown that if information is delayed long enough, a Hopf bifurcation can occur and cause unwanted oscillations in the queues. However, it is not known how large the oscillations are when a Hopf bifurcation occurs. To answer this question, we model queues with functional differential equations and implement two methods for approximating the amplitude of these oscillations. The first approximation is analytic and yields a closed-form approximation in terms of the model parameters. The second approximation uses a statistical technique, and delivers highly accurate approximations over a wider range of parameters.

Key words. Hopf bifurcation, delay-differential equation, perturbations method, Lindstedt's method, operations research, queueing theory

AMS subject classifications. 34K99, 35Q94, 41A10, 37G15

1. Introduction.

The omnipresence of smartphone and internet technologies has created new ways for corporations and service system managers to interact with their customers. One important and common example of such communication is the delay announcement, which has become the main tool for service system managers to inform customers of their estimated waiting time. Delay announcements are common in settings like customer support call centers, appointment scheduling in healthcare services, restaurants during busy hours, public transportation, and even online shopping at Amazon.com.

The reason why delay announcements are so popular among service providers is that they are vital to customer experience. Moreover, delay announcements can influence the decisions of customers, and consequently affect the dynamics of the queueing system as seen in [17]. As a result, delay announcements are of major interest among researchers who aim to quantify the impact of such announcements on the queue length process or the virtual waiting time process. The work of [3, 11, 15, 4, 12, 20, 21, 1, 2, 35] and references therein focus on this aspect of the delay announcements.

The analysis of this paper is similar to the main thrust of the delay announcement literature in that it is concerned with the impact of information on the dynamics of the queueing process. However, the current literature focuses only on services that give the delay announcements to their customers in real-time, while we consider the scenario when the information is delayed. Information delay is commonly experienced in services that inform their customers about the waiting times prior to the customers' arrival to the service location. One example is the Citibike bike-sharing network in

*Submitted to the editors on February 12, 2018. Accepted for publication on November 30, 2018.

†Center for Applied Mathematics, Cornell University, Ithaca, NY 14853 (sn574@cornell.edu, enw27@cornell.edu).

‡School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853 (jjp274@cornell.edu).

§Sibley School of Mechanical and Aerospace Engineering, Department of Mathematics, Cornell University, Ithaca, NY 14853 (rand@math.cornell.edu).

New York City [10, 33]. Riders can search the availability of bikes on a smartphone app. However, in the time that it takes for the riders to leave their home and get to a station, all of the bikes could have been taken from that station.

Typically, queueing theorists use ordinary differential equations to model the mean dynamics of the queue length processes, but the incorporation of delayed information leads us to utilize delay differential equations (DDE's) in our first model and functional differential equations (FDE's) in our second model. As a result, this paper introduces mathematical techniques that are new in the context of queueing literature. We would like to note, however, that there is a paper [26] which combines concepts from queueing theory with DDE's, and applies them to sizing router buffers in Internet infrastructure services.

The authors in [24] use DDE's and FDE's to develop two new two-dimensional fluid models of queues that incorporate customer choice based on delayed queue length information, and show that oscillations in queue lengths occur for certain lengths of delay. By comparison, in this paper we prove that the observed behavior is due to a supercritical Hopf bifurcation, and we use two techniques to approximate the size of the amplitude of oscillations. The first method is a classical perturbations technique called Lindstedt's method. The second method, which we call the *slope function method*, is a numerical technique that we develop specifically to extend the range of parameters for which the Lindstedt's approximation maintains accuracy. Based on numerical results, the slope function method successfully reduces the maximum error in approximation over a range of parameters by $60-75\%$ when compared to Lindstedt's method. In the context of queueing models, the accuracy of approximation matters because the amplitude of queue oscillations can provide valuable insights such as the average waiting time during busier hours, the longest waiting time a customer can experience, and the optimal moment for joining a queue that will guarantee the quickest service. Moreover, our method of approximation is not restricted to queueing models and can be applied to any system where Hopf bifurcations are observed.

1.1. Paper Outline. This paper considers two models of queues that were originally presented in [24] and [25] as fluid limits of stochastic queueing models. In each model, there are two queues and customers decide which queue to join based on information about the queue length that is delayed. In section 2 we present the first model that uses a constant delay. At first, subsection 2.1 describes the qualitative behavior of the queues, stating the conditions for a unique stable equilibrium as well as the conditions for supercritical Hopf bifurcations. Then, we focus on the behavior of the queues when the stable equilibrium transitions into a stable limit cycle, and approximate the amplitude of the resulting oscillations. In subsection 2.2 - subsection 2.3, we use Lindstedt's method, which is accurate on a limited range of parameters. To broaden this range, in subsection 2.4 - subsection 2.5, we implement the slope function method, which is a technique that uses known amplitude of a small subset of queues and extrapolates it for a larger set of parameters. Overall, this method achieves higher accuracy than Lindstedt's method over a range of model parameters.

In section 3, we present the second model of queues that uses a moving average of the queue lengths as the delay announcement. The structure of section 3 is identical to section 2, where we describe the qualitative behavior of queue lengths and later approximate the amplitude of oscillations via Lindstedt's method and the slope function method. Finally, we compare the performance of the two techniques, and conclude by highlighting the strengths and weaknesses of each method.

2. Constant Delay Model. In a model with two infinite-server queues visualized by Figure 2.1, customers arrive at a rate $\lambda > 0$. Each customer is given a choice of joining either queue. The customer is told the length of each queue, and is likely to prefer the shorter queue. The probability

p_i of a customer joining the i^{th} queue is given by the Multinomial Logit Model (MNL)

$$p_i(q(t), \Delta) = \frac{\exp\big(-q_i(t-\Delta)\big)}{\exp\big(-q_1(t-\Delta)\big) + \exp\big(-q_2(t-\Delta)\big)}, \tag{2.1}$$

where $q_i(t)$ is the length of i^{th} queue at time t. The MNL is commonly used to model customer choice in fields of operations research, economics, and applied psychology [32, 16, 23, 34]. The delay $\Delta > 0$ accounts for the customers' travel time to the service location, or for the time lag between when the service manager measures the queue length and discloses this information to customers. The model assumes an infinite-server queue, which is customary in operations research literature [9, 18, 29]. This assumption implies that the departure rate for a queue is the service rate $\mu > 0$ multiplied by the total number of customers in that queue. Therefore the queue lengths can be described by

$$\dot{q}_1(t) = \lambda \cdot \frac{\exp\big(-q_1(t-\Delta)\big)}{\exp\big(-q_1(t-\Delta)\big) + \exp\big(-q_2(t-\Delta)\big)} - \mu q_1(t) \tag{2.2}$$

$$\dot{q}_2(t) = \lambda \cdot \frac{\exp\big(-q_2(t-\Delta)\big)}{\exp\big(-q_1(t-\Delta)\big) + \exp\big(-q_2(t-\Delta)\big)} - \mu q_2(t) \tag{2.3}$$

for $t > 0$, with initial conditions specified by nonnegative continuous functions f_1 and f_2

$$q_1(t) = f_1(t), \quad q_2(t) = f_2(t), \quad t \in [-\Delta, 0]. \tag{2.4}$$

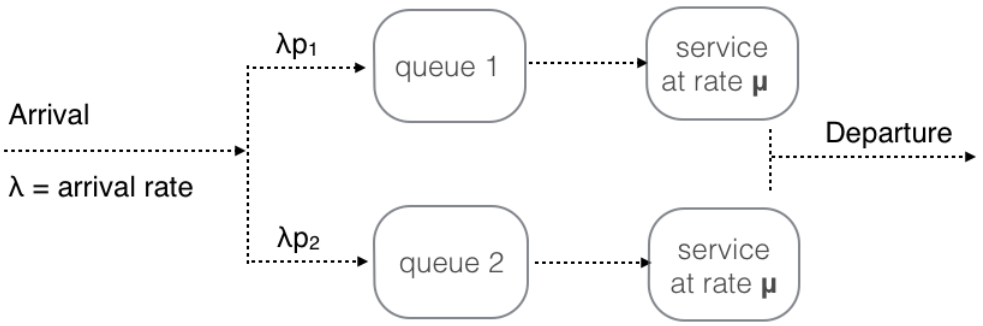

Figure 2.1: Customers going through a two-queue service system.

It is worth to noting that Equations (2.2) - (2.3) can be uncoupled when the sum and the difference of q_1 and q_2 is taken. The system is then reduced to the equations

$$\dot{v}_1(t) = \dot{q}_1(t) - \dot{q}_2(t) = \lambda \tanh\Big(-\frac{1}{2} v_1(t-\Delta)\Big) - \mu v_1(t), \tag{2.5}$$

$$\dot{v}_2(t) = \dot{q}_1(t) + \dot{q}_2(t) = \lambda - \mu v_2(t), \tag{2.6}$$

where $v_2(t)$ is solvable, and the equation for $v_1(t)$ is of a form commonly studied in the literature. Many papers, such as [19, 36, 38, 37, 6], prove properties for models similar to ours. In [38], the author uses asymptotic analysis to prove uniqueness and stability of the slowly oscillating periodic solutions that occur under certain parameter restrictions. The authors in [30] study the floquet multipliers. We complement these results by developing an approximation for the amplitude of the oscillations near the first bifurcation point.

2.1. Hopf Bifurcations in the Constant Delay Model. In this section, we discuss the qualitative behavior of the queueing system given by Equations (2.2) - (2.3). We will begin by establishing the existence and uniqueness of the equilibrium.

Theorem 2.1. *For sufficiently small Δ, the unique equilibrium to the system of N equations*

$$\dot{q}_i(t) = \lambda \cdot \frac{\exp\big(-q_i(t-\Delta)\big)}{\sum_{j=1}^{N} \exp\big(-q_j(t-\Delta)\big)} - \mu q_i(t) \qquad \forall i = 1, 2, \cdots, N \tag{2.7}$$

is given by

$$q_i^* = \frac{\lambda}{N\mu} \quad \forall i = 1, 2, \cdots, N. \tag{2.8}$$

Proof. See the Appendix for the proof. ∎

Therefore the equilibrium of the queues from Equations (2.2) - (2.3) is given by

$$q_1^* = q_2^* = \frac{\lambda}{2\mu}. \tag{2.9}$$

Next, we consider the stability of the equilibrium, which can be determined by the stability of the linearized system of equations [13, 31]. Hence, subsection 5.1.3 and subsection 5.1.4 in the Appendix linearize the system of Equations (2.2) - (2.3) and separate the variables, reducing the system from two unknown functions to one:

$$\dot{\tilde{v}}_2(t) = -\frac{\lambda}{2} \cdot \tilde{v}_2(t-\Delta) - \mu \tilde{v}_2(t). \tag{2.10}$$

Assuming a solution of the form $\tilde{v}_2(t) = \exp(\Lambda t)$, the characteristic equation is

$$\Phi(\Lambda, \Delta) = \Lambda + \frac{\lambda}{2} \exp(-\Lambda\Delta) + \mu = 0. \tag{2.11}$$

The equilibrium is stable whenever the real part of every eigenvalue Λ is negative. It is evident from the characteristic equation that any real root Λ must be negative. However, there are infinitely many complex roots. In the next result, we will show that for a sufficiently small delay, all complex eigenvalues have negative real parts.

Proposition 2.2. *For Equations (2.2) - (2.3), as the delay approaches 0, i.e. $\Delta \to 0^+$, the real part of any complex eigenvalue approaches negative infinity.*

Proof. When $\Delta = 0$, the characteristic equation (2.11) has only one eigenvalue, namely $\Lambda = -\frac{\lambda}{2} - \mu$. When the delay is raised above 0, the characteristic equation becomes transcendental and an infinite sequence of roots is born. Since $\Phi(\Lambda, \Delta)$ is continuous with respect to both Λ and Δ, each eigenvalue Λ must be continuous with respect to Δ. Hence the real part of Λ must go to positive infinity or to negative infinity as the delay approaches 0. However, any root with positive real part is bounded as shown in the Appendix by Proposition 5.1, so the real part of any complex eigenvalue must go to negative infinity. ∎

By Proposition 2.2, all eigenvalues have negative real parts when Δ is small, so the equilibrium is stable until a pair of complex eigenvalues reaches the imaginary axis. To find when the equilibrium loses stability, we assume $\Lambda = i\omega_{cr}$ with $\omega_{cr} > 0$, plug Λ into the characteristic equation (2.11),

and separate the real and imaginary parts into two equations. We use the trigonometric identity $\cos^2(\omega\Delta) + \sin^2(\omega\Delta) = 1$ to find

$$\Delta_{cr}(\lambda, \mu) = \frac{2\arccos(-2\mu/\lambda)}{\sqrt{\lambda^2 - 4\mu^2}}, \quad \omega_{cr} = \sqrt{\frac{\lambda^2}{4} - \mu^2}. \tag{2.12}$$

For ω_{cr} to be real and nonzero the condition $\frac{\lambda^2}{4} - \mu^2 > 0$ must hold, so $\lambda > 2\mu$. If this condition is met, the equilibrium becomes unstable when Δ exceeds the **smallest positive root** of Δ_{cr} from Equation (2.12).

Theorem 2.3. *If $\lambda < 2\mu$, the equilibrium is locally stable for all $\Delta > 0$. If $\lambda > 2\mu$, the equilibrium is locally stable when Δ is less than the smallest positive root of Δ_{cr}.*

Proof. As discussed above, all eigenvalues of the characteristic equation are on the negative real side of the complex plane, unless $0 \neq \omega_{cr} \in \mathbb{R}$, and the delay reaches Δ_{cr}. ■

Figure 2.2 - Figure 2.3 show the behavior of the queues before and after the equilibrium loses stability. As suggested by Figure 2.3 and proved by the next result, the conditions (2.12) specify where the Hopf bifurcations occur. We note that if $\lambda > 2\mu$, there will be infinitely many Hopf bifurcations as the delay grows, since the expression for Δ_{cr} has infinitely many roots.

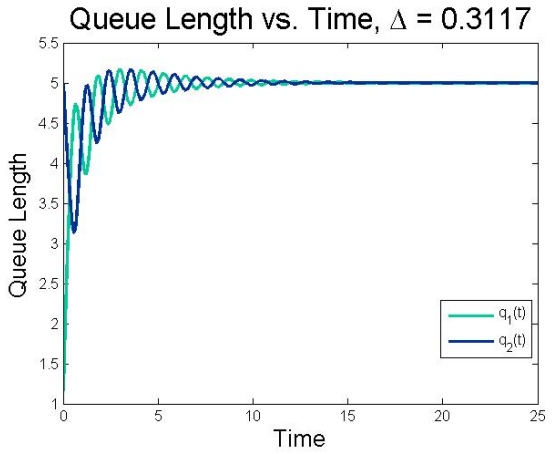

Figure 2.2: $\lambda = 10$, $\mu = 1$, $\Delta < \Delta_{cr}$.

