IBM Thomas J. Watson Research Center

Decentralized Optimization, Stochastic-Process Limits, and System Dynamics

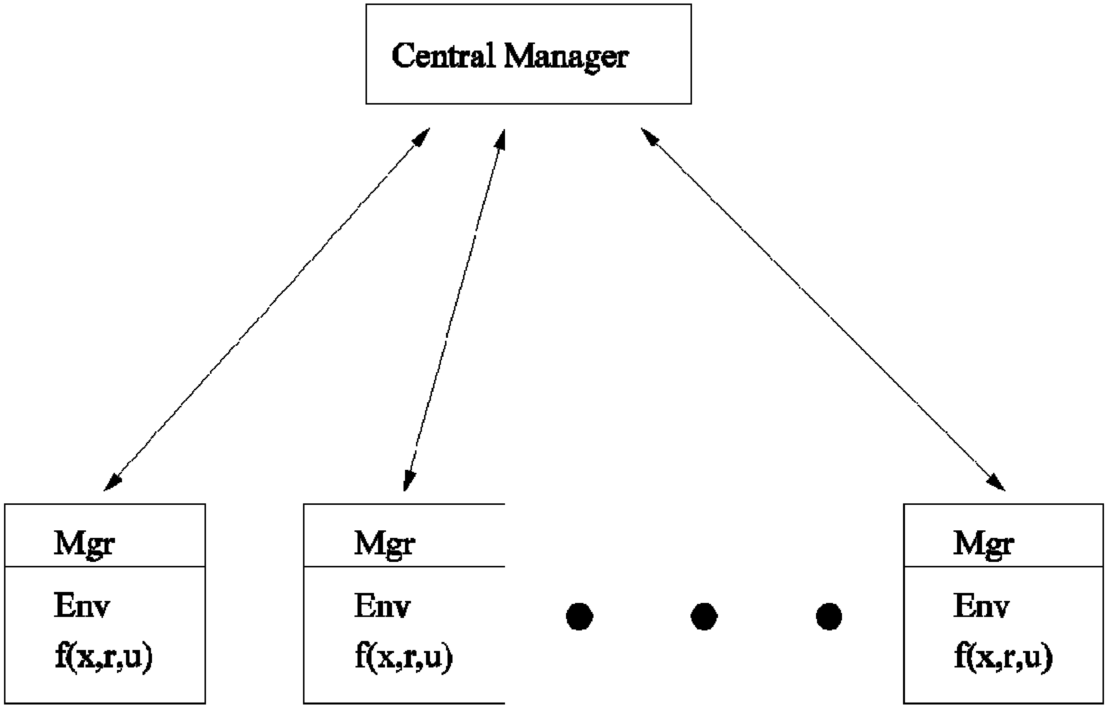
Mark S. Squillante*
Mathematical Sciences Department

January 19, 2005

* Based in part on joint works with T. Nowicki, C. Wu and X. Guo, Y. Lu

Problem Motivation

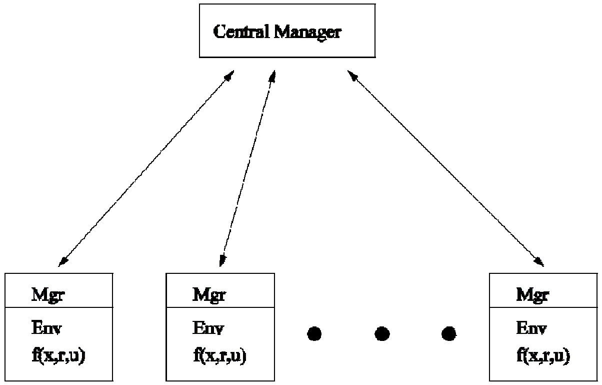
- Autonomic Systems: Computing systems, Sense-and-Respond systems, etc.
- Consider framework for decentralized optimization and dynamic optimal control
  - Decentralized approach is natural for large-scale autonomic systems due to overheads and delays
  - But is there any loss in optimality by using a decentralized approach vs. a centralized approach



General Overview

- Decentralized Optimization
  - Conditions for same quality of solution under decentralized as centralized
  - Hierarchical algorithmic issues
  - Representative application: Central Manager (CM) with multiple Envs
- Stochastic-Process Limits
  - Optimal routing problem for each Env (E_i) under renewal arrivals
  - Optimal routing problem for each Env (E_i) under correlated arrivals
- System Dynamics
  - Dynamics of optimal solutions for both

