Delay-Differential Equations Applied to Queueing Theory

Jamol Pender*, Richard H. Rand** and Elizabeth Wesson***
*School of Operations Research and Information Engineering, Cornell University, Ithaca, NY, USA
**Department of Mathematics and Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY, USA
***Department of Mathematics, Cornell University, Ithaca, NY, USA

Summary. Queueing theory is the mathematical theory of customers waiting in line. The term “customer" could refer to a human customer, but it could also refer to, for example, an automobile waiting in line to pass through a toll plaza. In the case that there are muliple lines, the customer often has the choice of which line to wait in. Delay of queue length information has the potential to influence the decision of a customer as which line to wait in. In this work we introduce the reader to the typical structure of the differential equations of queueing theory, and we study a model involving two lines in which the information regarding the length of a line provided to the customer is not current, but rather is delayed. That is, the provided information gives the state of the system at a previous time rather than at the current time. Thus the governing equations become delay-differential equations. In particular we show that the delay can cause oscillations in the length of the lines due to Hopf bifurcations.

Introduction

Queueing theory represents a new area of application for nonlinear dynamics. It is expected that the typical nonlinear dynamics researcher is unfamiliar with the mathematical models of queueing theory, and so we present a brief summary of the model equations. The interested reader is referred to [1, 2] for an introduction to basic queueing models. In what is called a “fluid model", the dependent variable $q(t)$ represents the length of a line (or queue). The time history of $q(t)$ is governed by a differential equation of the form:

$$\dot{q}(t) = \text{arrival rate of customers} - \text{rate at which customers are serviced} \tag{1}$$
$$= \lambda - \mu\, q(t) \tag{2}$$

A simple model might have the arrival rate λ and the service coefficient μ taken as constants. In a system with two queues, $q_1(t)$ and $q_2(t)$, the total arrival rate to both queues could be modeled as equal to the constant rate λ, but each queue could receive customers at a rate proportional to $\exp(-q_i t)$, so that the first term in eq.(2) would become:

$$\lambda \cdot \frac{\exp(-q_i(t))}{\exp(-q_1(t)) + \exp(-q_2(t))}. \tag{3}$$

In this work we will be interested in a model for customer choice in which the arrival rates are based on delayed information, leading to the following pair of delay-differential equations, where Δ represents the delay:

$$\dot{q}_1(t) = \lambda \cdot \frac{\exp(-q_1(t-\Delta))}{\exp(-q_1(t-\Delta)) + \exp(-q_2(t-\Delta))} - \mu q_1(t) \tag{4}$$
$$\dot{q}_2(t) = \lambda \cdot \frac{\exp(-q_2(t-\Delta))}{\exp(-q_1(t-\Delta)) + \exp(-q_2(t-\Delta))} - \mu q_2(t). \tag{5}$$

Hopf bifurcations

Simulation of eqs.(4),(5) shows that if the delay is larger than a certain critical delay Δ_{cr}, the length of the queues will oscillate periodically in time.
Theorem: For the queueing model (4),(5), the critical delay is given by the following expression

$$\Delta_{cr}(\lambda, \mu) = \frac{2 \arccos(-2\mu/\lambda)}{\sqrt{\lambda^2 - 4\mu^2}}. \tag{6}$$

Proof: The equilibrium in eqs.(4),(5) is given by $q_1(t) = q_2(t) = \frac{\lambda}{2\mu}$.

Setting
$$q_1(t) = \frac{\lambda}{2\mu} + u_1(t), \qquad q_2(t) = \frac{\lambda}{2\mu} + u_2(t) \tag{7}$$

and linearizing in u_1 and u_2, we obtain

$$\dot{u}_1(t) = -\frac{\lambda}{4} \cdot (u_1(t-\Delta) - u_2(t-\Delta)) - \mu \cdot u_1(t) \tag{8}$$
$$\dot{u}_2(t) = -\frac{\lambda}{4} \cdot (u_2(t-\Delta) - u_1(t-\Delta)) - \mu \cdot u_2(t). \tag{9}$$

Next we uncouple eqs.(8),(9) by setting $v_1(t) = u_1(t) + u_2(t)$ and $v_2(t) = u_1(t) - u_2(t)$ giving:

$$\dot{v}_1(t) = -\mu \cdot v_1(t) \tag{10}$$

$$\dot{v}_2(t) = -\frac{\lambda}{2} \cdot v_2(t-\Delta) - \mu \cdot v_2. \tag{11}$$

The general solution to eq.(10) is $v_1(t) = c_1 \exp(-\mu t)$ and is bounded and stable. For eq.(11), we set $v_2(t) = \exp(rt)$ and obtain

$$r = -\frac{\lambda}{2} \cdot \exp(-r\Delta) - \mu \tag{12}$$

For a Hopf bifurcation, we set $r = i\omega$ giving $i\omega = -\frac{\lambda}{2}(\cos\omega\Delta - i\sin\omega\Delta) - \mu$. Equating the real and imaginary parts to zero, we get:

$$\cos\omega\Delta_{cr} = -\frac{2 \cdot \mu}{\lambda} \quad \text{and} \quad \sin\omega\Delta_{cr} = \frac{2 \cdot \omega}{\lambda} \tag{13}$$

Squaring both equations and adding them together, we get that $\omega = \frac{1}{2}\sqrt{\lambda^2 - 4\mu^2}$, which when substituted into the first of eqs.(13) yields the required condition (6) stated in the Theorem.

Conclusions

In this work we have studied a deterministic queueing model that incorporates customer choice and delayed queue length information. We have derived an expression for an explicit threshold for the critical delay where below the threshold the two queues are balanced and converge to the equilibrium. However, when Δ is larger than the threshold, the equilibrium point is unstable and an oscillation in queue length results.

This analysis is the first of its kind in the queueing and operations research literature. It is important for businesses and managers to determine and know these thresholds since using delayed information can have such a large impact on the dynamics of the business. Even small delays can cause oscillations and it is of great importance for managers of these service systems to understand when oscillations can arise based on the arrival and service parameters of the queueing model.

The reader is referred to [3] for a more complete treatment of this subject which also analyzes a moving average delay differential equation. The reader is also referred to [4] for an introduction to delay-differential equations.

References

[1] "Approximations for the moments of nonstationary and state dependent birth-death queues", Stefan Engblom and Jamol Pender. arXiv preprint arXiv:1406.6164 (2014).

[2] "Gaussian skewness approximation for dynamic rate multi-server queues with abandonment", William A. Massey and Jamol Pender. Queueing Systems 75.2-4 (2013): 243-277.

[3] “Managing Information in Queues: The Impact of Giving Delayed Information to Customers", Jamol Pender, Richard H. Rand and Elizabeth Wesson, arXiv: 1610.01972v1.pdf (2016)

[4] “Lecture Notes in Nonlinear Vibrations", R.H.Rand, Published on-line by The Internet-First University Press (2012) http://ecommons.library.cornell.edu/handle/1813/28989