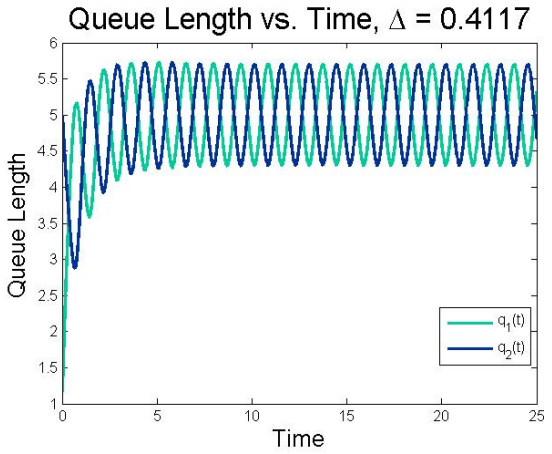

Figure 2.3: $\lambda = 10$, $\mu = 1$, $\Delta > \Delta_{cr}$.

Theorem 2.4. *If $\lambda > 2\mu$, a Hopf bifurcation occurs at $\Delta = \Delta_{cr}$, where Δ_{cr} is given by*

$$\Delta_{cr}(\lambda, \mu) = \frac{2\arccos(-2\mu/\lambda)}{\sqrt{\lambda^2 - 4\mu^2}}. \tag{2.13}$$

Proof. When $\Delta = \Delta_{cr}$, there is a a pair of purely imaginary eigenvalues Λ and $\bar{\Lambda}$. Further, $\mathrm{Re}\,\Lambda'(\Delta_{cr}) > 0$. We show this by introducing $\Lambda = \alpha(\Delta) + i\omega(\Delta)$ into the characteristic equation (2.11), separating the real and imaginary parts into two equations, and implicitly differentiating with respect to delay. We find $\frac{d\omega}{d\Delta}(\Delta_{cr})$ to be

$$\frac{d\omega}{d\Delta}(\Delta_{cr}) = \frac{\frac{\lambda}{2}e^{-\alpha\Delta}\Big(\cos(\omega\Delta)\omega - \sin(\omega\Delta)(\alpha'\Delta + \alpha)\Big)}{1 - \frac{\lambda\Delta}{2}\cos(\omega\Delta)e^{-\alpha\Delta}} = -\frac{\omega_{cr}(\alpha'\Delta_{cr} + \mu)}{1 + \mu\Delta_{cr}}, \tag{2.14}$$

This result is used to determine Re $\Lambda'(\Delta_{cr}) = \frac{d\alpha}{d\Delta}(\Delta_{cr})$:

$$\alpha' - \frac{\lambda}{2}e^{-\alpha\Delta}(\alpha'\Delta + \alpha)\cos(\omega\Delta) - \frac{\lambda}{2}e^{-\alpha\Delta}\sin(\omega\Delta)(\omega'\Delta + \omega) = 0, \tag{2.15}$$

$$\frac{d\alpha}{d\Delta}(\Delta_{cr}) = \frac{\omega_{cr}^2}{(1+\mu\Delta_{cr})^2 + \omega_{cr}^2\Delta_{cr}^2} > 0 \quad \forall \Delta_{cr} > 0. \tag{2.16}$$

where we use that at Δ_{cr}, $\alpha = 0$, $\omega = \omega_{cr}$, $\sin(\Delta_{cr}\omega_{cr}) = \frac{2\omega_{cr}}{\lambda}$, and $\cos(\Delta_{cr}\omega_{cr}) = -\frac{2\mu}{\lambda}$.

At each root of Δ_{cr} there is one purely imaginary pair of eigenvalues, but all other eigenvalues necessarily have a nonzero real part. This implies that all roots $\Lambda_j \neq \Lambda, \bar{\Lambda}$ satisfy $\Lambda_j \neq m\Lambda$ for any integer m. Hence, all conditions of the infinite-dimensional version of the Hopf theorem from [13] are satisfied, so a Hopf bifurcation occurs at every root of Δ_{cr}. ∎

Once the equilibrium loses stability, a limit cycle emerges. We now show that the resulting limit cycle is stable.

Theorem 2.5. *The Hopf bifurcations given by Theorem 2.4 are supercritical, i.e. each Hopf produces a stable limit cycle in its center manifold.*

Proof. One way to establish stability of limit cycles is by the method of *slow flow*, or the method of multiple scales. This method has previously been applied to systems of DDE's [7, 5, 22]. Another standard way to determine the stability of limit cycles is by showing that the floquet exponent has negative real part, as outlined in Hassard et al. [14]. In this theorem, we follow the first approach (the method of slow flow), but for the interest of the reader we include the floquet exponent method in the Appendix, subsection 5.1.5. We note that the results of the two methods agree.

We consider the third order polynomial expansions of q_1 and q_2 about the equilibrium. The resulting equations can be uncoupled, with the function of our interest given by

$$\dot{\tilde{v}}_2(t) = \lambda\left(-\frac{\tilde{v}_2(t-\Delta)}{2} + \frac{\tilde{v}_2^3(t-\Delta)}{24}\right) - \mu\tilde{v}_2(t). \tag{2.17}$$

For the details, see subsection 5.1.3 - subsection 5.1.4 of the Appendix. We set $\tilde{v}_2(t) = \sqrt{\epsilon}x(t)$ in order to prepare the DDE for perturbation treatment, and replace the independent variable t by two new time variables $\xi = \omega t$ (stretched time) and $\eta = \epsilon t$ (slow time). The delay and frequency are expanded about the critical Hopf values, $\Delta = \Delta_{cr} + \epsilon\alpha, \quad \omega = \omega_{cr} + \epsilon\beta$, so $\dot{x}$ becomes

$$\dot{x} = \frac{dx}{dt} = \frac{\partial x}{\partial \xi}\frac{d\xi}{dt} + \frac{\partial x}{\partial \eta}\frac{d\eta}{dt} = \frac{\partial x}{\partial \xi}\cdot(\omega_{cr} + \epsilon\beta) + \frac{\partial x}{\partial \eta}\cdot\epsilon. \tag{2.18}$$

The expression for $x(t-\Delta)$ may be simplified by Taylor expansion for small ϵ:

$$x(t-\Delta) = x(\xi - \omega\Delta, \eta - \epsilon\Delta) = \tilde{x} - \epsilon(\omega_{cr}\alpha + \Delta_{cr}\beta)\cdot\frac{\partial\tilde{x}}{\partial\xi} - \epsilon\Delta_{cr}\frac{\partial\tilde{x}}{\partial\eta} + O(\epsilon^2), \tag{2.19}$$

where $x(\xi - \omega_{cr}\Delta_{cr}, \eta) = \tilde{x}$. The function x is represented as $x = x_0 + \epsilon x_1 + \ldots$, yielding

$$\frac{dx}{dt} = \omega_{cr}\frac{\partial x_0}{\partial\xi} + \epsilon\beta\frac{\partial x_0}{\partial\xi} + \epsilon\frac{\partial x_0}{\partial\eta} + \epsilon\omega_{cr}\frac{\partial x_1}{\partial\xi}. \tag{2.20}$$

After the proposed transformations are carried out, the DDE (2.17) can be separated into two equations by collecting the terms with like powers of ϵ,

$$\omega_{cr}\frac{\partial x_0}{\partial\xi} + \frac{\lambda}{2}\tilde{x}_0 + \mu x_0 = 0, \tag{2.21}$$

$$\omega_{cr}\frac{\partial x_1}{\partial\xi} + \frac{\lambda}{2}\tilde{x}_1 + \mu x_1 = -\beta x_{0\xi} - x_{0\eta} + \frac{\lambda}{2}\big(\beta\Delta_{cr} + \alpha\omega_{cr}\big)\cdot\tilde{x}_{0\xi} + \frac{\lambda}{24}\tilde{x}_0^3. \tag{2.22}$$

Equation (2.21) shows that x_0 can be written as $x_0(t) = A(\eta)\cos(\xi) + B(\eta)\sin(\xi)$. Eliminating the secular terms $\sin(\xi)$ and $\cos(\xi)$ in Equation (2.22), we get two equations that involve $\frac{d}{d\eta}A(\eta)$ and $\frac{d}{d\eta}B(\eta)$, and we remove the delay terms by using Equation (2.21). We switch into polar coordinates by introducing $R(\eta) = \sqrt{A(\eta)^2 + B(\eta)^2}$, and we find $\frac{dR}{d\eta}$:

$$\frac{dR}{d\eta} = -\frac{R\Big((\Delta_{cr}\lambda^2 + 4\mu)R^2 - 16\alpha(\lambda^2 - 4\mu^2)\Big)}{16(4 + \Delta_{cr}^2\lambda^2 + 8\Delta_{cr}\mu)}. \tag{2.23}$$

Since $R \geq 0$ by definition, the two equilibrium points are $R_1 = 0$, which is unstable, and $R_2 = \sqrt{\frac{16\alpha(\lambda^2-4\mu^2)}{(\Delta_{cr}\lambda^2+4\mu)}}$, which is stable. Thus the limit cycle born when Δ exceeds any root of Δ_{cr} is locally stable in its center manifold. ■

To summarize, the queues converge to an equilibrium regardless of the delay when $\lambda < 2\mu$. However, when $\lambda > 2\mu$, infinitely many pairs of complex eigenvalues will (one by one) cross the imaginary axis from negative to positive real half of the complex plane as the delay increases. Each point of the delay where a pair of eigenvalues reaches the imaginary axis results in a supercritical Hopf bifurcation, and is denoted by the critical delay Δ_{cr}. Figure 2.4 displays the curves along which the Hopf bifurcations occur, as a function of the arrival rate λ. For any λ, the queues lose stability when the delay exceeds the first Hopf curve, at which point a stable limit cycle is established. We will now approximate the amplitude of the limit cycle near the bifurcation point via Lindstedt's method.

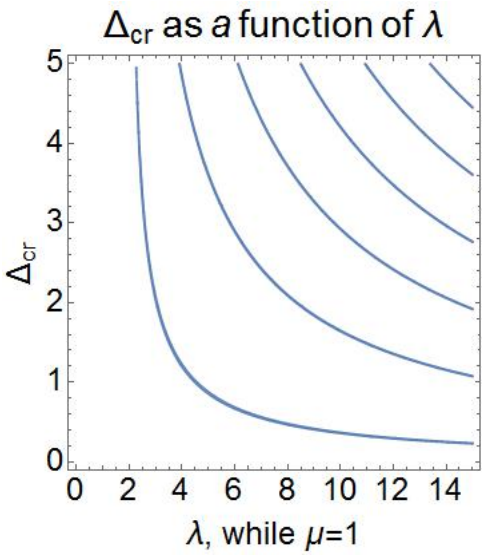

Figure 2.4: The Hopf curves for $\mu = 1$.

2.2. Main Steps of Lindstedt's Method. Lindstedt's method was originally formulated for finite-dimensional differential equations, but has been later extended to delay differential equations. Texts such as [8] and [27] apply Lindstedt's method for equations with delays. We synthesize the main steps into four essential parts. These steps provide clarity to the reader who might be unfamiliar with asymptotic techniques and outline a complete methodology for replicating our results for other types of models.

1. The third order Taylor expansions of the DDE's (2.2) - (2.3) can be uncoupled, yielding $\tilde{v}_2$ from Equation (5.14) as our function of interest. The details are provided in the Appendix,

subsection 5.1.3 - subsection 5.1.4. We stretch the time t and scale the function $\tilde{v}_2$:

$$\tau = \omega t, \quad \tilde{v}_2(t) = \sqrt{\epsilon} v(t). \tag{2.24}$$

2. We approximate the unknown function $v(t)$, the delay Δ, and the oscillation frequency ω by performing asymptotic expansions in ϵ:

$$v(t) = v_0(t) + \epsilon v_1(t) + ..., \quad \Delta = \Delta_0 + \epsilon\Delta_1 + ..., \quad \omega = \omega_0 + \epsilon\omega_1 + ... \tag{2.25}$$

3. After the expansions from (2.25) are made, the resulting equation can be separated by the terms with like powers of ϵ (ϵ^0 and ϵ^1). The resulting equations are

$$\mu v_0(\tau) + \frac{\lambda}{2} v_0(\tau - \Delta_0\omega_0) + \omega_0 v_0'(\tau) = 0, \tag{2.26}$$

$$\begin{aligned}\mu v_1(\tau) + \frac{\lambda}{2} v_1(\tau - \Delta_0\omega_0) + \omega_0 v_1'(\tau) \\ + \omega_1 v_0'(\tau) - \frac{1}{24}\lambda v_0^3(\tau - \Delta_0\omega_0) - \frac{1}{2}\lambda(\Delta_1\omega_0 + \Delta_0\omega_1) v_0'(\tau - \Delta_0\omega_0) = 0.\end{aligned} \tag{2.27}$$

Equation (2.26) is satisfied by the solution $v_0(\tau) = A_v \sin(\tau)$, which is expected since v_0 describes the queue behavior at the Hopf bifurcation where a limit cycle is born. It can be verified by substitution of $\Delta_0 = \Delta_{cr}$ and $\omega_0 = \omega_{cr}$. Further, the equation for $v_1(\tau)$ has a homogeneous and a non-homogeneous parts to it. The homogeneous part $v_1^H(\tau)$ satisfies an equation which is identical to the Equation (2.26), so any linear combination of $\sin(\tau)$ and $\cos(\tau)$ will satisfy the equation for $v_1^H(\tau)$. To avoid secular terms in the non-homogeneous solution, the coefficients of $\sin(\tau)$ and $\cos(\tau)$ resulting from v_0 in Equation (2.27) must vanish. This gives two equations with two unknowns, A_v and ω_1.

4. The resulting equations can be solved for A_v and ω_1. Substituting in $\Delta_0 = \Delta_{cr}$ and $\omega_0 = \omega_{cr}$, the results are

$$\omega_1 = -\frac{(\Delta - \Delta_{cr})\lambda^2(\lambda^2 - 4\mu^2)^{3/2}}{4\left(2\lambda^2\mu - 8\mu^3 + \lambda^2\sqrt{\lambda^2 - 4\mu^2}\arccos(-\frac{2\mu}{\lambda})\right)}, \tag{2.28}$$

$$A_v(\Delta) = \sqrt{\Delta - \Delta_{cr}} \cdot \sqrt{\frac{8(\lambda^2 - 4\mu^2)^2}{2\lambda^2\mu - 8\mu^3 + \lambda^2\sqrt{\lambda^2 - 4\mu^2}\arccos(-\frac{2\mu}{\lambda})}}. \tag{2.29}$$

Amplitude of the Queues. The function $\tilde{v}_2$ from Equation (5.14) attains a steady state amplitude approximately given by A_v. A change of variables reveals the amplitude of q_1 and q_2, showing that the steady state of queues up to a phase shift is given by

$$q_1(t) \to \frac{\lambda}{2\mu} + \frac{1}{2}A_v \sin(\omega t), \qquad q_2(t) \to \frac{\lambda}{2\mu} - \frac{1}{2}A_v \sin(\omega t), \tag{2.30}$$

where ω is the frequency of oscillations and the amplitude is $\frac{1}{2}A_v$.