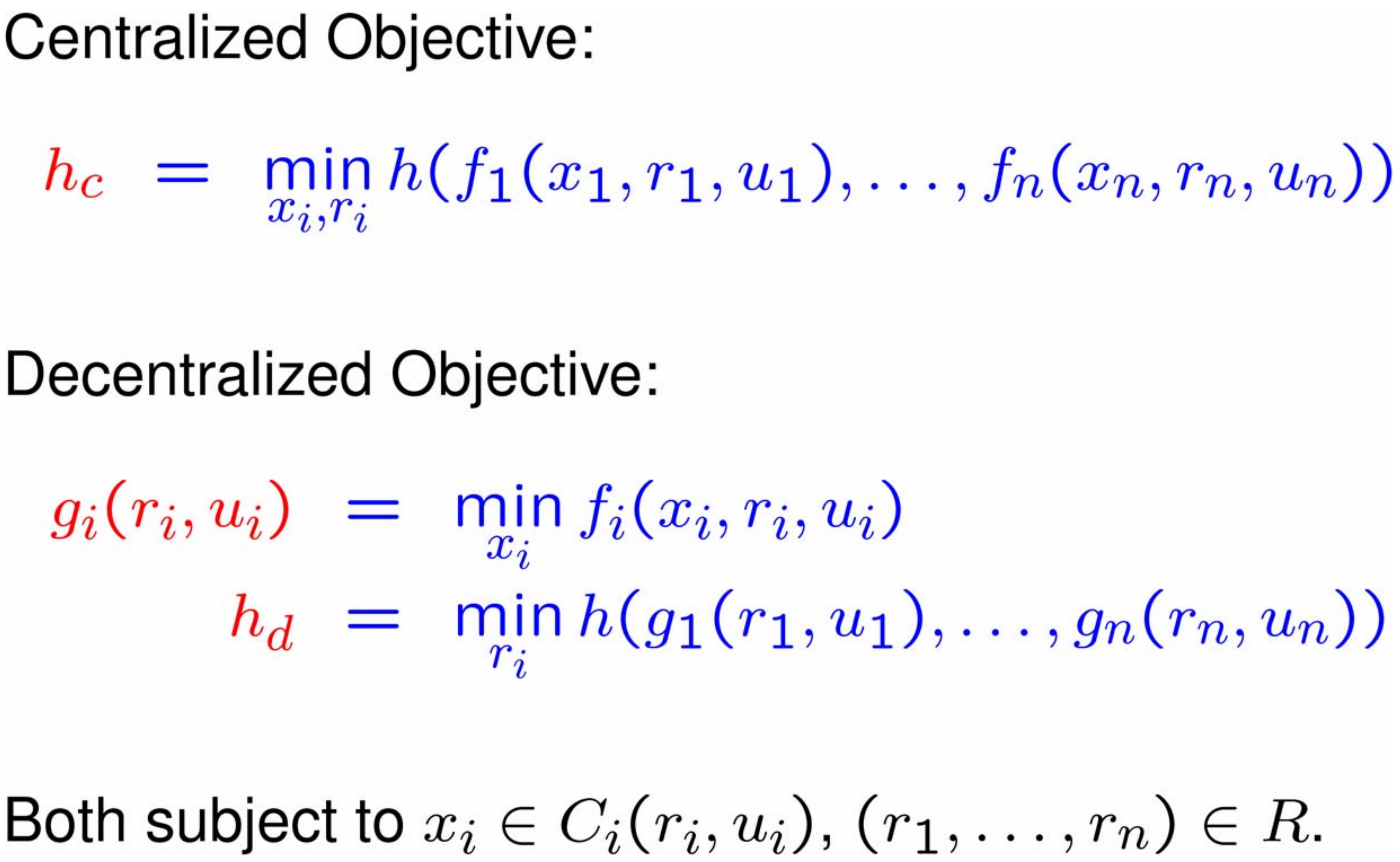
  CM and E_i under time-varying workloads



Decentralized Optimization: Total Cost

- Consider without any loss of generality minimizing a cost function
  - Maximizing f(¢) is equivalent to minimizing –f(¢)
- Cost function $f_i(x_i,r_i,u_i)$ is associated with each E_i, where
  - x_i is the set of variables that can be changed in E_i, e.g., routing parameters
  - r_i is the set of resources allocated to E_i by CM, e.g., server assignments
  - u_i is the set of external variables that affect E_i, e.g., workloads
- Set of variables x_i must satisfy the set of constraints $C_i(r_i,u_i)$
- Set of resources $(r_1,\ldots,r_n)$ assigned to Envs must satisfy the set of constraints R
- Total cost function for the entire system aggregates the cost of each E_i

Decentralized Optimization: Total Cost



Centralized Objective:

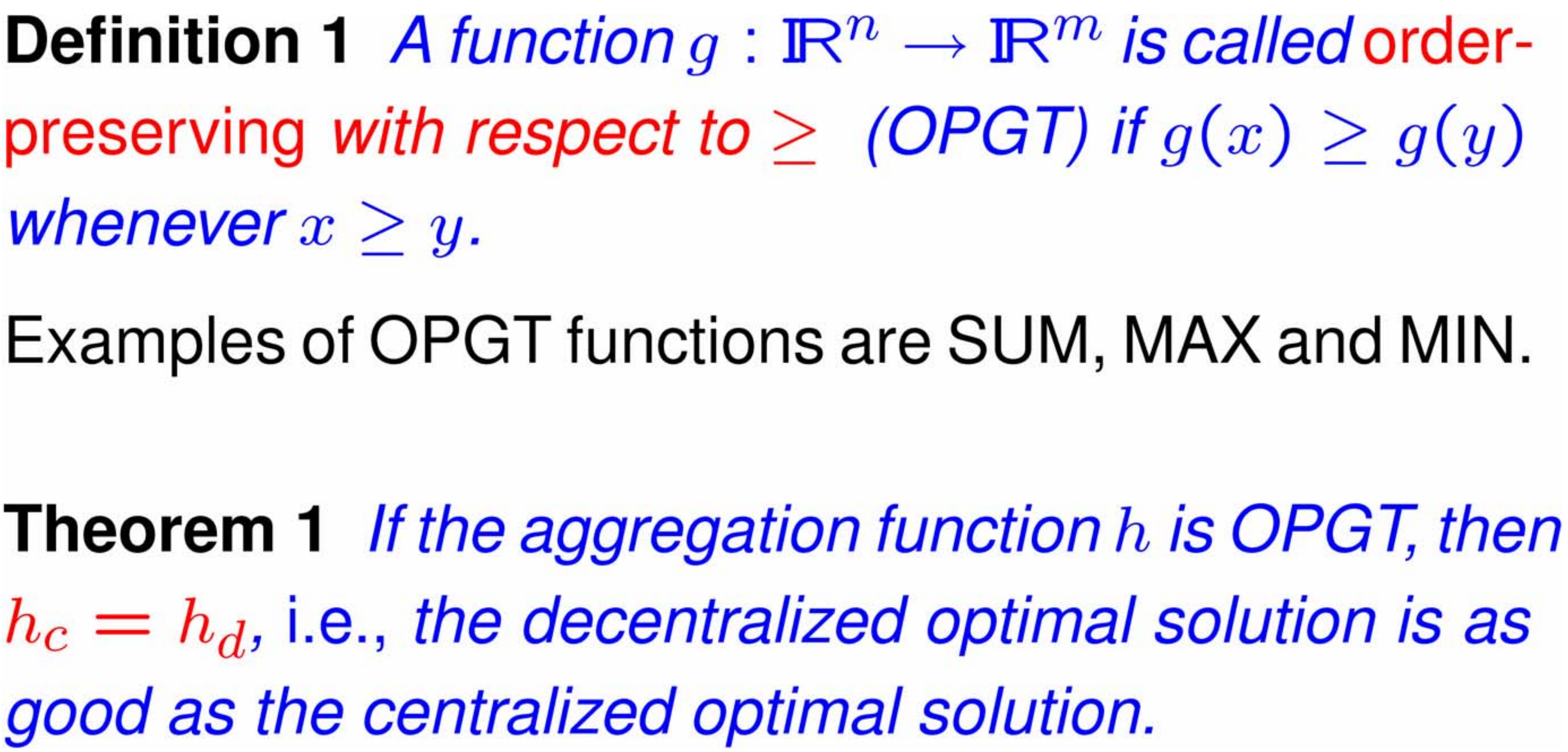
$$h_c \;=\; \min_{x_i, r_i} h(f_1(x_1, r_1, u_1), \ldots, f_n(x_n, r_n, u_n))$$

