The Value of Cooperation: From AIMD to Flipped Classroom Teaching

Chee Wei Tan

This paper is in honor of Professor Dah-Ming Chiu on the occasion of his 70th birthday

November 12, 2021

Talk Outline

- The Value of Cooperation in AIMD Abstraction
- Teaching Examples via Convex Optimization Theory and Perron-Frobenius Theory
- Integrate Classical Performance Evaluation and Modern Data Science Techniques
- A Parallel between AIMD and Flipped Classroom Teaching
- Technology-enabled Pedagogy and Blended Learning
- Thanks to Dah-Ming Chiu, Y. C. Tay, Vittoria de Nitto Persone, R. Srikant, Geoffrey M. Voelker, Lorenzo Alvisi, Mung Chiang, Steven Low and Stephen Boyd

Abstraction and its Importance to Teaching

- Performance analysis often depends on a mathematical model.
- Elegant performance models such as the queuing models and stochastic models are good approximations of real systems
- Performance model choice often a trade-off between mathematical tractability and real-world relevance.
- Finding an appropriate model becomes even more crucial.
- Right "abstraction" helps to focus on most interesting and crucial aspect of a problem, and opens the door to a variety of performance analysis techniques.
- Allows one abstraction to relate to other abstractions, and to explore the trade-off between tractability and approximation.

AIMD Abstraction

- The Additive-Increase-Multiplicative-Decrease (AIMD) algorithm abstraction proposed by Dah-Ming Chiu and Raj Jain [1] is influential in design of Transmission Control Protocol (TCP) protocol for Internet congestion control.
- In the early days of the Internet, many different models to study congestion control, some from the viewpoint of a single flow while others assume synchronous events or fluid flows.
- AIMD abstracts the congestion control problem as one with social consequences, whereby the parties encountering congestion work cooperatively to arrive at a good solution.

[1] Dah-Ming Chiu and Raj Jain. Analysis of the increase and decrease algorithms for congestion avoidance in computer networks. Computer Networks and ISDN Systems, 17(1):1–14, 1989.

Illustrating an Abstraction

- Ingenuity behind the AIMD abstraction can be appreciated through a classical illustration of two flows (Figure 1)
- Usher in the development of robust TCP algorithms [2]
- Classical illustration found in a number of popular computer networking textbooks [3] [4]
- Insightful illustration to accompany the right abstraction to a complex problem can be helpful to teaching
- Usher in new performance analysis techniques like convex optimization theory and Perron-Frobenius theory

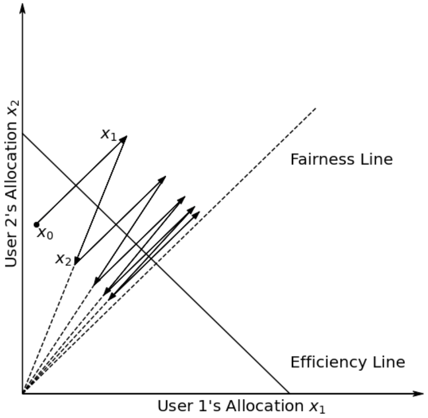
---

[2] Van Jacobson. Congestion avoidance and control. ACM SIGCOMM computer communication review, 18(4):314–329, 1988.

[3] Jim Kurose and Keith Ross. Computer Networking: A Top-Down Approach, 8th edition. Pearson, 2020.

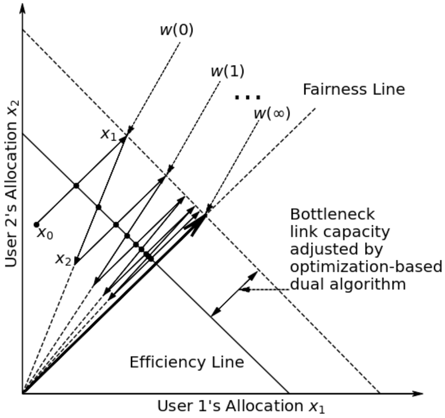
[4] Andrew Tanenbaum and David J. Wetherall. Computer Networks, 5th edition. Prentice Hall, 2010.

Classical Visual Proof of AIMD



A visual proof of the AIMD abstraction

Visualizing the Theories Behind AIMD



Teaching convex optimization theory and Perron-Frobenius Theory.