2.3. Numerical Results of Lindstedt's Method. Although the Figure 2.5 - Figure 2.6 demonstrate that the amplitude approximation from Equation (2.30) matches the behavior of the queues quite well, they do not reveal whether the approximation remains equally accurate when the model parameters vary. Hence, in this section, we wish to know under what conditions the approximation of the steady state amplitude is accurate. We consider the queue lengths to be determined with

sufficient accuracy by numerical integration of Equations (2.2) - (2.3) using MATLAB's 'dde23' function, and will use numerical integration to assess the validity of the approximation.

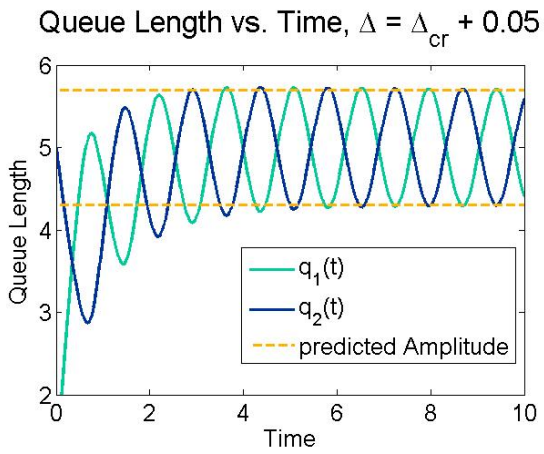

Figure 2.5: $\lambda = 10$, $\mu = 1$.

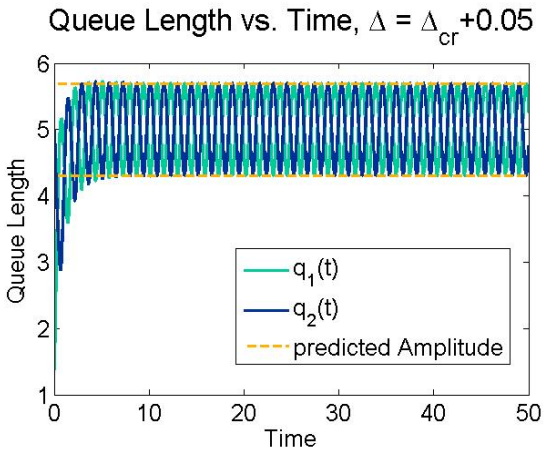

Figure 2.6: $\lambda = 10$, $\mu = 1$.

Lindstedt's method perturbs the system about Δ_{cr}, so the approximated amplitude must approach the true amplitude as $\Delta \to \Delta_{cr}$. This is consistent with our numerical results, and is evident from Figure 2.7 - Figure 2.8. The two plots compare the numerically found amplitude with Lindstedt's amplitude while treating each as a function of delay for parameters $(\lambda, \mu) = (10, 1)$ for the ranges $\Delta \in [\Delta_{cr}, \Delta_{cr} + 0.2]$ and $\Delta \in [\Delta_{cr}, \Delta_{cr} + 1]$, respectively. In both cases the approximation is highly accurate when $\tau = \Delta - \Delta_{cr} \to 0$. However, Lindstedt's method cannot provide theoretical guarantees as the gap between Δ and Δ_{cr} increases, and as seen from Figure 2.7 - Figure 2.8 the approximation loses accuracy.

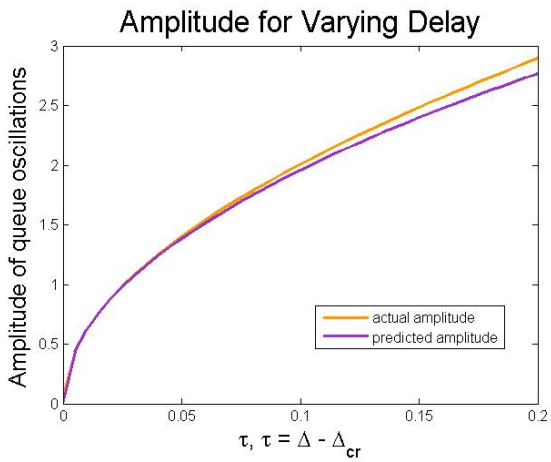

Figure 2.7: $\lambda = 10$, $\mu = 1$.

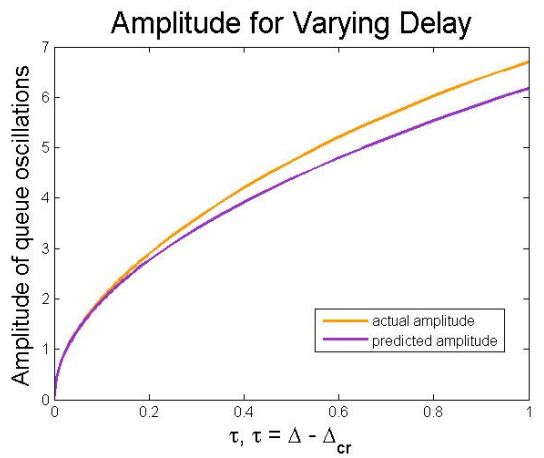

Figure 2.8: $\lambda = 10$, $\mu = 1$.

The method's performance is also affected by the choice of parameters λ and μ. Lindstedt's method works better for smaller λ, as shown by the surface plot Figure 2.9 of the absolute error of Lindstedt's approximation across a range of λ and Δ. Based on the plot, the error of approximation monotonically increases with respect to both λ and Δ. While the absolute error in Figure 2.9 is

constructed for $\mu = 1$, the same holds for other choices of μ. The performance of Lindstedt's method depends on μ in a similar fashion. The accuracy of the method improves when μ increases, and the error is monotone with respect to both μ and Δ. This trend is exemplified by the surface plot in Figure 2.10, which shows the absolute error of Lindstedt's method as a function of μ and Δ, for $\lambda = 10$.

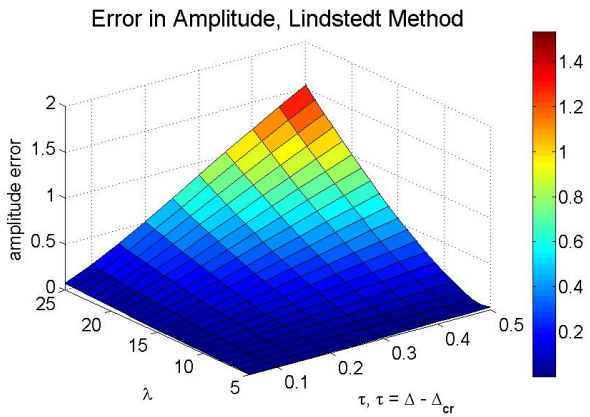

Figure 2.9: Absolute error, varying λ.

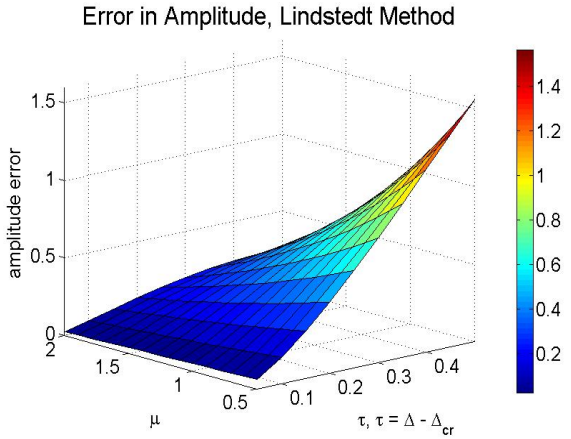

Figure 2.10: Absolute error, varying μ.

The observation that Lindstedt's method works differently for varying values of λ, μ, and Δ leads to two points. The first point is that even though the parameters depend on the physical circumstances and cannot be easily manipulated, it is beneficial to know when to expect a larger error in approximation. The second point is that the limitations of Lindstedt's method motivate us to develop a different numerical technique with the objective of decreasing the maximum error over a larger set of parameter values. Specifically, we would like to eliminate the peaks of error observed in Figure 2.9 - Figure 2.10 when λ is large or μ is small, and therefore obtain a more accurate approximation of the amplitude. With this in mind, we introduce the slope function method.

2.4. The Slope Function Method. The theory of Hopf bifurcation together with numerical examples highlight that the amplitude is approximately proportional to the square root of the difference of the actual delay and the critical delay, i.e.

$$\text{Amplitude} \approx C(\lambda, \mu) \cdot \sqrt{\Delta - \Delta_{cr}}, \tag{2.31}$$

where the $C(\lambda, \mu)$ does not depend on Δ. We call $C(\lambda, \mu)$ the **slope function** as it characterizes the slope of the amplitude as a function of system's parameters. In this section, we propose a statistical way to fit the slope function, which turns out to approximate the amplitude in some cases better than Lindstedt's method.

The Slope Function Algorithm.

1. For a fixed pair of parameters λ_1 and μ_1, we find the amplitude $A(\tau)$ via numerical integration for a finite number of points $\tau = \Delta - \Delta_{cr} := 0, d, 2d, ..., (K-1)d$, where $d > 0$ and $K \in \mathbb{N}$. Then $C(\lambda_1, \mu_1)$ is defined to be such coefficient C that for $A_p(\tau) = C\sqrt{\tau}$, the error $A_p(\tau) - A(\tau)$ is minimized in the least squares sense.
   The sum of squared errors for the K points of delay is given by the function $F(c) =$

$\sum_{j=0}^{K-1}\Big(c\sqrt{jd} - A(jd)\Big)^2$, which by definition reaches its minimum at C. Hence

$$\frac{dF(C)}{dc} = \sum_{j=0}^{K-1} 2\sqrt{jd}\Big(C\sqrt{jd} - A(jd)\Big) = 0. \tag{2.32}$$

The closed-form solution for C is found to be

$$C = \frac{\sum_{j=0}^{K-1}\sqrt{jd}A(jd)}{\sum_{j=0}^{K-1} jd}. \tag{2.33}$$

This gives us the value of the slope function at (λ_1, μ_1). To see how this approximation compares to the Lindstedt's method, consider Figure 2.11 and Figure 2.12, which show the amplitude as a function of delay for $\lambda = 10$ and $\lambda = 20$, respectively. The slope function offers a relatively good approximation for the fixed λ and μ, and it is left to determine the function for the other values of λ and μ.

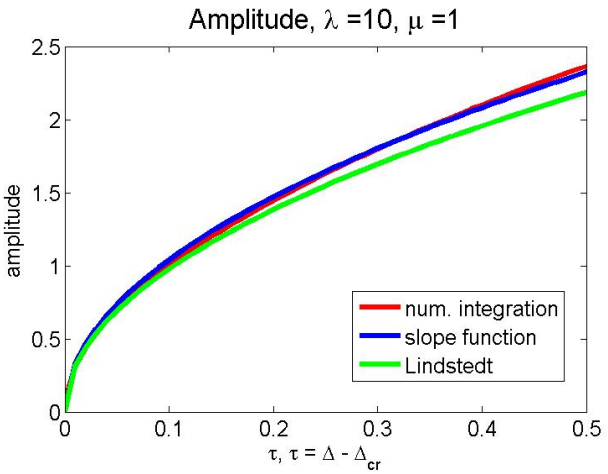

Figure 2.11: Approximation comparison.

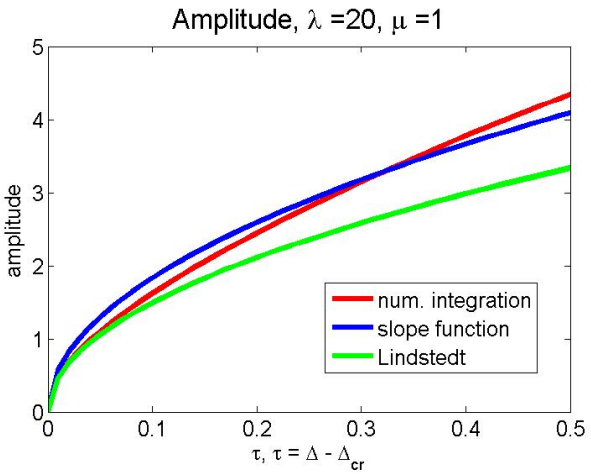

Figure 2.12: Approximation comparison.

2. We extrapolate to find the slope function at arbitrary λ and μ based on the function's values computed for a few points. We assume that $C(\lambda, \mu)$ is a separable function,

$$C(\lambda, \mu) = \Lambda(\lambda)M(\mu), \tag{2.34}$$

and then approximate the functions Λ and M by first degree polynomials

$$\Lambda(\lambda) \approx l_0 + l_1\lambda, \qquad M(\mu) \approx m_0 + m_1\mu, \qquad l_0, l_1, m_0, m_1 \in \mathbb{R}. \tag{2.35}$$

We cannot prove that C is a separable function because it depends on the unknown function A, the "true" amplitude, which is not necessarily separable. However, the separability assumption is a reasonable approximation based on numerical insight. Further, $C(\lambda, \mu)$ from Equation (2.33), as seen from experimental data, indeed is very close to a linear function of λ when μ is constant, and it is close to linear as a function of μ while λ is constant. This approximately linear behavior with respect to λ and μ is demonstrated in Figure 2.13

- Figure 2.14, respectively, where the blue line in each plot represents the values of $C(\lambda, \mu)$ computed according to Equation (2.33).

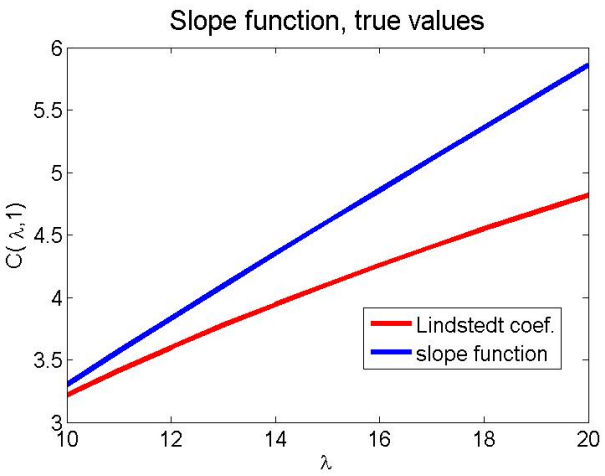

Figure 2.13: C is approximately linear in λ.

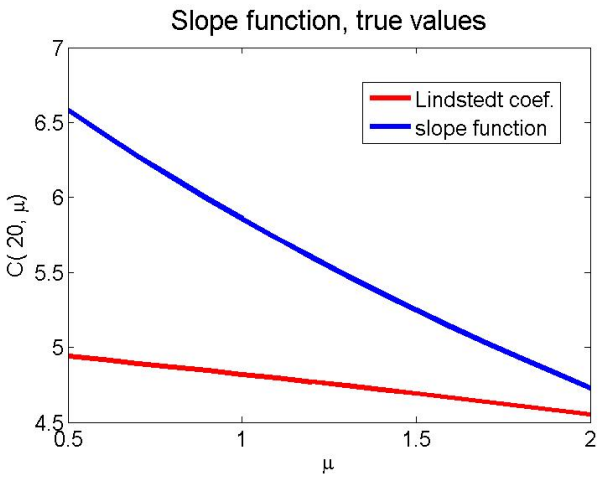

Figure 2.14: C is approximately linear in μ.