Decentralized Objective:

$$\begin{aligned} g_i(r_i, u_i) &= \min_{x_i} f_i(x_i, r_i, u_i) \\ h_d &= \min_{r_i} h(g_1(r_1, u_1), \ldots, g_n(r_n, u_n)) \end{aligned}$$

Both subject to $x_i \in C_i(r_i, u_i)$, $(r_1, \ldots, r_n) \in R$.

Decentralized Optimization: Simple Result

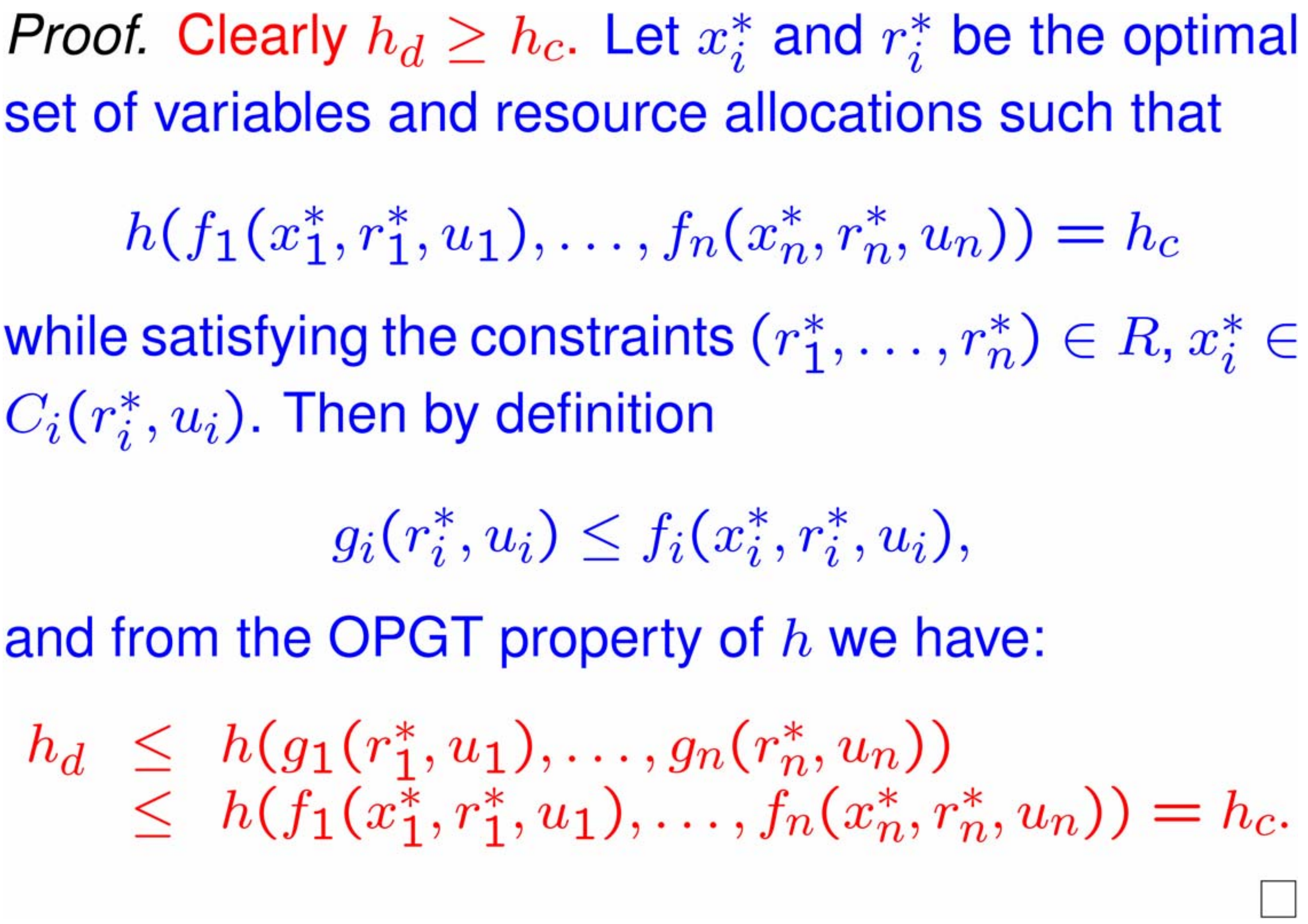


Definition 1 *A function* $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ *is called* order-preserving *with respect to* $\geq$ *(OPGT) if* $g(x) \geq g(y)$ *whenever* $x \geq y$.

Examples of OPGT functions are SUM, MAX and MIN.