An Optimization Theoretic Perspective of AIMD

- A theoretical framework called Network Utility Maximization (NUM) to analyze both the equilibrium and the dynamical nature of TCP algorithms [5] [6]
- Elegant framework brings tools and ideas from convex optimization theory to bear on the design of internet congestion control algorithms
- By leveraging Lagrange duality and gradient descent algorithms, TCP can be interpreted as a dual algorithm that maximizes the aggregate utilities in the network
- AIMD viewed as solving a network utility maximization where the sending source rates and network congestion measures are interpreted as primal variables and dual variables respectively

[5] Rayadurgam Srikant. The mathematics of Internet congestion control. Springer Science & Business Media, 2004.

[6] Steven H Low. Analytical methods for network congestion control. Synthesis Lectures on Communication Networks, 10(1):1–213, 2017.

Network Utility Maximization(NUM)

- The basic network utility maximization of n sending sources can be formulated as follows:

$$\text{maximize} \sum_{s=1}^{n} U_s(x_s) \ \text{ subject to } Rx \preceq c, \ x \succeq 0,$$

  where x_s is the sending rate of source s given a routing matrix R with entries $R_{ls} = 1$ if source s uses link l with a link capacity c_l or 0 otherwise. At the tth iteration, source s solves:

$$x_s^*(q_s) = \text{argmax} \left\{ U_s(x_s) - q_s x_s \right\},$$

- The lth link runs the algorithm:

$$p_l(t+1) = \max \left\{ p_l(t) - \alpha(t) \left(c_l - \sum_{s: l \in L(s)} x_s^*(q_s(t)) \right), 0 \right\}.$$

NUM

- Using AIMD with an end-to-end marking probability q_s and a total delay D_s, TCP Reno has the utility function [7] [8]:

$$\text{arctan utility}: U_s(x_s) = \frac{\sqrt{3/2}}{D_s} \arctan\left(\sqrt{2/3}x_s D_s\right)$$

- Mathematical basis for other fairness abstraction like proportional fairness in TCP Vegas and FAST TCP
- Other forms of TCP do not leverage Lagrange duality directly as they are instead interpreted as solutions to a penalty function formulation of the optimization problem

[7] Steven H Low. Analytical methods for network congestion control. Synthesis Lectures on Communication Networks, 10(1):1–213, 2017.

[8] Rayadurgam Srikant. The mathematics of Internet congestion control. Springer Science & Business Media, 2004.

Teaching Optimization Theory

- Convex optimization theory important to networking applications [9] [10]
- Teach students to optimize performance analysis and data science techniques, e.g., link capacity shaped by the abstraction of wireless network optimization[11]
- Guide students to design optimization algorithms, data analytics to reduce data-driven models into performance models with parameters to be optimized
- Teach students to appreciate how to connect abstractions across the network protocol layers (primal and dual decomposition in Lagrange duality theory)

---

[9] Stephen Boyd and Lieven Vandenberghe, Convex Optimization, Cambridge University Press, 2004

[10] Mung Chiang, Networked life 20 questions and answers - Cambridge University Press, 2012

[11] Chee Wei Tan. Wireless network optimization by Perron-Frobenius theory. Foundations and Trends in Networking, 9(2-3):107–218, 2015.

Perron-Frobenius Theory Approach to AIMD

- Perspective via positive linear systems theory [12] [13]
- Let $w_s(k)$ denote the congestion window size of source s immediately before the kth network congestion event is detected by all the sources as shown in Figure 2.
- Let α_s and $0 < \beta_s < 1$ be the additive and multiplicative parameters of source s using the AIMD algorithm (that are conventionally set as 1 and 0.5) respectively
- Let $q_{\max}$ and P be, respectively, the maximum queue length of the congested bottleneck link and the maximum instantaneous number of sent unacknowledged packets that are in transit (e.g., $P = q_{\max} + BT$ where B is the bottleneck link service rate in packets per second and T is the round-trip time)

[12] Abraham Berman, Robert Shorten, and Douglas Leith. Positive matrices associated with synchronised communication networks. Linear Algebra and its Applications, 393:47–54, 2004.

[13] Martin Corless, Christopher King, Robert Shorten, and Fabian Wirth. AIMD Dynamics and Distributed Resource Allocation. Society for Industrial and Applied Mathematics, 2016.