3. We reduce the number of coefficients by a change of variables $a_1 = l_1 m_1$, $l_0 = a_2 l_1$, and $m_0 = a_3 m_1$. Equation (2.34) then becomes

$$C(\lambda, \mu) = a_1(a_2 + \lambda)(a_3 + \mu). \tag{2.36}$$

Determining three unknown coefficients requires three data points $C(\lambda_1, \mu_1)$, $C(\lambda_2, \mu_1)$, and $C(\lambda_1, \mu_2)$ that are evaluated based on Equation (2.33) from Step 1 of the algorithm. Then Equation (2.36) allows us to solve for a_1, a_2, and a_3:

$$\frac{C(\lambda_1, \mu_1)}{C(\lambda_2, \mu_1)} = \frac{a_2 + \lambda_1}{a_2 + \lambda_2}, \quad \frac{C(\lambda_1, \mu_1)}{C(\lambda_1, \mu_2)} = \frac{a_3 + \mu_1}{a_3 + \mu_2}, \quad a_1 = \frac{C(\lambda_1, \mu_2)}{(a_2 + \lambda_1)(a_3 + \mu_2)}. \tag{2.37}$$

Therefore the coefficients of interest are

$$a_1 = \frac{C(\lambda_1, \mu_2)}{(a_2 + \lambda_1)(a_3 + \mu_2)}, \quad a_2 = \frac{\lambda_1 - x_1\lambda_2}{x_1 - 1}, \quad a_3 = \frac{\mu_1 - x_2\mu_2}{x_2 - 1}, \tag{2.38}$$

$$\text{where} \quad x_2 = \frac{C(\lambda_1, \mu_1)}{C(\lambda_1, \mu_2)}, \quad x_1 = \frac{C(\lambda_1, \mu_1)}{C(\lambda_2, \mu_1)}. \tag{2.39}$$

Remark. By this algorithm, the amplitude of the queues is estimated to be

$$\text{Amplitude} \approx a_1(a_2 + \lambda)(a_3 + \mu)\sqrt{\Delta - \Delta_{cr}}, \tag{2.40}$$

where the coefficients a_1, a_2, and a_3 are given by Equations (2.38) - (2.39). The specific values of these coefficients will slightly vary depending on the choice of parameters λ_1, λ_2, μ_1, and μ_2 because the linearity assumption of Equations (2.35) is only an approximation of the true behavior as shown in Figure 2.13 - Figure 2.14. Hence, for optimal results one should choose the data points $C(\lambda_1, \mu_1)$, $C(\lambda_2, \mu_1)$, and $C(\lambda_1, \mu_2)$ around the range of λ and μ that one is interested in.

2.5. Numerical Results for the Slope Function Method. We will now numerically compare the performance of the slope function method to Lindstedt's method. Figure 2.15 and Figure 2.16 show the absolute error of the amplitude for varying λ and Δ resulting from the slope function and Lindstedt's method, respectively. Note that overall the slope function results in a smaller error for a wide range of λ and Δ, with a maximum error of 0.4 compared with a maximum error of 1.5 in Lindstedt's approximation. However, unlike Lindstedt's technique the slope function does not guarantee to be accurate when Δ approaches Δ_{cr}. Thus, it is advantageous to use the slope function for predicting the amplitude when the delay is sufficiently greater than the critical value, while Lindstedt's method is preferable when the delay is close to the threshold. A similar observation holds in the case when λ is constant and μ varies. Surface plots Figure 2.17 and Figure 2.18 show that the slope function has a maximum error of less than a third of the error seen in Lindstedt's method, being outperformed mainly when the delay approaches the critical value.

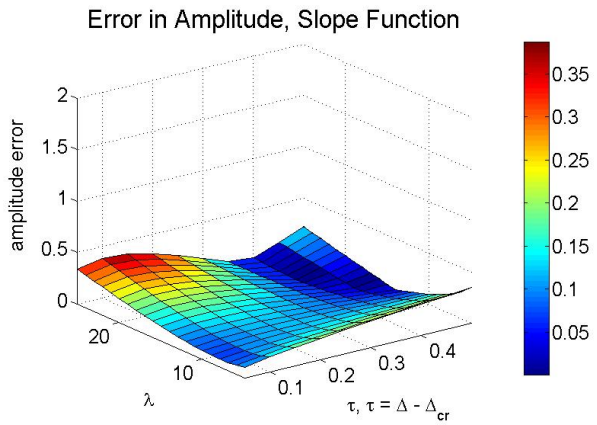

Figure 2.15: Absolute error from the slope function, with $\mu = 1$.

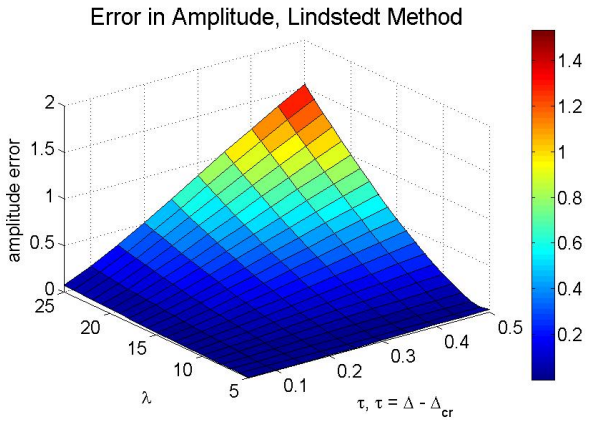

Figure 2.16: Absolute error from Lindtsedt's method, with $\mu = 1$.

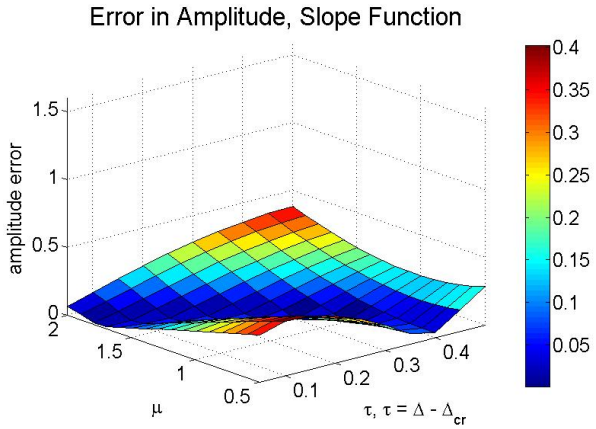

Figure 2.17: Absolute error from the slope function, with $\lambda = 20$.

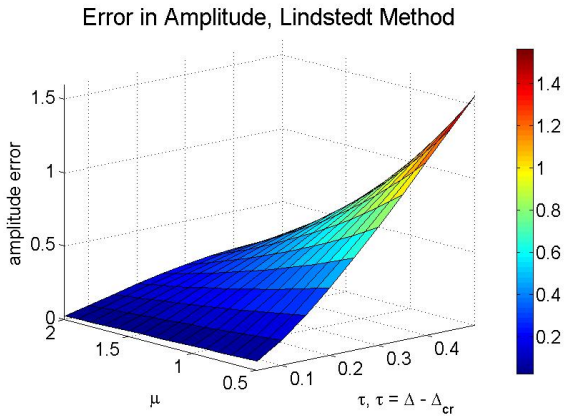

Figure 2.18: Absolute error from Lindtsedt's method, with $\lambda = 20$.

In conclusion to this analysis, we wish to emphasize that neither numerical method comes with

an analytic expression for an error bound. Therefore a comparison of numeric results provides a valuable insight and gives intuition about the performance of the two methods. However, our insight from numerics is of course limited because we do not have guarantees that the numerical trends observed for this one queueing model can be extended to other models. In the next section we will introduce a different queueing model not only to study the model's behavior (which is interesting in of itself given the model's relevance to applications), but also to verify that the numerical trends of the method performance are consistent with the trends we observed so far.

3. Moving Average Fluid Model. In this section, we present a queueing model similar to the constant delay model from section 2, except here, the information given to the customer is the average length of each queue measured over the last Δ time units, or the *moving average*. Figure 2.1 still accurately represents the overall system: the customers appear at a rate λ, join one of the two queues with probabilities p_1 and p_2, and get service at a rate μ with an infinite number of servers. Customers join the queues according to the Multinomial Logit Model, giving higher preference to the queue with a smaller average length

$$p_1 = \frac{\exp\Big(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_1(s)ds\Big)}{\exp\Big(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_1(s)ds\Big) + \exp\Big(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_2(s)ds\Big)} \tag{3.1}$$

$$p_2 = \frac{\exp\Big(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_2(s)ds\Big)}{\exp\Big(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_1(s)ds\Big) + \exp\Big(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_2(s)ds\Big)}. \tag{3.2}$$

Here p_i is the probability of i^{th} queue being joined, $q_i(t)$ is the i^{th} queue length, and the integral expressions in the exponents are the moving average lengths of the queues.

Given these probabilities we can describe the queue lengths as

$$\dot{q}_1 = \lambda \cdot \frac{\exp\Big(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_1(s)ds\Big)}{\exp\Big(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_1(s)ds\Big) + \exp\Big(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_2(s)ds\Big)} - \mu q_1(t) \tag{3.3}$$

$$\dot{q}_2 = \lambda \cdot \frac{\exp\Big(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_2(s)ds\Big)}{\exp\Big(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_1(s)ds\Big) + \exp\Big(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_2(s)ds\Big)} - \mu q_2(t), \tag{3.4}$$

where $\Delta, \lambda, \mu > 0$. The equations are simplified by the notation for the moving average m_i, which itself satisfies a delay differential equation:

$$m_i(t,\Delta) = \frac{1}{\Delta}\int_{t-\Delta}^{t} q_i(s)ds, \tag{3.5}$$

$$\dot{m}_i(t,\Delta) = \frac{1}{\Delta}\cdot\big(q_i(t) - q_i(t-\Delta)\big), \quad i \in \{1,2\}. \tag{3.6}$$

The functional differential equations (3.3) - (3.4) can now be expressed as a system of DDE's

$$\dot{q}_1 = \lambda \cdot \frac{\exp(-m_1(t))}{\exp(-m_1(t)) + \exp(-m_2(t))} - \mu q_1(t) \tag{3.7}$$

$$\dot{q}_2 = \lambda \cdot \frac{\exp(-m_2(t))}{\exp(-m_1(t)) + \exp(-m_2(t))} - \mu q_2(t) \tag{3.8}$$

$$\dot{m}_1 = \frac{1}{\Delta}\cdot(q_1(t) - q_1(t-\Delta)) \tag{3.9}$$

$$\dot{m}_2 = \frac{1}{\Delta}\cdot(q_2(t) - q_2(t-\Delta)). \tag{3.10}$$

Since the functions m_i represent the averages of q_i, the initial conditions of m_i must reflect this. With $f_1(t)$ and $f_2(t)$ being continuous and nonnegative functions on $t \in [-\Delta, 0]$, the initial conditions are

$$q_1(t) = f_1(t), \quad q_2(t) = f_2(t), \quad t \in [-\Delta, 0]; \tag{3.11}$$

$$m_1(0) = \frac{1}{\Delta}\int_{-\Delta}^{0} f_1(s)ds, \quad m_2(0) = \frac{1}{\Delta}\int_{-\Delta}^{0} f_2(s)ds. \tag{3.12}$$

3.1. Hopf Bifurcation in the Moving Average Model. The behavior of the queues (3.7) - (3.10) depends on the delay parameter Δ, but the dependence itself is more nuanced than in the Constant Delay model. To provide a qualitative understanding of the behavior, we will begin by establishing the existence and uniqueness of the equilibrium.

Theorem 3.1. *The unique equilibrium of Equations* (3.7) - (3.10) *is given by*

$$q_1^*(t) = q_2^*(t) = m_1^*(t) = m_2^*(t) = \frac{\lambda}{2\mu}. \tag{3.13}$$

Proof. See the proof in the Appendix. ∎

The stability of the equilibrium comes from the eigenvalues of the characteristic equation that is determined by the linearized system of equation. In subsection 5.2.3 and subsection 5.2.4 in the Appendix, we linearize the system of Equations (3.7) - (3.10) and separate the variables, reducing the system from four unknown functions to two:

$$\dot{\tilde{v}}_2(t) = -\frac{\lambda}{2}\tilde{v}_4(t) - \mu\tilde{v}_2(t) \tag{3.14}$$

$$\dot{\tilde{v}}_4(t) = \frac{1}{\Delta}\Big(\tilde{v}_2(t) - \tilde{v}_2(t-\Delta)\Big). \tag{3.15}$$

To determine the characteristic equation, we need to first consider a special scenario with the trivial eigenvalue. Under the assumption that $\tilde{v}_2 = e^{\Lambda t}$ with $\Lambda = 0$, both functions must be constant, so for some $c_2, c_4 \in \mathbb{R}$, $\tilde{v}_2(t) = c_2$, $\tilde{v}_4(t) = c_4$. By Equation (3.15), the initial condition for $\tilde{v}_2(t)$ must be a constant function on $t \in [-\Delta, 0]$ so $\tilde{v}_2(t) = c_2$ for all $t \geq -\Delta$. The initial condition for $\tilde{v}_4$ then implies that $\tilde{v}_4(0) = c_4 = \int_{-\Delta}^{0} c_2 ds = \Delta c_2$. Therefore $c_4 = \Delta c_2$, but from Equation (3.14) we also find that $c_2 = -\frac{\lambda c_4}{2\mu}$. The only way both equalities can hold is if $c_2 = c_4 = 0$. Thus the trivial eigenvalue can only exist as a solution when the initial conditions are exactly zero, meaning that both queues must be of equal length $q_1(t) = q_2(t) = \frac{\lambda}{2\mu}$ for all $t \in [-\Delta, 0]$.

Now we determine the characteristic equation assuming that $\tilde{v}_2 = e^{\Lambda t}$, $\Lambda \neq 0$:

$$\Phi(\Lambda, \Delta) = \Lambda + \mu + \frac{\lambda}{2\Delta\Lambda} - \frac{\lambda}{2\Delta\Lambda} \cdot e^{-\Lambda\Delta} = 0. \tag{3.16}$$

The equilibrium is stable as long as all eigenvalues Λ have negative real parts. Proposition 5.2 in the Appendix shows that any real eigenvalue must be negative. However, since $\Delta > 0$ there are also infinitely many pairs of complex eigenvalues. The following proposition shows that, regardless of the parameters λ and μ, all complex eigenvalues have negative real parts when the delay is sufficiently small.