Theorem 1 *If the aggregation function* h *is OPGT, then* $h_c = h_d$, i.e., *the decentralized optimal solution is as good as the centralized optimal solution.*

Decentralized Optimization: Simple Result



Proof. Clearly $h_d \geq h_c$. Let x_i^* and r_i^* be the optimal set of variables and resource allocations such that

$$h(f_1(x_1^*, r_1^*, u_1), \ldots, f_n(x_n^*, r_n^*, u_n)) = h_c$$

while satisfying the constraints $(r_1^*, \ldots, r_n^*) \in R$, $x_i^* \in C_i(r_i^*, u_i)$. Then by definition

$$g_i(r_i^*, u_i) \leq f_i(x_i^*, r_i^*, u_i),$$

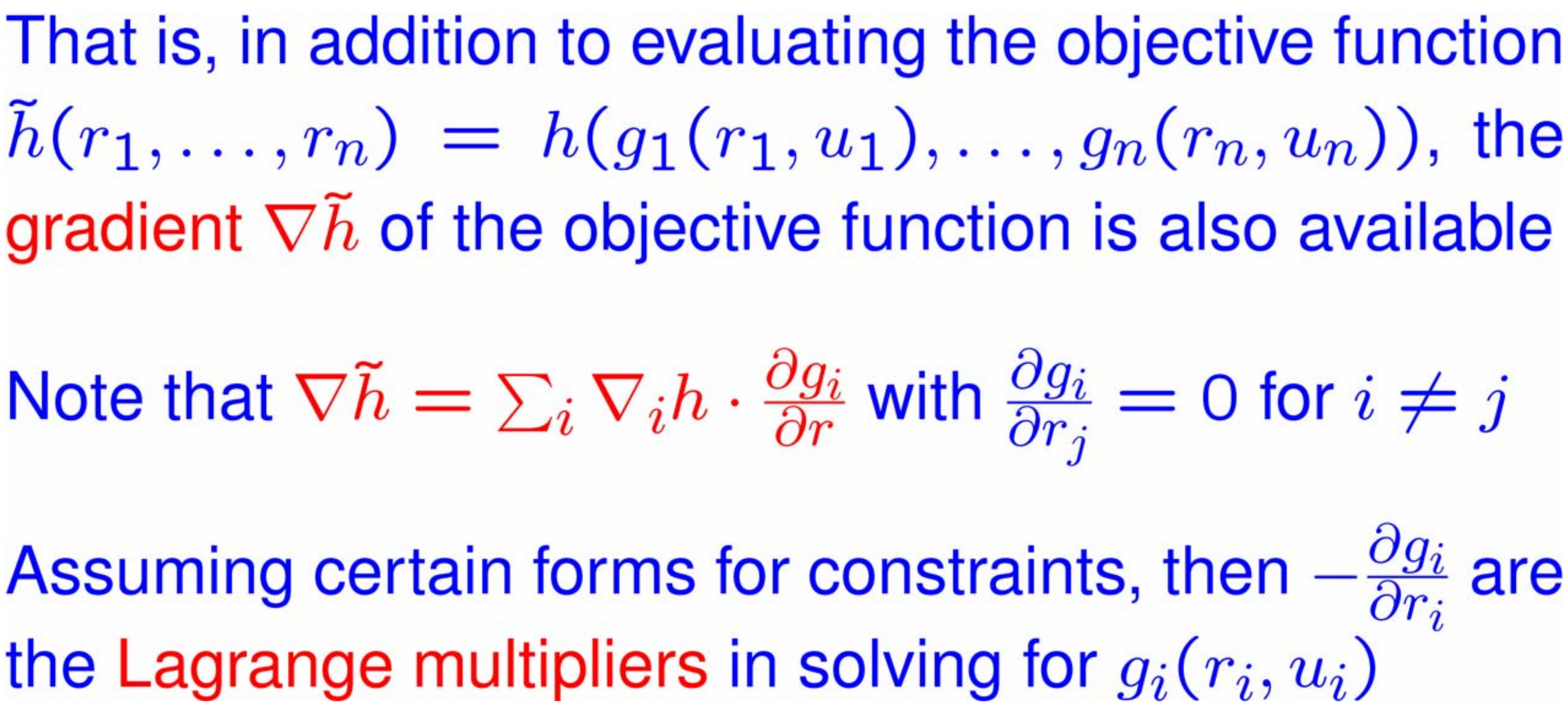
and from the OPGT property of h we have:

$$\begin{aligned} h_d &\leq h(g_1(r_1^*, u_1), \ldots, g_n(r_n^*, u_n)) \\ &\leq h(f_1(x_1^*, r_1^*, u_1), \ldots, f_n(x_n^*, r_n^*, u_n)) = h_c. \end{aligned}$$

□

Decentralized Optimization: Hierarchical Algorithm

- Continuous optimization algorithms generally perform much better if in addition to evaluating objective function, the gradient of objective function is also available



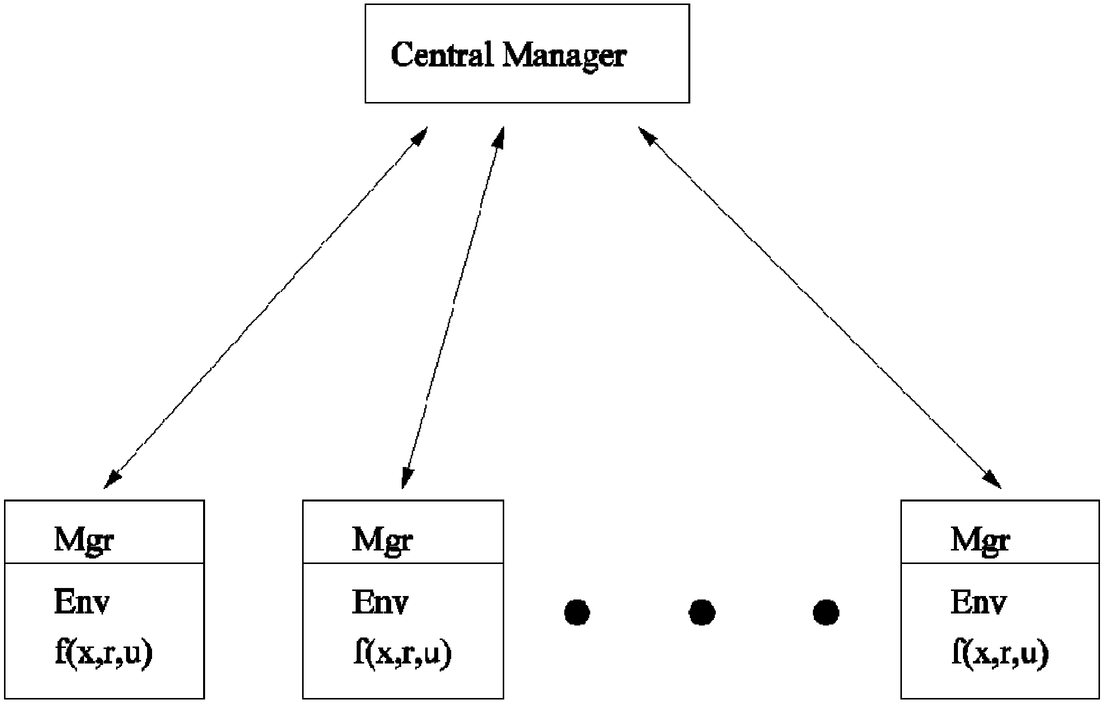
That is, in addition to evaluating the objective function $\tilde{h}(r_1,\ldots,r_n) \;=\; h(g_1(r_1,u_1),\ldots,g_n(r_n,u_n))$, the gradient $\nabla\tilde{h}$ of the objective function is also available