Positive Linear System

- At the $(k+1)$th congestion event, source s's window satisfies

$$w_s(k+1) = \beta_s w_s(k) + \left(\frac{\alpha_s}{\sum_{i=1}^n \alpha_i} \right) \sum_{i=1}^{n} (1-\beta_i)\, w_i(k).$$

Let $w(k) = (w_1(k), \ldots, w_n(k))^T$ and write a positive system:

$$w(k+1) = Aw(k),$$

where

$$A = \begin{bmatrix} \beta_1 & 0 & \cdots & 0 \\ 0 & \beta_2 & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \beta_n \end{bmatrix} + \frac{1}{\sum_{i=1}^n \alpha_i} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \cdots \\ \alpha_n \end{pmatrix} (1-\beta_1, \; \cdots \;, \; 1-\beta_n)$$

Positive Linear System

- The spectrum of the matrix A (e.g., the Perron-Frobenius eigenvalue and eigenvectors) provides insights on fairness, rate of convergence and transient response: [14]

$$\lim_{k\to\infty} w(k) = \left(\frac{\alpha_1}{1-\beta_1}, \ldots, \frac{\alpha_n}{1-\beta_n}\right)^T,$$

  which, if specialized to the case of $\alpha_i = 1$ and $\beta_i = 0.5$ for all i, is proportional to the all-ones vector as it should be
- Fairness line as Perron-Frobenius right eigenvector
- Classical power method algorithm can simulate AIMD and to visualize the iterates as shown in Figure 2

[14] Abraham Berman, Robert Shorten, and Douglas Leith. Positive matrices associated with synchronised communication networks. Linear Algebra and its Applications, 393:47–54, 2004.

Teaching Perron-Frobenius Theory to Undergraduates

- Networking courses for undergraduates often use Figure 1 to show the fairness guarantee of AIMD
- Adapt curriculum to cover the classical linear Perron-Frobenius theory and the aforementioned eigenvector interpretation of the fairness line
- Combine performance analysis with data science techniques to study the transient behavior and convergence of flow rates in a bottleneck link
- Students can conduct experiments using data flows with different AIMD parameters $\{\alpha_s, \beta_s\}$ for all s to compete with the conventional TCP sources to empirically deduce the *cooperative* configuration (i.e., "TCP-friendly"):

$$\alpha_s = 2(1 - \beta_s) \quad \forall \ s$$

Teaching Perron-Frobenius Theory to Graduate Students

- Teaching curriculum can cover more general AIMD involving the nonlinear versions of the Perron-Frobenius theory[15] [16]
- Design data analytics based on TCP data measurement and apply machine learning to data-driven models to create computer-generated TCP algorithms
- Data analytics to validate assumptions, e.g., single bottleneck link
- Machine learning to automate this process when the bottleneck link may shift around in the network due to variable traffic conditions

[15] Chee Wei Tan. Wireless network optimization by Perron-Frobenius theory. Foundations and Trends in Networking, 9(2-3):107–218, 2015.

[16] Roger D. Nussbaum and Bas Lemmens. Nonlinear Perron-Frobenius Theory. Cambridge University Press, 2012.

Flipped Classroom Teaching

- In traditional university teaching, the teacher tends to just give lectures and hand out homework to students – there is less cooperation between teacher and students as well as between students
- A flipped classroom approach leverages the power of cooperation to improve the interaction between the teacher and students using feedback [17]
- Can we engineer flipped classroom teaching tools to enhance teacher-student interaction and cultivate the spirit of cooperation between the teacher and students?

[17] John Hattie and Helen Timperley. The power of feedback. Review of Educational Research, 77(1):81–112, 2007.

A parallel between AIMD and Classroom Teaching

- Classroom flipping requires active feedback through cooperation
- In AIMD/TCP, the users may wish to send as much data packets as possible without knowing the link capacity ahead in time. In teaching, what is the "*content capacity*" (related to teachers asking "*am I teaching too much?*" or "*am I teaching this not enough?*")?
- Back-off when teachers realize that they are teaching too much content?
- In AIMD/TCP, the actual information of link capacity and degree of fairness are not known to the users sending the data packets. In teaching, what is the "*comprehension capacity*" (related to teachers asking "*are some students being overwhelmed and dropping behind?*" or "*is this material accessible to all the students?*")?
- Slow start when teachers introduce new or more advanced concepts in class?