Proposition 3.2. *Let $\lambda, \mu, \Delta > 0$. There exists $\Delta^* > 0$ such that for any $\Delta < \Delta^*$, all complex eigenvalues of the characteristic equation* (3.16) *have negative real parts.*

Proof. Let $\Lambda = a + ib$ be a solution of Equation (3.16). Then a and b must satisfy

$$\cos(b\Delta) = \frac{e^{a\Delta}}{\lambda}(2a^2\Delta - 2b^2\Delta + \lambda + 2a\mu\Delta) \tag{3.17}$$

$$\sin(b\Delta) = -\frac{e^{a\Delta}}{\lambda} \cdot 2b\Delta(2a + \mu). \tag{3.18}$$

If b satisfies these equations, then $-b$ is a solution too. Hence without loss of generality we will assume that $b > 0$. Summing the squares of the two equations, we get

$$e^{-2a\Delta}\lambda^2 = \left(2a^2\Delta - 2b^2\Delta + \lambda + 2a\mu\Delta\right)^2 + \left(2b\Delta(2a + \mu)\right)^2, \tag{3.19}$$

from which b can be expressed as a continuous function of a and Δ, namely $b(a, \Delta)$. If $a = 0$ then $b(0, \Delta) = \sqrt{\frac{\lambda}{\Delta} - \mu^2}$, and when plugged into Equation (3.18) we get

$$\sin(b(0, \Delta)\Delta) = -\frac{2\mu}{\lambda} \cdot b(0, \Delta)\Delta \tag{3.20}$$

$$\sin(x(0, \Delta)) = -\frac{2\mu}{\lambda} \cdot x(0, \Delta) \tag{3.21}$$

$$x(a, \Delta) = b(a, \Delta)\Delta, \quad x(0, \Delta) = \Delta\sqrt{\frac{\lambda}{\Delta} - \mu^2}. \tag{3.22}$$

The function x will be helpful in the proof. Note that x is a continuous function of b and therefore of a. Let us define $\Delta^* > 0$ as

$$\Delta^* = \begin{cases} \frac{\lambda}{2\mu^2}, & \frac{\lambda}{2\mu} \leq \pi \\ \frac{\lambda - \sqrt{\lambda^2 - 4\mu^2\pi^2}}{2\mu^2}, & \text{otherwise.} \end{cases} \tag{3.23}$$

This choice of Δ^* guarantees that for all $\Delta < \Delta^*$, the functions $b(0, \Delta)$ and $x(0, \Delta)$ are real. Further, Δ^* ensures that $0 < x(0, \Delta) < \min(\pi, \frac{\lambda}{2\mu})$ for all $\Delta < \Delta^*$, which can be checked from Equation (3.22). The condition $0 < x(0, \Delta) < \pi$ implies that

$$\sin(x(0, \Delta)) > 0 > -\frac{2\mu}{\lambda} \cdot x(0, \Delta). \tag{3.24}$$

However, for any $a \geq 0$, Equation (3.18) gives the inequality

$$\sin(x(a, \Delta)) = -\frac{e^{a\Delta}}{\lambda} \cdot 2x(a, \Delta)(2a + \mu) \leq -\frac{2\mu}{\lambda} \cdot x(a, \Delta), \tag{3.25}$$

therefore when $a = 0$ the inequality remains

$$\sin(x(0, \Delta)) \leq -\frac{2\mu}{\lambda} \cdot x(0, \Delta), \tag{3.26}$$

which is in contradiction with Equation (3.24). Hence a must be negative to satisfy the characteristic equation for $\Delta < \Delta^*$. ∎

The stability of the equilibrium is lost when a pair of complex eigenvalues crosses the imaginary axis. If for some $\Delta = \Delta_{cr}$ there are purely imaginary eigenvalues, $\Lambda = \pm i\omega_{cr}$, $\omega_{cr} > 0$, then the characteristic equation gives the equalities

$$\sin(\omega_{cr}\Delta_{cr}) = -\frac{2\Delta_{cr}\mu\omega_{cr}}{\lambda}, \quad \cos(\omega_{cr}\Delta_{cr}) = 1 - \frac{2\Delta_{cr}\omega_{cr}^2}{\lambda}. \tag{3.27}$$

From the trigonometric identity $\sin^2(\omega_{cr}\Delta_{cr}) + \cos^2(\omega_{cr}\Delta_{cr}) = 1$, ω_{cr} can be found

$$\omega_{cr} = \sqrt{\frac{\lambda}{\Delta_{cr}} - \mu^2}. \tag{3.28}$$

Since ω_{cr} must be real and nonzero, the condition $\Delta_{cr} < \frac{\lambda}{\mu^2}$ must hold. When ω_{cr} is substituted into Equation (3.27), we find that Δ_{cr} must satisfy the equation

$$\sin\left(\Delta_{cr} \cdot \sqrt{\frac{\lambda}{\Delta_{cr}} - \mu^2}\right) + \frac{2\mu\Delta_{cr}}{\lambda} \cdot \sqrt{\frac{\lambda}{\Delta_{cr}} - \mu^2} = 0. \tag{3.29}$$

We are now ready to formulate the conditions that determine the stability of the equilibrium.

Theorem 3.3. *If the Equation (3.29) has no positive roots Δ_{cr} then the equilibrium of Equations (3.7) - (3.10) is stable for all $\Delta > 0$. If there exists $\Delta_{cr} > 0$ satisfying Equation (3.29) then the equilibrium is stable when Δ is less than the smallest positive root Δ_{cr} or greater than the largest root Δ_{cr}. Further, the largest root Δ_{cr} is less than $\frac{\lambda}{\mu^2}$.*

Proof. See the proof in the Appendix. ∎

If and when Δ exceeds the smallest positive root Δ_{cr} of Equation (3.29), the equilibrium becomes unstable and a stable limit cycle emerges. Figure 3.1 and Figure 3.2 show the transition. The change of behavior is due to a Hopf bifurcation, as shown in the next theorem. Further, since there can be multiple roots Δ_{cr} to Equation (3.29) for fixed parameters λ and μ, multiple Hopf bifurcations may occur.

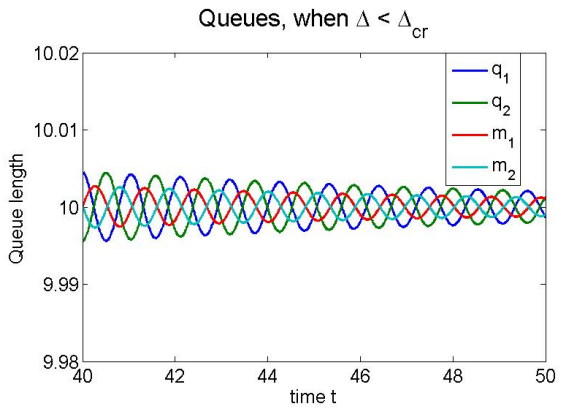

Figure 3.1: Before bifurcation.

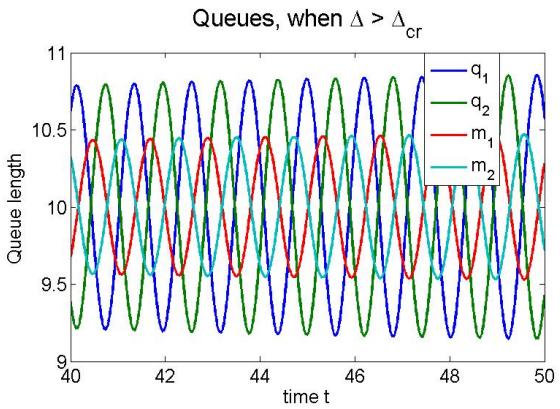

Figure 3.2: After bifurcation.

Theorem 3.4. *If Δ_{cr} satisfies Equation (3.29) and $\Delta_{cr} \neq \frac{\lambda - 2\mu}{2\mu^2}$, then the queues from Equations (3.7) - (3.10) undergo a Hopf bifurcation at Δ_{cr}.*

Proof. For each Δ_{cr} satisfying Equation (3.29), the characteristic equation (3.16) has two simple roots $\Lambda = \pm i\omega_{cr}$. Further, through implicit differentiation of Equation (3.16), it can be shown that $\mathrm{Re}[\Lambda'(\Delta_{cr})] \neq 0$:

$$\mathrm{Re}\ \Lambda'(\Delta_{cr}) = \frac{2\omega_{cr}^2(\lambda - 2\mu - 2\mu^2\Delta_{cr})}{4\omega_{cr}^2\Delta_{cr}(3 + 2\Delta_{cr}\mu) + \lambda(4 + \Delta_{cr}\lambda + 4\Delta_{cr}\mu)}. \tag{3.30}$$

The denominator of $\mathrm{Re}[\Lambda'(\Delta_{cr})]$ is positive, and the assumption $\Delta_{cr} \neq \frac{\lambda - 2\mu}{2\mu^2}$ guarantees the numerator to be nonzero. Further, all other eigenvalues Λ^* are complex with a nonzero real part, so $\Lambda^* \neq m\Lambda$. Therefore, a Hopf bifurcation occurs at Δ_{cr}. ∎

As was suggested by Figure 3.2, the limit cycle is stable. In fact, the following theorem shows that any Hopf bifurcation in our queueing system is supercritical.

Theorem 3.5. *Any Hopf bifurcation from Theorem 3.4 is supercritical.*

Proof. We will use the method of slow flow to determine whether the limit cycle is stable. The third order expansion of Equations (3.7) - (3.8) can be uncoupled, and the resulting equations of interest are given by subsection 5.2.3 and subsection 5.2.4:

$$\dot{\tilde{v}}_2 = \lambda\Big(-\frac{\tilde{v}_4(t)}{2} + \frac{\tilde{v}_4(t)^3}{24}\Big) - \mu\tilde{v}_2(t) \tag{3.31}$$

$$\dot{\tilde{v}}_4 = \frac{1}{\Delta}\Big(\tilde{v}_2(t) - \tilde{v}_2(t-\Delta)\Big). \tag{3.32}$$

The two variables are scaled by $\sqrt{\epsilon}$

$$\tilde{v}_2(t) = \sqrt{\epsilon}v(t), \qquad \tilde{v}_4(t) = \sqrt{\epsilon}u(t), \tag{3.33}$$

the delay and the frequency are expanded close to their critical values, and two time scales are introduced:

$$\Delta = \Delta_{cr} + \epsilon\alpha, \quad \omega = \omega_{cr} + \epsilon\beta, \quad \xi = \omega t, \quad \eta = \epsilon t. \tag{3.34}$$

The functions $v(t)$ and $u(t)$ are also expanded

$$v(\xi,\eta) = v_0(\xi,\eta) + \epsilon v_1(\xi,\eta), \qquad u(\xi,\eta) = u_0(\xi,\eta) + \epsilon u_1(\xi,\eta). \tag{3.35}$$

When the suggested transformations are made to the equations for $\dot{v}(t)$ and $\dot{u}(t)$, we can separate the resulting equations by collecting all the terms with the like orders of ϵ. The equations for the zeroth order terms are satisfied with a solution of the form

$$v_0(\xi,\eta) = A(\eta)\cos(\xi) + B(\eta)\sin(\xi), \tag{3.36}$$

which allows us to find the form of $u_0(\xi,\eta)$:

$$u_0(\xi,\eta) = -\frac{2(A(\eta) + B(\eta)\omega_{cr})}{\lambda}\cos(\xi) - \frac{2(B(\eta) - A(\eta)\omega_{cr})}{\lambda}\sin(\xi). \tag{3.37}$$

The terms involving the first order of ϵ comprise of (i) the differential operator acting on x_1, (ii) the non-resonant terms $\cos(3\xi)$ and $\sin(3\xi)$, and (iii) the resonant terms involving $\cos(\xi)$ and $\sin(\xi)$. For no secular terms, the coefficients of $\cos(\xi)$ and $\sin(\xi)$ must vanish, giving a slow flow on $A(\eta)$ and $B(\eta)$. By introducing the polar coordinates

$$A = R\cos(\Theta), \qquad B = R\sin(\Theta) \tag{3.38}$$

we find equation for the radial component $\frac{d}{d\eta}R(\eta)$

$$\frac{dR}{d\eta} = \frac{R(\lambda - \Delta_{cr}\mu^2)(R^2(\lambda+2\mu) - 4\alpha\lambda(\lambda - 2\mu - 2\Delta_{cr}\mu^2))}{2\Delta_{cr}\lambda(-\Delta_{cr}\lambda^2 + 4\Delta_{cr}\mu^2(3 + 2\Delta_{cr}\mu) - 4\lambda(4 + 3\Delta_{cr}\mu))} \tag{3.39}$$

Assuming $R \geq 0$, the equilibrium points are

$$R_0 = 0, \quad R_1 = \sqrt{\frac{4\alpha\lambda(\lambda - 2\mu - 2\Delta_{cr}\mu^2)}{\lambda + 2\mu}}. \tag{3.40}$$

From Theorem 3.4, $\Delta_{cr} \neq \frac{\lambda-2\mu}{2\mu^2}$, so R_1 and R_0 are always two distinct points. When $\Delta_{cr} < \frac{\lambda-2\mu}{2\mu^2}$ then in order for R_1 to be real, α must be positive. On the other hand, if $\Delta_{cr} > \frac{\lambda-2\mu}{2\mu^2}$ then α must be negative for R_1 to be real. In both cases, the assumption $\frac{\lambda}{\Delta_{cr}} - \mu^2 > 0$ that arose from the frequency ω_{cr} being positive, guarantees that $\frac{dR}{d\eta}$ is positive on the interval $R \in (0, R_1)$ and negative when $R > R_1$. Therefore the Hopf bifurcation is supercritical. ■

To summarize, for any fixed parameters λ and μ the queues converge to a stable equilibrium when the delay is sufficiently small. However, as the delay increases up to $\Delta = \frac{\lambda-2\mu}{2\mu^2}$, finitely many pairs of complex eigenvalues may cross to the positive real side of the imaginary axis of the complex plane. Every pair of eigenvalues reaching the imaginary axis is indicated on Figure 3.3 by a Hopf curve. Note that the dashed orange line $\Delta = \frac{\lambda-2\mu}{2\mu^2}$ from Figure 3.3 passes through the minimum of each Hopf curve, where each minimum represents a pair of eigenvalues that reaches the imaginary axis at $\Delta = \frac{\lambda-2\mu}{2\mu^2}$ and then returns back to the negative real side of complex plane without crossing the imaginary axis.

Once the delay exceeds $\frac{\lambda-2\mu}{2\mu^2}$ and the parameters are in the region to the right of the dashed orange line from Figure 3.3, every pair of eigenvalues with positive real parts will inevitably cross back the imaginary axis in the negative real direction. In fact, all eigenvalues will obtain negative real parts before the delay reaches a. This is guaranteed by the condition $0 \neq \omega_{cr} \in \mathbb{R}$ together with Proposition 5.3 in the Appendix. The condition $\Delta = \frac{\lambda}{\mu^2}$ is indicated on Figure 3.3 by the non-Hopf curve, and it is clear that the Hopf curves cannot cross the non-Hopf curve.

The equilibrium is stable whenever λ is below the Hopf 1 curve from Figure 3.3. To quantitatively describe the behavior of the queues after Hopf 1 curve is crossed, we will approximate the amplitude of the queue oscillations via Lindstedt's method.