Note that $\nabla\tilde{h} = \sum_i \nabla_i h \cdot \frac{\partial g_i}{\partial r}$ with $\frac{\partial g_i}{\partial r_j} = 0$ for $i \neq j$

Assuming certain forms for constraints, then $-\frac{\partial g_i}{\partial r_i}$ are the Lagrange multipliers in solving for $g_i(r_i,u_i)$

Decentralized Optimization: Hierarchical Algorithm

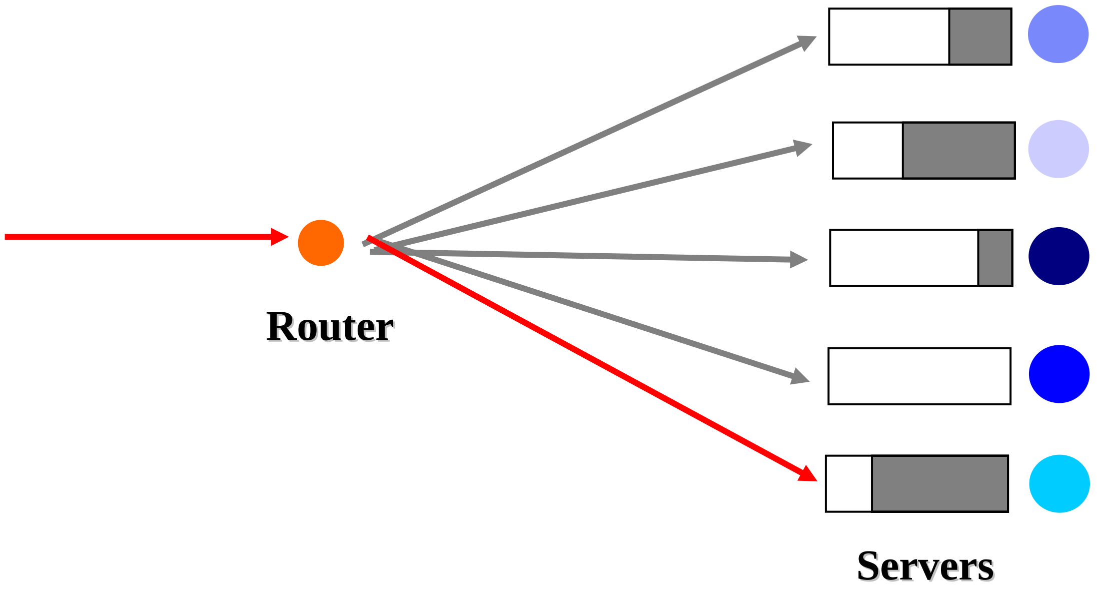
- Efficient (logical) hierarchical scheme between the CM and the Envs
  - CM determines $(r_1,\ldots,r_n)$ and sends r_i to each E_i
  - Each E_i computes and sends $g_i(r_i,u_i)$ to CM along with additional information:
    - Corresponding Lagrange multipliers
    - Trust region radius and model function used in computing $g_i(r_i,u_i)$
  - CM uses this information to compute the objective function and find next $(r_1,\ldots,r_n)$



Decentralized Optimization: Application

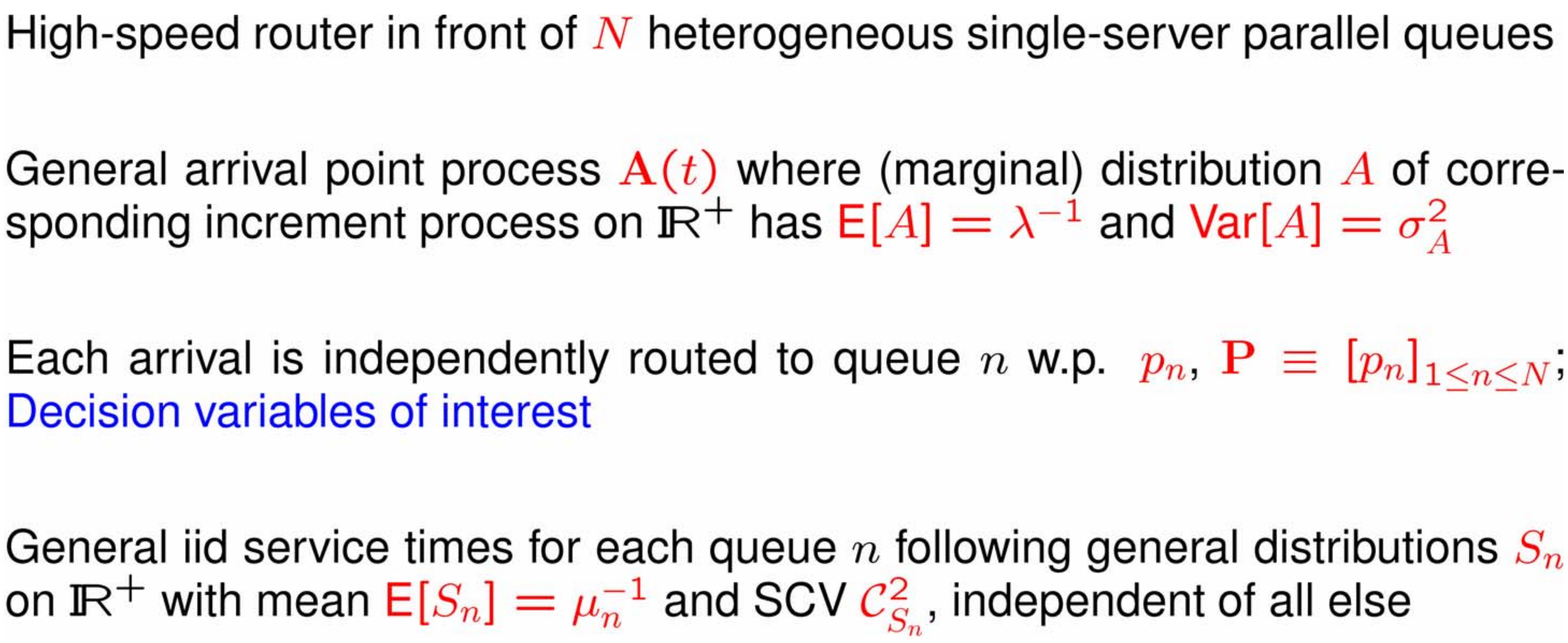
- Consider a representative application consisting of
  - Set of N client environments $E_1,\ldots,E_N$ hosted by common provider on
  - Set of M heterogeneous computing servers $S_1,\ldots,S_M$
  - Set of N routers, one for each E_i
- Decentralized optimization in such an autonomic system includes
  - Allocation of servers among the set of Envs (r_i)
  - Routing of requests among the servers within each Env (x_i)
  - Scheduling of requests at each server within an Env
- SLA defines QoS requirements with revenues and penalties for each Env
  - Focus on typical scenario in which QoS requirements based on response times
- Goal: Minimize global objective function based on the collection of SLAs
  - Simplify presentation by considering SLAs with a single QoS class within each Env