Technology-driven Pedagogy for Classroom Flipping

- Our mobile chatbot software technologies blend together Peer Instructions and Just-in-Time Teaching for in-person or remote instructions[18]
- Poll-quiz routine to regulate content delivery for the whole class to meet the "*content capacity*" and to use outside-class quizzes of varying difficulty levels to meet the "*comprehension capacity*" of individual students
- An optimization framework optimizes the frequency and difficulty levels of quizzes
- Data analytics collate the data from the series of poll-quiz routines to identify students' weakness in learning and the instructors' blind spots in online teaching

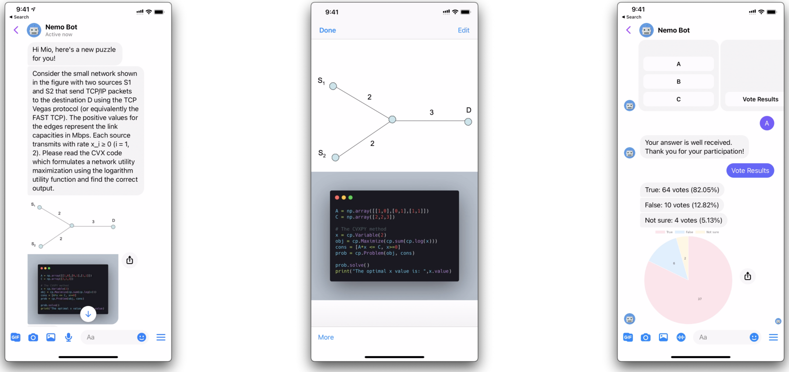
[18] Jingting Li, Lin Ling, and Chee Wei Tan. Blending peer instruction with just-in-time teaching: Jointly optimal task scheduling with feedback for classroom flipping. ACM Conference on Learning at Scale, 2021.

Technology-driven Pedagogy for Classroom Flipping

- The teacher first issues a poll whose instantaneous response outcome can be observed by the entire class, and a short discussion (e.g., peer discussions) ensues
- Then conclude by a time-limited quiz whose content is related to the prior poll
- Polls and quizzes are typically multiple-choice questions but they can be enriched with multimedia contents
- Enrich Human-Computer Interaction via automated hints or interactive human-computer input like touch-screen annotation with auto-grading capability [19] [20]

[19] Lin Ling and Chee Wei Tan. Human-assisted computation for auto-grading. IEEE ICDM Workshop on Machine Learning for Education, 2018.

[20] Lin Ling and Chee Wei Tan. Peer-grading at scale with rank aggregation. ACM Conference on Learning at Scale, 2021.



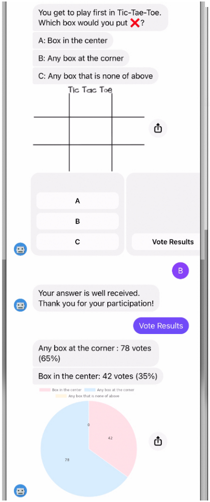
(a) (b) (c)

Figure: Students receive a poll via a mobile chatbot software on the topic of network utility maximization in (a) testing their understanding of the CVX optimization software [21] in (b). In (c), the student once having voted on the poll gets to see the whole class response to the poll before engaging in peer discussions and answering a timed quiz.

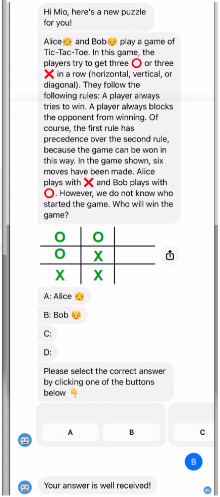
[21] Michael Grant and Stephen Boyd. CVX: http://cvxr.com/cvx, Mar 2014.

Poll-Quiz Routine

- In-class synchronous Peer Instruction activity
- Process:
  - Poll: 5 minutes
  - Statistical visualization
  - Discussion: 10 to 20 minutes
  - Quiz with scoring (content related or identical to poll)