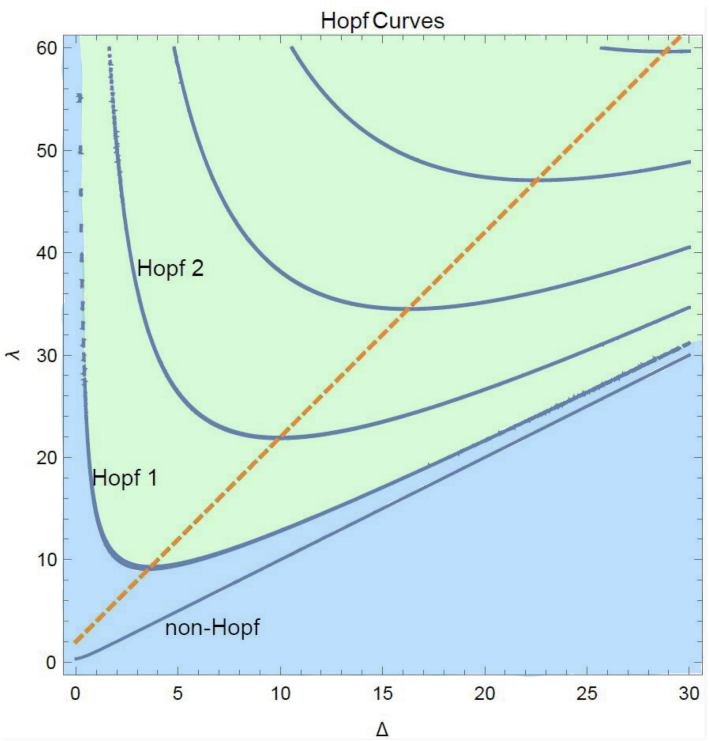

Figure 3.3: The Hopf curves for $\mu = 1$; green area - limit cycles; blue area - stable equilibrium; dashed orange line $\rightarrow \lambda = 2\mu^2\Delta + 2\mu$; the non-Hopf curve $\rightarrow \Delta = \frac{\lambda}{\mu^2}$.

3.2. Lindstedt's Method. We apply Lindstedt's method according to the steps shown in subsection 2.2. However, instead of working with one unknown function we are now working with two.

1. We start with the variables that represent third order polynomial expansion of q_1, q_2, m_1, and m_1 about the equilibrium. These four variables can be reduced to two by a change of variables. The details are provided in the Appendix, subsection 5.2.3 - subsection 5.2.4. The functions of interest become $\tilde{v}_2$ and $\tilde{v}_4$ from Equation (5.43). We stretch the time and scale both functions by $\sqrt{\epsilon}$:

$$\tau = \omega t, \quad \tilde{v}_2 = \sqrt{\epsilon} v(t), \quad \tilde{v}_4 = \sqrt{\epsilon} u(t). \tag{3.41}$$

This ensures that the cubic terms will have one higher order of ϵ than linear terms,

$$\omega \dot{v}(\tau) = \lambda\Big(-\frac{u(\tau)}{2} + \frac{\epsilon u(\tau)^3}{24}\Big) - \mu v(\tau) \tag{3.42}$$

$$\omega \dot{u}(\tau) = \frac{1}{\Delta}\Big(v(\tau) - v(\tau - \omega\Delta)\Big). \tag{3.43}$$

2. We approximate the variables by performing asymptotic expansions in ϵ:

$$v(t) = v_0(t) + \epsilon v_1(t) + ..., \quad u(t) = u_0(t) + \epsilon u_1(t) + ..., \tag{3.44}$$

$$\Delta = \Delta_0 + \epsilon\Delta_1 + ..., \quad \omega = \omega_0 + \epsilon\omega_1 + ... \tag{3.45}$$

3. We separate each of the resulting equations by collecting all the terms of the like powers of ϵ. The terms of order ϵ^0 yield equalities

$$0 = \frac{1}{2}\lambda m_0(\tau) + \mu v_0(\tau) + \omega_0 \dot{v}_0(\tau) \tag{3.46}$$

$$0 = -v_0(\tau) + v_0(\tau - \Delta_0\omega_0) + \Delta_0\omega_0 \dot{m}_0(\tau), \tag{3.47}$$

and the terms of order ϵ^1 yield

$$0 = -\frac{1}{24}\lambda m_0(\tau)^3 + \frac{1}{2}\lambda m_1(\tau) + \mu v_1(\tau) + \omega_1 \dot{v}_0(\tau) + \omega_0 \dot{v}_1(\tau) \tag{3.48}$$

$$0 = \Delta_1\Big(v_0(\tau) - v_0(\tau - \Delta_0\omega_0)\Big) + \Delta_0^2\omega_1 \dot{m}_0(\tau) + \Delta_0^2\omega_0 \dot{m}_1(\tau)$$

$$- \Delta_0\Big(v_1(\tau) - v_1(\tau - \Delta_0\omega_0) + (\Delta_1\omega_0 + \Delta_0\omega_1)\dot{v}_0(\tau - \Delta_0\omega_0)\Big). \tag{3.49}$$

The function m_0 can be expressed through v_0 by Equation (3.46), and m_1 can be expressed through v_0 and v_1 from Equation (3.48). It can be verified that $v_0(\tau) = A_v \sin(\tau)$ satisfies Equations (3.46) - (3.47). Further, the homogeneous part of solution for v_1 is satisfied by $v_1^H(\tau) = a\sin(\tau) + b\cos(\tau)$. Therefore to avoid secular terms $\sin(\tau)$ and $\cos(\tau)$, the coefficients of $\sin(\tau)$ and $\cos(\tau)$ from Equation (3.49) must vanish. This condition gives two equations for two unknowns, w_1 and A_v.

4. After some algebra we determine the amplitude A_v as a function of delay:

$$A_v(\Delta) = \sqrt{\Delta - \Delta_{cr}} \cdot \sqrt{\frac{4\lambda^2(-\lambda - 2\mu + 2\Delta_{cr}\omega_{cr}^2)}{\Delta_{cr}\big(\mu^2 + \omega_{cr}^2\big)\big(-\lambda + 2(\mu + \Delta_{cr}\mu^2 + \Delta_{cr}\omega_{cr}^2)\big)}}. \tag{3.50}$$

Amplitude of the Queues. The function A_v approximates the amplitude of oscillations for $v(t)$ from Equation (3.42). A change of variables reveals the amplitude of q_1 and q_2, showing that the steady state of queues is given up to a phase shift by

$$q_1(t) \to \frac{\lambda}{2\mu} + \frac{1}{2}A_v \sin(\omega t), \qquad q_2(t) \to \frac{\lambda}{2\mu} - \frac{1}{2}A_v \sin(\omega t), \tag{3.51}$$

where the amplitude is $\frac{1}{2}A_v$ and ω is the frequency of oscillations. Figure 3.4 - Figure 3.5 use the predicted amplitude to bound the oscillations of queues near the bifurcation point, providing some validation to Lindstedt's method as well as our calculations.

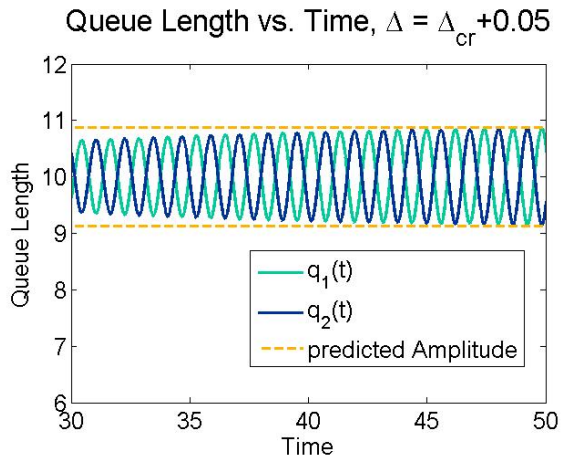

Figure 3.4: $\lambda = 20, \mu = 1$.

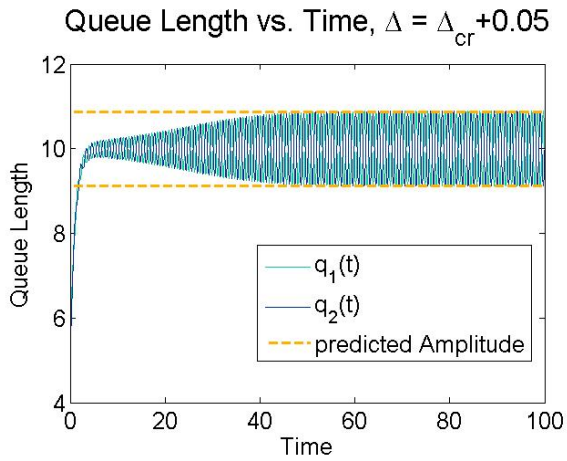

Figure 3.5: $\lambda = 20, \mu = 1$.

3.3. Numerical Results. In this section we compare the approximations of amplitude from Lindstedt's method and the slope function method to the true behavior of the queueing system. Note that the slope function is provided by the algorithm in subsection 2.4 and Equation (2.40), so no additional work is needed. Also, we consider the queue lengths to be determined with sufficient accuracy by numerical integration of Equations (3.7) - (3.10) using MATLAB's 'dde23' function, so we will test our approximations against the numerical integration results.

Our key finding is that *the trends of the method performance are consistent* with those that were observed for the constant delay model in subsection 2.3, both for Lindstedt's method and the slope function method. Hence, we avoid repeating the analysis of subsection 2.3, and instead provide relevant figures with a summary of the key differences between the two methods.

- Lindstedt's method tends to be more accurate than the slope function method when $\Delta \to \Delta_{cr}$. For example, see Figure 3.7, where the amplitude is shown as a function of delay.
- Lindstedt's method loses accuracy when the delay increases, and it is outperformed by the slope function method for larger delay. See Figure 3.6 - Figure 3.11.
- The error of Lindstedt's approximation is monotonic in λ, Δ, and μ. Hence, over the parameter space the error function has predictable and significant peaks around large λ and Δ and around small μ. See Figure 3.9 and Figure 3.11.
- The error of the slope function method is relatively evenly distributed over the parameter space, and therefore there are no significant peaks in error. See Figure 3.8 and Figure 3.10.
- The maximum error for slope function over a neighborhood of parameters is 3 - 4 times smaller than it is for Lindstedt's method. Specifically, the maximum error is three times

smaller for the constant delay model, and 4 times smaller for the moving average model. See Figure 3.8 - Figure 3.11.

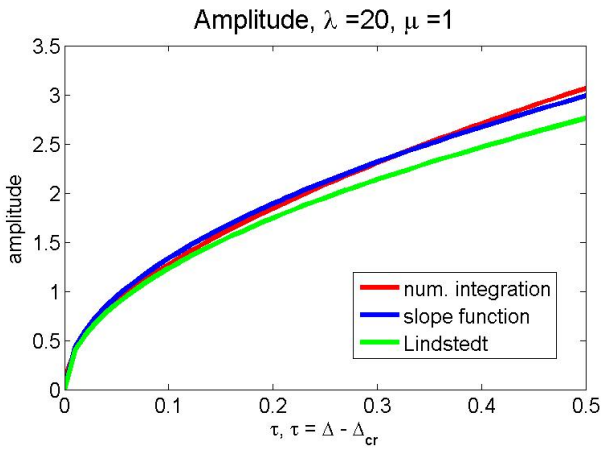

Figure 3.6: Comparison of approximations.

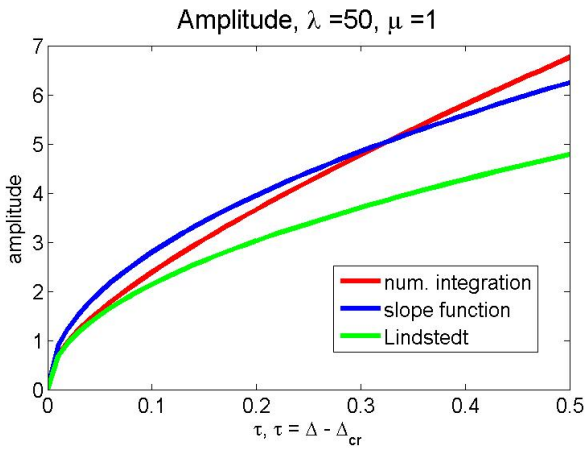

Figure 3.7: Comparison of approximations.

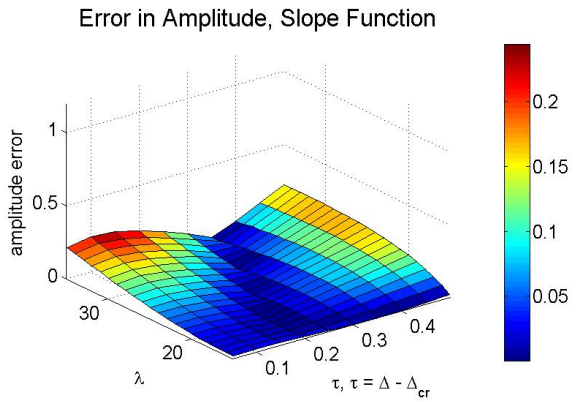

Figure 3.8: Absolute error, $\mu = 1$.

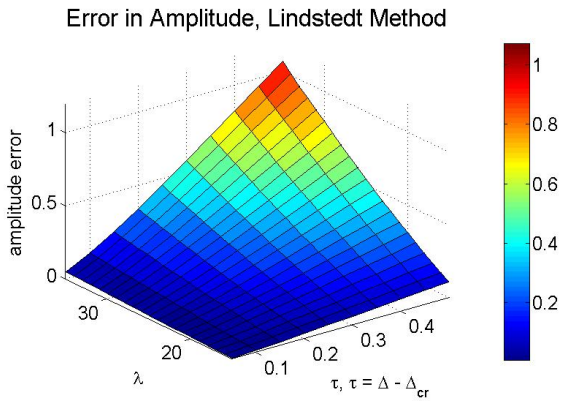

Figure 3.9: Absolute error, $\mu = 1$.

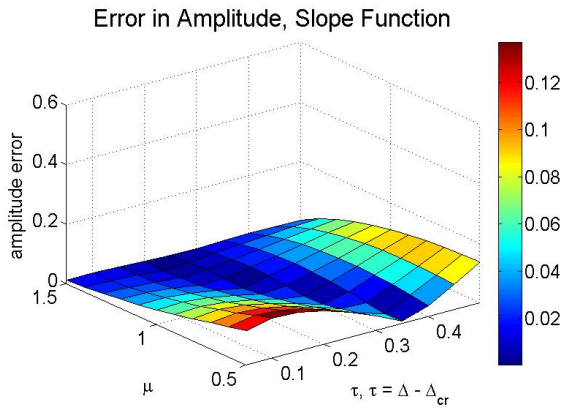

Figure 3.10: Absolute error, $\lambda = 20$.

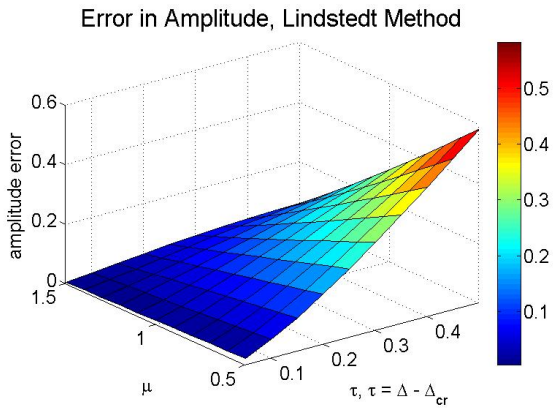

Figure 3.11: Absolute error, $\lambda = 20$.