Stochastic-Process Limits: E_i Routing Problem



- Route customers among distributed heterogeneous single-server queues
- Minimize an objective function based on equilibrium sojourn times
- General assumptions for the arrival and service processes
- Customers are routed to distributed queues in a probabilistic manner
- Each single-server queue independently serves customers under FCFS discipline
- Obtain explicit solutions that can be efficiently evaluated in real time
- Static scheduling strategy, but can use in a continual optimization manner

Stochastic-Process Limits: Mathematical Model



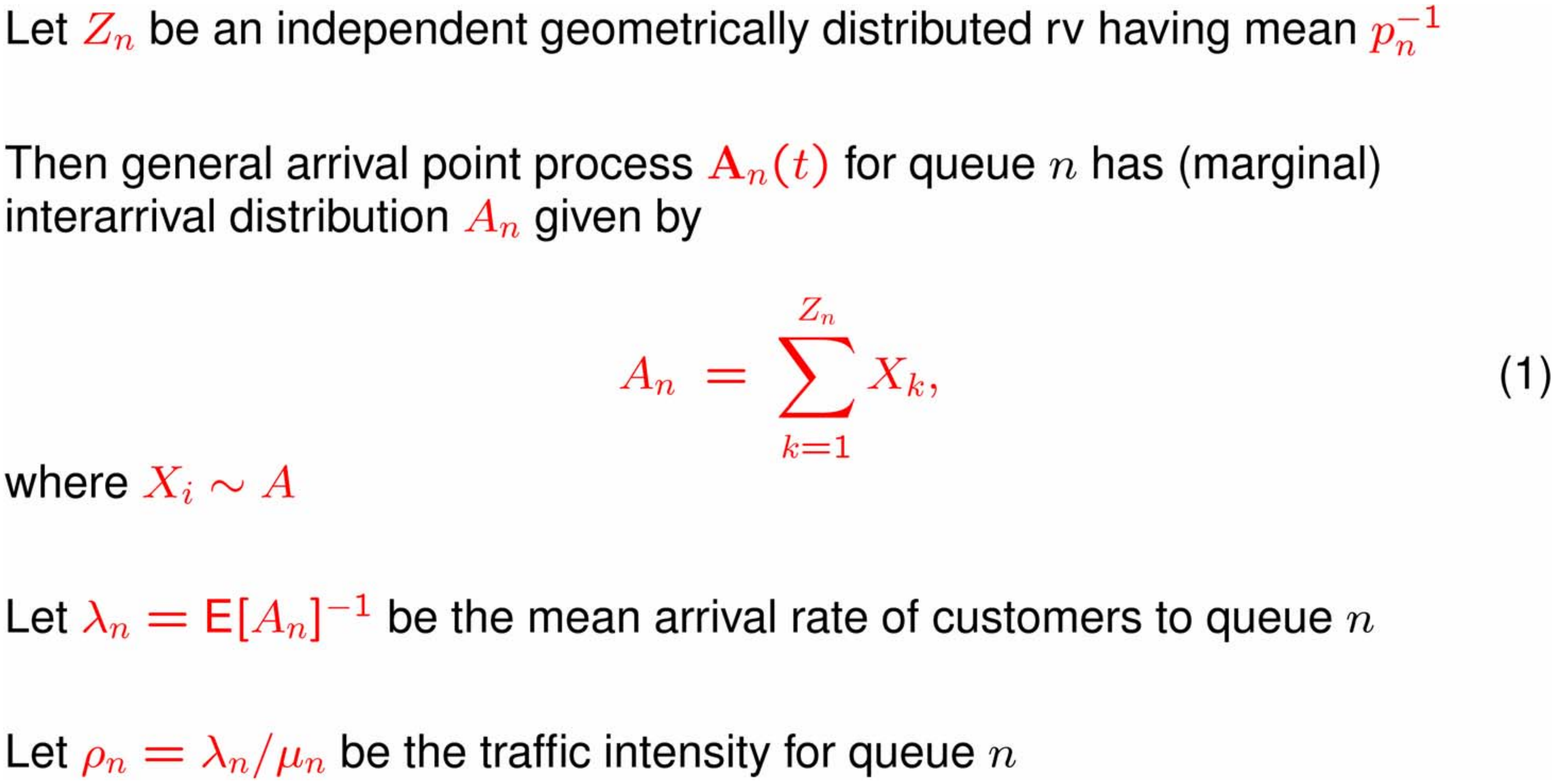
High-speed router in front of N heterogeneous single-server parallel queues