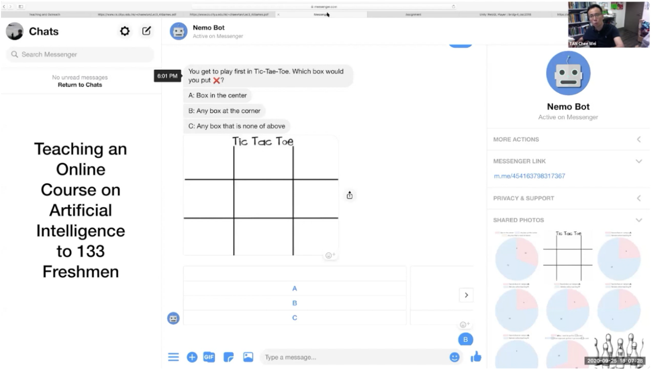


(a) Poll



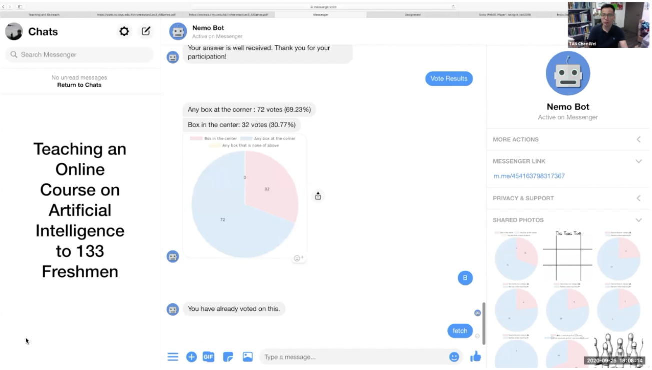
(b) Quiz

Online Classroom Flipping for Remote Teaching



A poll typically allows any student to respond intuitively or with the help of pre-class reading assignment.

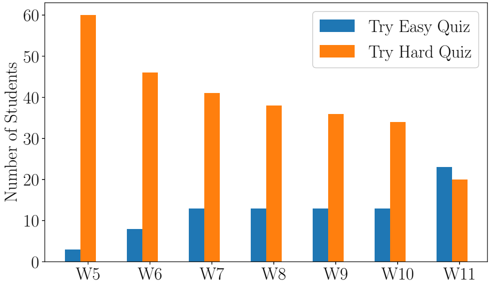
Instant Feedback to Students and Educator



A poll response that makes for great discussion among peers.

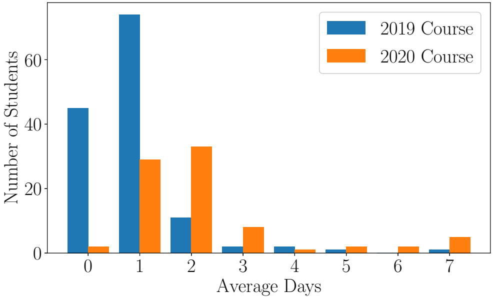
Usability Profile on Quiz Difficulty Level

- Students are willing to opt for harder quizzes
- How to incentivize students is key, e.g., more points for a harder quiz
- Open up new avenues to auto-track students' abilities



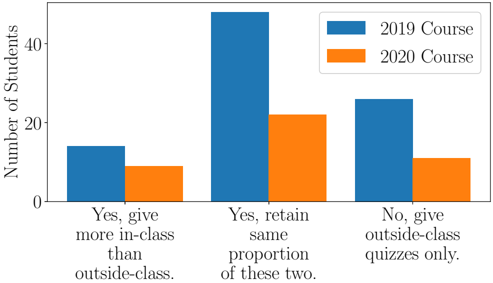
Usability Profile on Pre-class Quiz Activities

- In 2019, 96% subscribers answer the quizzes within two days.
- In 2020, 99% subscribers answer the quizzes within three days.



Students Vote on Blended Learning by End of Course

- More in-class quiz. (16%, 21%)
- Retain same proportion. (55%, 52%)
- Outside-class quiz only. (30%, 26%)



Conclusions

- Teaching performance evaluation and data science via AIMD
- Nonlinear mathematics like optimization theory and Perron-Frobenius theory are fundamental to AIMD abstraction and can enrich teaching curriculum
- A parallel between AIMD/TCP and flipped classroom teaching
- Technology-enabled pedagogy with poll-quiz routines to enhance cooperation between teachers and students
- Allow teachers to find suitable operating points in the "*content capacity*" and to understand limitations due to the "*comprehension capacity*"
- Classroom flipping dataset available at Kaggle `https://bit.ly/3y4YKAZ`
- Useful for fully online instruction setting, e.g. during COVID-19 pandemic, to "read a room" and blended learning [22]

[22] Maxwell Bigman and John Mitchell. Teaching online in 2020: Experiments, empathy, discovery. IEEE 7th Learning with MOOCS (LWMOOCS) 2020.