4. Conclusion. In this paper, we analyze two queueing models that incorporate customer choice and delayed queue length information. The first model assumes a constant delay and while the second one uses a moving average. We analyze the qualitative behavior of these queueing models and show the occurrence of supercritical Hopf bifurcations. Using Lindstedt's method, we construct an analytic approximation for the amplitude of oscillations that the queueing system exhibits after a Hopf bifurcation. Lindstedt's method works well where the delay is close the critical delay value, but the method becomes less accurate for larger values of delay. We address this by proposing a new numerical technique, the slope function method, that estimates the slope of the amplitude as a function of the system's parameters.

The slope function method is conceptually intuitive and elementary in implementation. It can be used in a wide variety of models where a Hopf bifurcation is observed. Unlike the perturbations method, the slope function does not require complicated analytical work and can be implemented without a substantial mathematical background. Limit cycles are known to occur in models studied by social scientists and biologists, for which the slope function method can provide an easy way to numerically approximate the amplitude of oscillations. Although we give no theoretical guarantees on the method's performance, our paper demonstrates on two different models that the slope function method maintains a low error across a much wider range of the parameters than does Lindstedt's method. For our models, the maximum error in approximation is $3-4$ times smaller over a large neighborhood of parameters than the maximum error from Lindstedt's method.

Lastly, it is worthy to note that this paper connects the field of queueing theory to nonlinear dynamics and, in particular, delay-differential equations. Our work opens doors for many other queueing models to be considered with mathematical techniques that may be new to the queueing community. Simultaneously, our work places queueing theory on the radar of the dynamical systems experts as a potential application area for their research with direct relevance to industry.

References.

[1] G. Allon and A. Bassamboo, *The impact of delaying the delay announcements*, Operations research, 59 (2011), pp. 1198–1210.

[2] G. Allon, A. Bassamboo, and I. Gurvich, *We will be right with you: Managing customer expectations with vague promises and cheap talk*, Operations Research, 59 (2011), pp. 1382–1394.

[3] M. Armony and C. Maglaras, *On customer contact centers with a call-back option: Customer decisions, routing rules, and system design*, Operations Research, 52 (2004), pp. 271–292.

[4] M. Armony, N. Shimkin, and W. Whitt, *The impact of delay announcements in many-server queues with abandonment*, Operations Research, 57 (2009), pp. 66–81.
[5] M. Belhaq and S. M. Sah, *Fast parametrically excited van der pol oscillator with time delay state feedback*, International Journal of Non-Linear Mechanics, 43 (2008), pp. 124–130.
[6] S. N. Chow and H. O. Walther, *Characteristic multipliers and stability of symmetric periodic solutions of* $\dot{x}(t) = g(x(t-1))$, Transactions of the American Mathematical Society, 307 (1988), pp. 127–142.
[7] S. L. Das and A. Chatterjee, *Multiple scales without center manifold reductions for delay differential equations near hopf bifurcations*, Nonlinear Dynamics, 30 (2002), pp. 323–335.
[8] T. Erneux, *Applied Delay Differential Equations*, Springer Science, 2009.
[9] B. H. Fralix and I. J. B. F. Adan, *An infinite-server queue influenced by a semi-markovian environment*, Queueing Systems, 61 (2009), pp. 65–84.
[10] D. Freund, S. Henderson, and D. Shmoys, *Minimizing multimodular functions and allocating capacity in bike-sharing systems*, arXiv preprint arXiv:1611.09304, (2016).
[11] P. Guo and P. Zipkin, *Analysis and comparison of queues with different levels of delay information*, Management Science, 53 (2007), pp. 962–970.
[12] P. Guo and P. Zipkin, *The impacts of customers' delay-risk sensitivities on a queue with balking*, Probability in the Engineering and Informational Sciences, 23 (2009), pp. 409–432.
[13] J. Hale and V. Lunel, *Introduction to Functional Differential Equations*, Springer Science, 1993.
[14] B. D. Hassard, N. D. Kazarinoff, and Y. H. Wan, *Theory and Applications of Hopf Bifurcation*, Cambridge University Press, 1981, https://books.google.com/books?id=3wU4AAAAIAAJ.
[15] R. Hassin, *Information and uncertainty in a queuing system*, Probability in the Engineering and Informational Sciences, 21 (2007), pp. 361–380.
[16] J. Hausman and D. McFadden, *Specification tests for the multinomial logit model*, Econometrica, 52 (1984), pp. 1219–1240.
[17] R. Ibrahim, M. Armony, and A. Bassamboo, *Does the past predict the future? the case of delay announcements in service systems*, Management Science, 63 (2017), pp. 1762–1780.
[18] D. L. Iglehart, *Limiting diffusion approximations for the many server queue and the repairman problem*, Journal of Applied Probability, 2 (1965), pp. 429–441.
[19] A. Ivanov, B. Lani-Wayda, and H. O. Walther, *Unstable hyperbolic periodic solutions of differential delay equations*, World Scientific Series in Applicable Analysis, 1 (1992), pp. 301–316.
[20] O. Jouini, Z. Akşin, and Y. Dallery, *Queueing models for full-flexible multi-class call centers with real-time anticipated delays*, International Journal of Production Economics, 120 (2009), pp. 389–399.
[21] O. Jouini, Z. Akşin, and Y. Dallery, *Call centers with delay information: Models and insights*, Manufacturing & Service Operations Management, 13 (2011), pp. 534–548.
[22] L. Lazarus, M. Davidow, and R. Rand, *Periodically forced delay limit cycle oscillator*, International Journal of Non-Linear Mechanics, 94 (2017), pp. 216–222.
[23] D. McFadden, *Modelling the choice of residential location*, Cowles Foundation Discussion Papers 477, Cowles Foundation for Research in Economics, Yale University, 1977, https://EconPapers.repec.org/RePEc:cwl:cwldpp:477.
[24] J. Pender, R. H. Rand, and E. Wesson, *Queues with choice via delay differential equations*, International Journal of Bifurcation and Chaos, 27 (2017).
[25] J. Pender, R. H. Rand, and E. Wesson, *An asymptotic analysis of queues with delayed information and time varying arrival rates*, Nonlinear Dynamics, 91 (2018), pp. 2411–2427.

[26] G. Raina and D. Wischik, *Buffer sizes for large multiplexers: Tcp queueing theory and instability analysis*, Next Generation Internet Networks, 2005, (IEEE, 2005).
[27] R. Rand, *Differential-delay equations*, in Complex Systems: Fractionality, Time-Delay and Synchronization, A.C.J. Luo and J-Q Sun, eds., Springer, 2011, ch. 3, pp. 83–117.
[28] R. Rand, *Lecture Notes on Nonlinear Vibrations*, University Readers, version 53, 2012, http://www.math.cornell.edu/~rand/randdocs/.
[29] S. Resnick and G. Samorodnitsky, *Activity periods of an infinite server queue and performance of certain heavy tailed fluid queues*, Queueing Systems, 33 (1999), pp. 43–71.
[30] A. Skubachevskii and H. O. Walther, *On the floquet multipliers of periodic solutions to non-linear functional differential equations*, Journal of Dynamics and Differential Equations, 18 (2006), pp. 257–355.
[31] H. Smith, *An Introduction to Delay Differential Equations with Applications to the Life Sciences*, Springer Science, 2011.
[32] Y. So and W. Kuhfeld, *Multinomial logit models*, SUGI 20 conference proceedings, (1995).
[33] S. Tao and J. Pender, *A stochastic analysis of bike sharing systems*, arXiv preprint arXiv:1708.08052, (2017).
[34] K. Train, *Discrete Choice Methods with Simulation*, Cambridge University Press, 2009.
[35] W. Whitt, *Improving service by informing customers about anticipated delays*, Management Science, 45 (1999), pp. 192–207.
[36] X. Xie, *Uniqueness and stability of slowly oscillating periodic solutions of delay equations with bounded nonlinearity*, Journal of Dynamics and Differential Equations, 3 (1991), pp. 515–540.
[37] X. Xie, *The multiplier equation and its application to s-solutions of a delay equation*, Journal of Differential Equations, 95 (1992), pp. 259–280.
[38] X. Xie, *Uniqueness and stability of slowly oscillating periodic solutions of delay equations with unbounded nonlinearity*, Journal of Differential Equations, 103 (1993), pp. 350–374.

5. Appendix.

5.1. Constant Delay Model.

5.1.1. Showing the existence and uniqueness of equilibrium.

Proof of Theorem (2.1). When $q_i(t) = q_i(t-\Delta) = \frac{\lambda}{N\mu}$ for each $1 \le i \le N$, all functions q_i are constant with respect to time

$$\dot{q}_i(t) = \lambda \cdot \frac{\exp(-\frac{\lambda}{N\mu})}{\sum_{j=1}^{N} \exp(-\frac{\lambda}{N\mu})} - \mu \frac{\lambda}{N\mu} = 0. \tag{5.1}$$

Therefore $q_i^* = \frac{\lambda}{N\mu}$ is an equilibrium.

To show uniqueness, we will argue by contradiction. Suppose there is another equilibrium given by $\bar{q}_i$, $1 \le i \le N$, and for some i we have $q_i^* \neq \bar{q}_i$. Without loss of generality, let us assume that it is the N'th queue, so $q_N^* \neq \bar{q}_N$. Also, without loss of generality let us assume that $q_N^* > \bar{q}_N$, and since both are constants with respect to time, we can conclude that $\bar{q}_N(t) = \frac{\lambda}{N\mu} + \epsilon$ for some $\epsilon > 0$.

From the condition $0 = \sum_{i=1}^{N} \dot{\bar{q}}_i$, the sum of the queues has to be $\sum_{i=1}^{N} \bar{q}_i = \frac{\lambda}{\mu}$, so the average queue length is $\frac{1}{N}\sum_{i=1}^{N} \bar{q}_i = \frac{\lambda}{N\mu}$. Since $\bar{q}_N$ is greater than the average, then there must be some queue $\bar{q}_k$, $1 \le k \le N-1$, that is less than the average, so $\bar{q}_k = \frac{\lambda}{\mu N} - \delta$ for some $\delta > 0$. We can use this together with the condition $\dot{\bar{q}}_i = 0$ to get an expression

$$\sum_{i=1}^{N} \exp\big(-\bar{q}_i(t-\Delta)\big) = \frac{\lambda}{\mu} \cdot \frac{\exp\big(-\frac{\lambda}{N\mu} + \delta\big)}{(\frac{\lambda}{N\mu} - \delta)}, \tag{5.2}$$

which can now be used to show contradiction:

$$\dot{\bar{q}}_N(t) = \lambda \frac{\exp\left(-\frac{\lambda}{N\mu} - \epsilon\right)}{\frac{\lambda}{\mu} \cdot \frac{\exp(-\frac{\lambda}{N\mu}+\delta)}{(\frac{\lambda}{N\mu}-\delta)}} - \mu\left(\frac{\lambda}{N\mu} + \epsilon\right) \tag{5.3}$$

$$= -\frac{\lambda}{N}\left(1 - e^{-\epsilon-\delta}\right) - \mu\left(\epsilon + \delta e^{-\epsilon-\delta}\right) < 0. \tag{5.4}$$

Hence $\bar{q}_i$ is not an equilibrium, and so the equilibrium must be unique. ∎

5.1.2. Showing stability of the equilibrium. The following proposition is used to prove the stability of the equilibrium.

Proposition 5.1. *If there is a root $r = x + iy$ of the characteristic equation*

$$r = \alpha + \beta e^{-r\Delta} \tag{5.5}$$

with positive real part ($x > 0$) then it is bounded by $x \leq \alpha + |\beta|$ and $|y| \leq |\beta|$.

Proof. Plug $r = x + iy$ into Equation (5.5) and separate real and imaginary parts to get

$$\cos(y\Delta) = \frac{e^{x\Delta}(x-\alpha)}{\beta}, \quad \sin(y\Delta) = -\frac{e^{x\Delta}y}{\beta} \tag{5.6}$$

These equations give the inequalities

$$-1 \leq \frac{e^{x\Delta}(x-\alpha)}{\beta} \leq 1, \quad -1 \leq -\frac{e^{x\Delta}y}{\beta} \leq 1 \tag{5.7}$$

Assuming that $x > 0$ and $\Delta \geq 0$, we know that $e^{x\Delta} \geq 1$. Therefore inequalities reduce to

$$-1 \leq \frac{(x-\alpha)}{\beta} \leq 1, \quad -1 \leq -\frac{y}{\beta} \leq 1, \tag{5.8}$$

and give the desired bounds $x \leq \alpha + |\beta|$ and $|y| \leq |\beta|$. ∎

5.1.3. Third order Taylor expansion. A third order Taylor expansion of $\dot{q}_1(t)$ and $\dot{q}_2(t)$ is used to approximate the deviation of the queues from the equilibrium. This is required both by the Lindstedt's method and by the the slow flow method. To find the expansion, we define new functions $\tilde{u}_1$ and $\tilde{u}_2$ that represent the deviation of the queues q_1 and q_2 from the equilibrium state at $\frac{\lambda}{2\mu}$

$$q_1(t) = \frac{\lambda}{2\mu} + \tilde{u}_1(t), \quad q_2(t) = \frac{\lambda}{2\mu} + \tilde{u}_2(t). \tag{5.9}$$

Equations (2.2) - (2.3) give expressions for $\dot{\tilde{u}}_1(t)$ and $\dot{\tilde{u}}_2(t)$, which can be approximated by with a third degree polynomial about the equilibrium point $\tilde{u}_1(t) = \tilde{u}_2(t) = 0$. We denote the approximations by $w_1(t)$ and $w_2(t)$

$$\dot{w}_1(t) = \lambda\left(-\frac{w_1 - w_2}{4} + \frac{w_1^3 - 3w_2w_1^2 + 3w_1w_2^2 - w_2^3}{48}\right)(t-\Delta) - \mu w_1(t) \tag{5.10}$$

$$\dot{w}_2(t) = \lambda\left(-\frac{w_2 - w_1}{4} + \frac{w_2^3 - 3w_1w_2^2 + 3w_2w_1^2 - w_1^3}{48}\right)(t-\Delta) - \mu w_2(t). \tag{5.11}$$

5.1.4. Reduction to one cubic delay equation. The symmetry of Equations (5.10) - (5.11) allows the equations to become uncoupled. We consider sum and the difference of w_1 and w_2,

$$\tilde{v}_1(t) = w_1(t) + w_2(t), \quad \tilde{v}_2(t) = w_1(t) - w_2(t). \tag{5.12}$$

This change of variables leads to the differential equations

$$\dot{\tilde{v}}_1(t) = -\mu\big(w_1(t) + w_2(t)\big) = -\mu\tilde{v}_1(t) \tag{5.13}$$

$$\dot{\tilde{v}}_2(t) = \lambda\left(-\frac{\tilde{v}_2(t-\Delta)}{2} + \frac{\tilde{v}_2^3(t-\Delta)}{24}\right) - \mu\tilde{v}_2(t), \tag{5.14}$$

which are uncoupled. Equation (5.13) has the solution $\tilde{v}_1(t) = Ce^{-\mu t}$ so $\tilde{v}_1(t)$ decays to 0 regardless of what the delay parameter is, making $\tilde{v}_2(t)$ the function of interest.