General arrival point process $\mathbf{A}(t)$ where (marginal) distribution A of corresponding increment process on $\mathbb{R}^+$ has $\mathsf{E}[A] = \lambda^{-1}$ and $\mathsf{Var}[A] = \sigma_A^2$

Each arrival is independently routed to queue n w.p. p_n, $\mathbf{P} \equiv [p_n]_{1 \leq n \leq N}$; Decision variables of interest

General iid service times for each queue n following general distributions S_n on $\mathbb{R}^+$ with mean $\mathsf{E}[S_n] = \mu_n^{-1}$ and SCV $\mathcal{C}_{S_n}^2$, independent of all else

Stochastic-Process Limits: Mathematical Model



Let Z_n be an independent geometrically distributed rv having mean p_n^{-1}

Then general arrival point process $\mathbf{A}_n(t)$ for queue n has (marginal) interarrival distribution A_n given by

$$A_n = \sum_{k=1}^{Z_n} X_k, \tag{1}$$

where $X_i \sim A$

Let $\lambda_n = \mathsf{E}[A_n]^{-1}$ be the mean arrival rate of customers to queue n

Let $\rho_n = \lambda_n / \mu_n$ be the traffic intensity for queue n

Stochastic-Process Limits: Mathematical Model

Let $\mathsf{E}[\mathcal{T}_n]$ be the equilibrium sojourn time of customers served at queue n

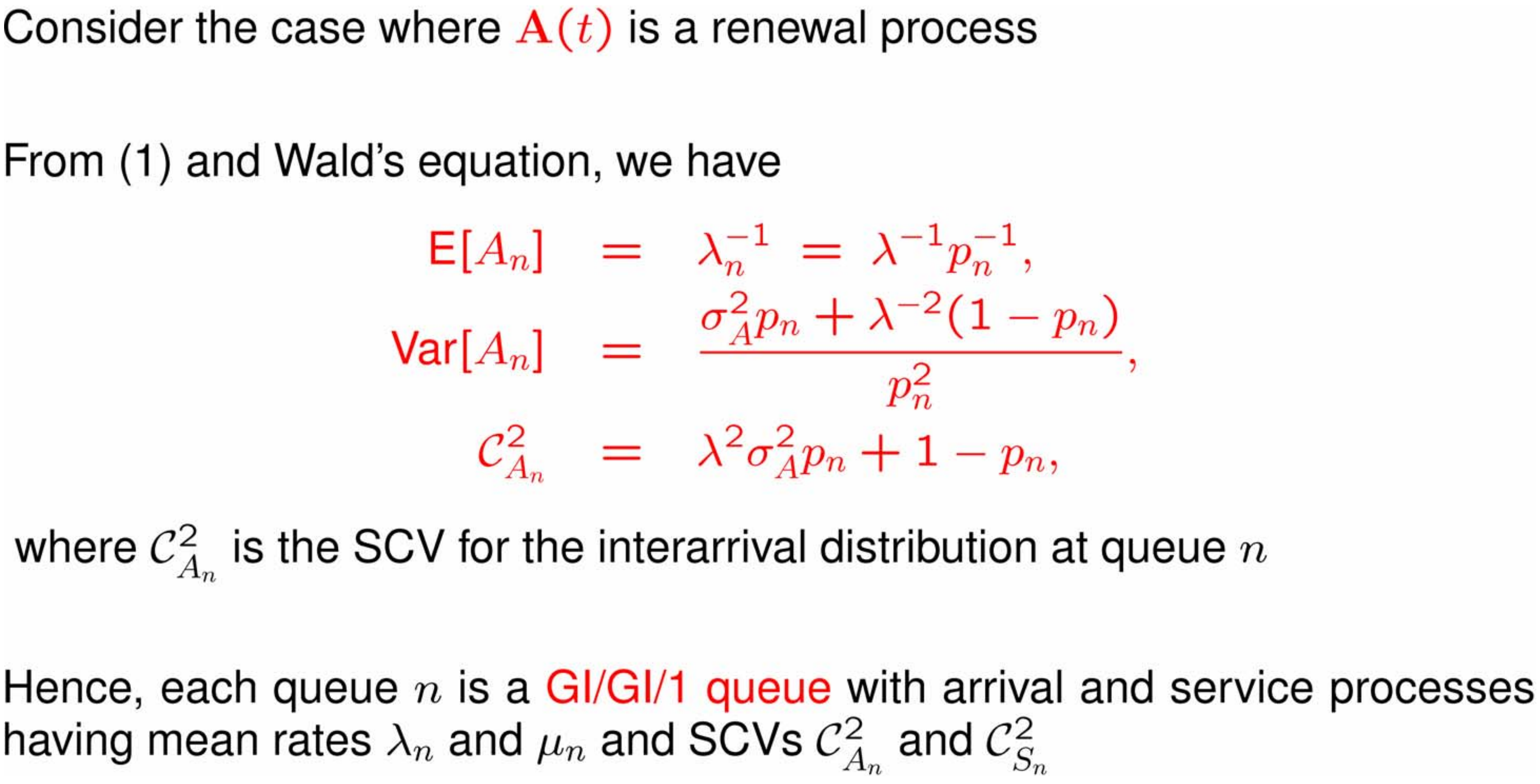
Let $h_n < \infty$ be the holding cost, or weight, per customer per unit time at queue n

(OR1) $$\min \sum_{n=1}^{N} h_n \mathsf{E}[\mathcal{T}_n],$$ [ShanXu97]

(OR2) $$\min \sum_{n=1}^{N} h_n \mathsf{E}[\mathcal{T}_n] p_n,$$ [Borst95,SethSqui98]

$$\text{s.t.} \quad \sum_{n=1}^{N} p_n = 1, \quad p_n \geq 0.$$

Stochastic-Process Limits: Renewal Arrivals



Consider the case where $\mathbf{A}(t)$ is a renewal process

From (1) and Wald's equation, we have

$$\begin{aligned} \mathsf{E}[A_n] &= \lambda_n^{-1} = \lambda^{-1} p_n^{-1}, \\ \mathsf{Var}[A_n] &= \frac{\sigma_A^2 p_n + \lambda^{-2}(1 - p_n)}{p_n^2}, \\ \mathcal{C}_{A_n}^2 &= \lambda^2 \sigma_A^2 p_n + 1 - p_n, \end{aligned}$$

where $\mathcal{C}_{A_n}^2$ is the SCV for the interarrival distribution at queue n