5.1.5. Limit cycle stability via floquet exponents. Theorem (2.5) shows that the Hopf bifurcations are supercritical by perturbing the system about the point of bifurcation. However, the stability of limit cycles can also be determined by projecting the infinite-dimensional DDE on a center manifold, and then finding the characteristic floquet exponent of the resulting system of ODE's. This approach is explained in detail by Hassard et al. [14]. In our case, the DDE is given by

$$\dot{v} = -\mu v(t) - \frac{\lambda}{2}v(t-\Delta) + f\big(v(t), v(t-\Delta)\big), \tag{5.15}$$

where $f\big(v(t), v(t-\Delta)\big)$ contains all the nonlinear terms. To project Equation (5.15) onto a center manifold, we follow Chapter 14.3 in [28] precisely. First, we get rid of delay in our equation by defining $v_t(\theta) = v(t+\theta)$ for $\theta \in [-\Delta, 0]$ and the operators

$$Av_t(\theta) = \begin{cases} \frac{\partial x_t(\theta)}{\partial \theta} & \text{for} \quad \theta \in [-\Delta, 0) \\ -\mu v_t(0) - \frac{\lambda}{2}v_t(-\Delta) & \text{for} \quad \theta = 0 \end{cases} \tag{5.16}$$

$$Fv_t(\theta) = \begin{cases} 0 & \text{for} \quad \theta \in [-\Delta, 0) \\ f\big(v_t(0), v_t(-\Delta)\big) & \text{for} \quad \theta = 0 \end{cases} \tag{5.17}$$

so that the DDE (5.15) can be written as

$$\frac{d}{d\theta}v_t(\theta) = Av_t(\theta) + Fv_t(\theta). \tag{5.18}$$

We assume that $\Delta = \Delta_{cr}$, so there is a pair of purely imaginary roots $\Lambda = \pm i\omega_{cr}$ with the corresponding eigenfunctions $s_1(\theta)$ and $s_2(\theta)$ such that

$$A\big(s_1(\theta) + is_2(\theta)\big) = i\omega_{cr}\big(s_1(\theta) + is_2(\theta)\big). \tag{5.19}$$

The solution v_t of Equation (5.18) can then be expressed as a sum of points lying in the center subspace spanned by $s_1(\theta)$ and $s_2(\theta)$, and the points that don't lie in the center subspace, which is the rest of the solution and we denote it by w:

$$v_t(\theta) = y_1(t)s_1(\theta) + y_2(t)s_2(\theta) + w(t, \theta). \tag{5.20}$$

The idea of the center manifold reduction is to approximate w as a function of y_1 and y_2 (the center manifold), therefore replacing the infinite dimensional system with a two dimensional approximation. After some algebra we can determine y_1 and y_2 to be

$$\dot{y}_1 = \omega y_2 - \frac{\big(4\omega y_2 + 4(\mu + \Delta\mu^2 + \Delta\omega^2)y_1\big)^3}{24\lambda^2\big((1+\Delta\mu)^2 + \Delta^2\omega^2\big)^3} + O\big(y_i^5\big) \tag{5.21}$$

$$\dot{y}_2 = -\omega y_1 + O\big(y_i^5\big). \tag{5.22}$$

Now we will follow the technique in Chapter 1 of Hassard et al. [14] to analyze the stability of y_1 and y_2. The system of ODE's (5.21) - (5.22) can be equivalently written as

$$\dot{z} = \omega z + \sum_{2\le i+j\le L} g_{ij}\frac{z^i\bar{z}^j}{i!j!} + O(|z|^{L+1}), \tag{5.23}$$

where z is a complex function $z = y_2 + iy_1$, and $\bar{z}$ is its complex conjugate. The coefficients g_{ij} can be determined from Equations (5.21) - (5.23), and we find that $g_{20} = g_{02} = g_{11} = 0$ and

$$g_{21} = -\frac{2(\mu - i\omega_{cr})(\mu + i\omega_{cr})^2}{\lambda^2(1+\Delta_{cr}(\mu - i\omega_{cr}))(1+\Delta_{cr}(\mu + i\omega_{cr}))^2}. \tag{5.24}$$

Further, if the floquet exponent is negative, then the bifurcating periodic solutions of Equation (5.23) are asymptotically, orbitally stable with asymptotic phase. When Δ is sufficiently close to Δ_{cr}, the floquet exponent is of the same sign as β_2 that is given by Equation (5.9) in Chapter 1 of Hassard et al. [14]:

$$\beta_2 = 2\ \mathrm{Re}\left[\frac{i}{2\omega_{cr}}\big(g_{20}g_{11} - 2|g_{11}|^2 - \frac{1}{3}|g_{02}|^2\big) + \frac{1}{2}g_{21}\right]. \tag{5.25}$$

Hence,

$$\beta_2 = -\frac{2(\mu^2+\omega_{cr}^2)(\mu + \Delta_{cr}\mu^2 + \Delta_{cr}\omega_{cr}^2)}{\lambda^2((1+\Delta_{cr}\mu)^2 + \Delta_{cr}^2\omega_{cr}^2)^2} < 0, \tag{5.26}$$

so the floquet exponent is negative and the limit cycle is stable in its center manifold.

5.2. Moving Average Model.

5.2.1. Showing the existence and uniqueness of equilibrium.

Proof of Theorem (3.3). Suppose the queues are in equilibrium. Then $q_1(t) = q_1^*$, $q_2(t) = q_2^*$, $m_1(t) = \frac{1}{\Delta}\int_{t-\Delta}^t q_1(s)ds = q_1^*$, and $m_2(t) = \frac{1}{\Delta}\int_{t-\Delta}^t q_2(s)ds = q_2^*$. By summing Equations (3.7) - (3.8) we find

$$\lambda - \mu(q_1^* + q_2^*) = 0, \quad q_1^* = \frac{\lambda}{\mu} - q_2^*. \tag{5.27}$$

Eliminating q_1^* from Equations (3.7) - (3.8) and subtracting one equation from the other, we find that for $x = 2q_2^* - \frac{\lambda}{\mu}$

$$x = \frac{\lambda}{\mu}\left(\frac{1-e^x}{1+e^x}\right). \tag{5.28}$$

Since $\frac{\lambda}{\mu} > 0$, when $x > 0$ the right-hand side of Equation (5.28) is negative so $x \le 0$. Similarly, when $x < 0$ then the right hand side of the equation is positive, which means that $x = 0$ is the only solution. Hence $q_2^* = \frac{\lambda}{2\mu}$ and $q_1^* = \frac{\lambda}{\mu} - q_2^* = \frac{\lambda}{2\mu}$ is the only equilibrium point of $q_1(t)$ and $q_2(t)$, which implies that $m_1(t) = m_2(t) = \frac{\lambda}{2\mu}$. ∎

5.2.2. Showing stability of the equilibrium. The equilibrium is stable whenever all eigenvalues of the characteristic equation (3.16) have negative real parts. The following propositions help to establish that.

Proposition 5.2. *Any real eigenvalue of the characteristic equation (3.16) is negative.*

Proof. Under the assumption $\Lambda \neq 0$ and $\Lambda \in \mathbb{R}$, the characteristic equation can be rewritten as

$$1 + \frac{2\Delta}{\lambda} \cdot \Lambda(\Lambda + \mu) = e^{-\Lambda\Delta}. \tag{5.29}$$

The left hand side (LHS) and the right hand side (RHS) intersect at $\Lambda = 0$, and for $\Lambda > 0$ the LHS is monotonically increasing while the RHS is monotonically decreasing. Hence when $\Lambda \in \mathbb{R}$, this equality can only hold for $\Lambda < 0$. ∎

Proposition 5.3. *If $\Delta \geq \frac{\lambda}{\mu^2}$, then any complex eigenvalue of the Equation (3.16) has a negative real part.*

Proof. We will argue by contradiction. Assume that $\Delta \geq \frac{\lambda}{\mu^2}$, $a \geq 0$, and $b \neq 0$, for some $\Lambda = a + ib$ where $a, b \in \mathbb{R}$. We substitute Λ into Equation (3.16) and separate the real and imaginary parts:

$$\cos(b\Delta)e^{-a\Delta}\lambda = 2a^2\Delta - 2b^2\Delta + \lambda + 2a\mu\Delta \tag{5.30}$$

$$\sin(b\Delta)e^{-a\Delta}\lambda = -2b\Delta(2a + \mu). \tag{5.31}$$

Summing the squares of the two equations, we get

$$e^{-2a\Delta}\lambda^2 = \left(2a^2\Delta - 2b^2\Delta + \lambda + 2a\mu\Delta\right)^2 + \left(2b\Delta(2a + \mu)\right)^2, \tag{5.32}$$

and after some algebra we find

$$b^2 \leq \frac{1}{\Delta}(\lambda + 2a\mu\Delta + 2a^2\Delta) - (2a + \mu)^2 \tag{5.33}$$

$$= \frac{\lambda}{\Delta} - \mu^2 - 2a(a + \mu) \leq -2a(a + \mu) \leq 0, \tag{5.34}$$

so b must be 0, which contradicts our assumption. Therefore $\mathrm{Re}[\Lambda] = a < 0$ for any complex eigenvalue when $\Delta > \frac{\lambda}{\mu^2}$. ∎

It is now left to establish the stability of the equilibrium.

Proof of Theorem (3.3). We will show that for the specified range of Δ, all eigenvalues of the characteristic equation (3.16) have negative real parts. Recall that prior to deriving the characteristic equation, we considered the case with the trivial eigenvalue separately, so to analyze the stability we now only need to look at the non-trivial eigenvalues. Proposition (5.2) shows that any nontrivial real eigenvalue must be negative. Hence, it remains to show that the complex eigenvalues have negative real parts.

Case 1. Suppose the characteristic equation (3.29) does not have positive roots Δ_{cr}. This implies that a complex eigenvalue Λ never reaches the imaginary axis as Δ varies. Since Λ is continuous as a function of Δ, then $\mathrm{Re}[\Lambda]$ must be of the same sign for all $\Delta > 0$. Proposition (3.2) shows that for sufficiently small Δ, all complex eigenvalues have negative real parts, which is therefore true for all $\Delta > 0$.

Case 2. Suppose Equation (3.29) has at least one positive root Δ_{cr}. By the continuity of Λ with respect to Δ, $\mathrm{Re}[\Lambda]$ must be of the same sign on the interval where Δ is less than the smallest

positive root of Equation (3.29), and by Proposition (3.2) the sign is negative. Same holds when Δ is greater than the largest root Δ_{cr} of Equation (3.29). Any root Δ_{cr} is less than $\frac{\lambda}{\mu^2}$ by the condition $0 \neq \omega_{cr} \in \mathbb{R}$, and for Δ exceeding $\frac{\lambda}{\mu^2}$ all complex eigenvalues have negative real parts by Proposition (5.3). Therefore, the continuity of Λ implies that $\mathrm{Re}[\Lambda] < 0$ whenever Δ exceeds the largest root Δ_{cr}.

We showed that for the specified ranges of Δ, all eigenvalues have negative real parts and therefore the equilibrium is stable. ∎

5.2.3. Third order polynomial expansion. We will perform a Taylor series expansion for the deviations about the equilibrium (3.13) of equations (3.7) - (3.10) and keep terms up to the third order. To start, we find the perturbations of our functions from the equilibrium,

$$q_1(t) = \frac{\lambda}{2\mu} + \tilde{u}_1(t), \quad q_2(t) = \frac{\lambda}{2\mu} + \tilde{u}_2(t), \tag{5.35}$$

$$m_1(t) = \frac{\lambda}{2\mu} + \tilde{u}_3(t), \quad m_2(t) = \frac{\lambda}{2\mu} + \tilde{u}_4(t), \tag{5.36}$$

and from Equations (3.7) - (3.10) we find their derivatives. A third order polynomial expansion of $\dot{\tilde{u}}_i(t)$ is given by $\dot{w}_i(t)$, where

$$\begin{aligned} \dot{w}_1(t) &= \lambda \cdot \left(- \frac{w_3(t) - w_4(t)}{4} \right) - \mu w_1(t) \\ &+ \lambda \cdot \left(\frac{w_3^3(t) - 3w_4(t)w_3^2(t) + 3w_3(t)w_4^2(t) - w_4^3(t)}{48} \right) \end{aligned} \tag{5.37}$$

$$\begin{aligned} \dot{w}_2(t) &= \lambda \cdot \left(- \frac{w_4(t) - w_3(t)}{4} \right) - \mu w_2(t) \\ &+ \lambda \cdot \left(\frac{w_4^3(t) - 3w_3(t)w_4^2(t) + 3w_4(t)w_3^2(t) - w_3^3(t)}{48} \right) \end{aligned} \tag{5.38}$$

$$\dot{w}_3(t) = \frac{1}{\Delta}\Big(w_1(t) - w_1(t-\Delta)\Big) \tag{5.39}$$

$$\dot{w}_4(t) = \frac{1}{\Delta}\Big(w_2(t) - w_2(t-\Delta)\Big). \tag{5.40}$$

5.2.4. Reduction to two cubic delay equations. We will utilize the symmetry of the Equations (5.37) - (5.40) to simplify our problem by uncoupling the four equations. To do so we introduce a change of variables

$$\tilde{v}_1 = w_1 + w_2, \quad \tilde{v}_2 = w_1 - w_2, \quad \tilde{v}_3 = w_3 + w_4, \quad \tilde{v}_4 = w_3 - w_4. \tag{5.41}$$

The expressions for variables $\tilde{v}_1$ and $\tilde{v}_3$ are uncoupled from $\tilde{v}_2$ and $\tilde{v}_4$

$$\dot{\tilde{v}}_1 = -\mu \tilde{v}_1(t), \qquad \dot{\tilde{v}}_3 = \frac{1}{\Delta}\Big(\tilde{v}_1(t) - \tilde{v}_1(t-\Delta)\Big), \tag{5.42}$$

$$\dot{\tilde{v}}_2 = \lambda\left(- \frac{\tilde{v}_4(t)}{2} + \frac{\tilde{v}_4(t)^3}{24} \right) - \mu \tilde{v}_2(t), \qquad \dot{\tilde{v}}_4 = \frac{1}{\Delta}\Big(\tilde{v}_2(t) - \tilde{v}_2(t-\Delta)\Big). \tag{5.43}$$

Furthermore, $\tilde{v}_1(t)$ and $\tilde{v}_3(t)$ can be solved directly and they converge to zero as $t \to \infty$. Hence we are left with only two functions of further interest, $\tilde{v}_2$ and $\tilde{v}_4$.