Hence, each queue n is a GI/GI/1 queue with arrival and service processes having mean rates λ_n and μ_n and SCVs $\mathcal{C}_{A_n}^2$ and $\mathcal{C}_{S_n}^2$

Stochastic-Process Limits: Renewal Arrivals

Define

$$\begin{aligned} U_{n,k} &\equiv u_{1,k} + \ldots + u_{n,k}, \quad V_{n,k} \equiv v_{1,k} + \ldots + v_{n,k}, \quad k \geq 1, \\ N_n^U(t) &\equiv \max\{\ell : U_{n,\ell} \leq t,\ \ell \geq 0\}, \quad N_n^V(t) \equiv \max\{\ell : V_{n,\ell} \leq t,\ \ell \geq 0\}, \quad t \geq 0. \end{aligned}$$

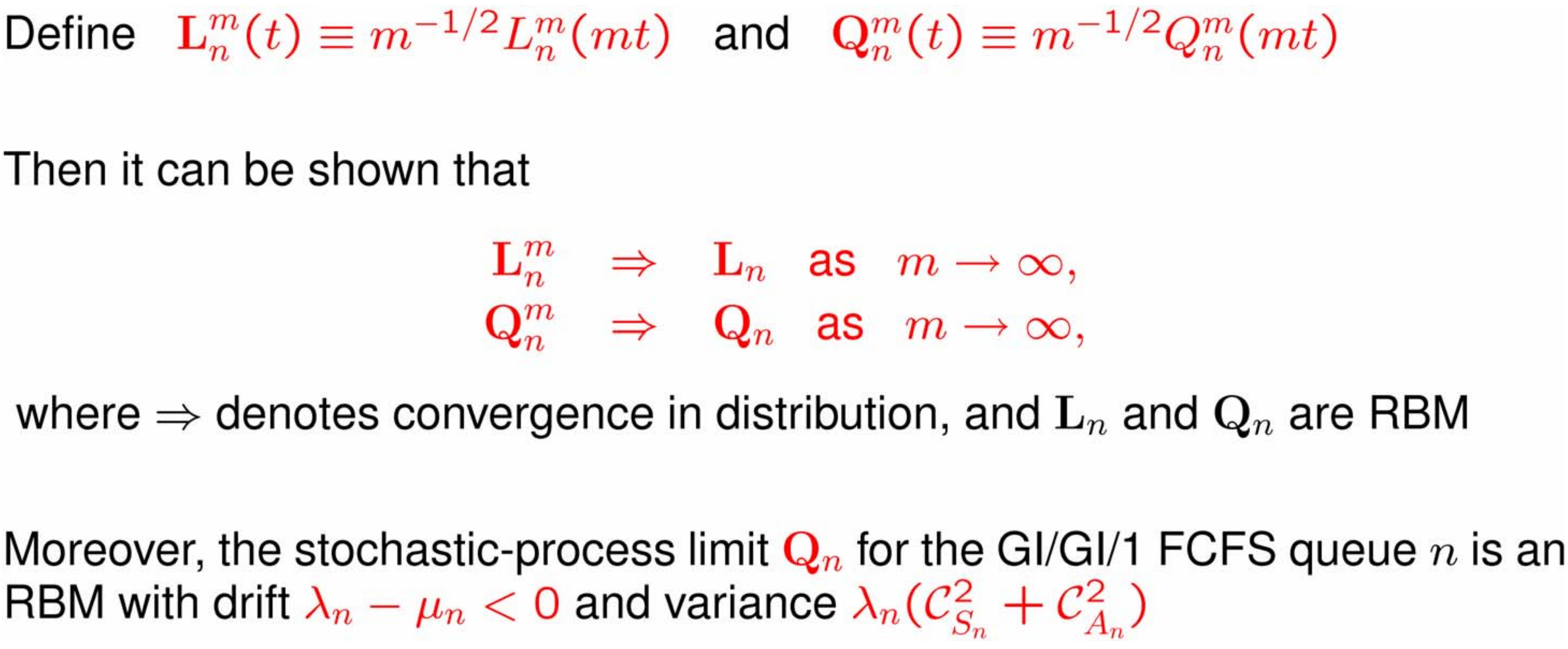
Let $C_n(t) = \sum_{\ell=1}^{N_n^U(t)} V_{n,\ell}$ be the cumulative input process for queue n

Let $X_n(t) = C_n(t) - t$ be the associated net-input process for queue n

Define workload process by $L_n(t) \equiv X_n(t) - \inf\{X(s) \wedge 0 : 0 \leq s \leq t\}$

Define queue length process by $Q_n(t) \equiv N_n^U(t) - N_n^V(C_n(t) - L_n(t))$

Stochastic-Process Limits: Renewal Arrivals



Define $\mathbf{L}_n^m(t) \equiv m^{-1/2} L_n^m(mt)$ and $\mathbf{Q}_n^m(t) \equiv m^{-1/2} Q_n^m(mt)$

Then it can be shown that

$$\begin{aligned} \mathbf{L}_n^m &\Rightarrow \mathbf{L}_n \quad \text{as} \quad m \to \infty, \\ \mathbf{Q}_n^m &\Rightarrow \mathbf{Q}_n \quad \text{as} \quad m \to \infty, \end{aligned}$$

where $\Rightarrow$ denotes convergence in distribution, and $\mathbf{L}_n$ and $\mathbf{Q}_n$ are RBM

Moreover, the stochastic-process limit $\mathbf{Q}_n$ for the GI/GI/1 FCFS queue n is an RBM with drift $\lambda_n - \mu_n < 0$ and variance $\lambda_n(\mathcal{C}^2_{S_n} + \mathcal{C}^2_{A_n})$

Stochastic-Process Limits: Renewal Arrivals

Using this diffusion approximation, we have

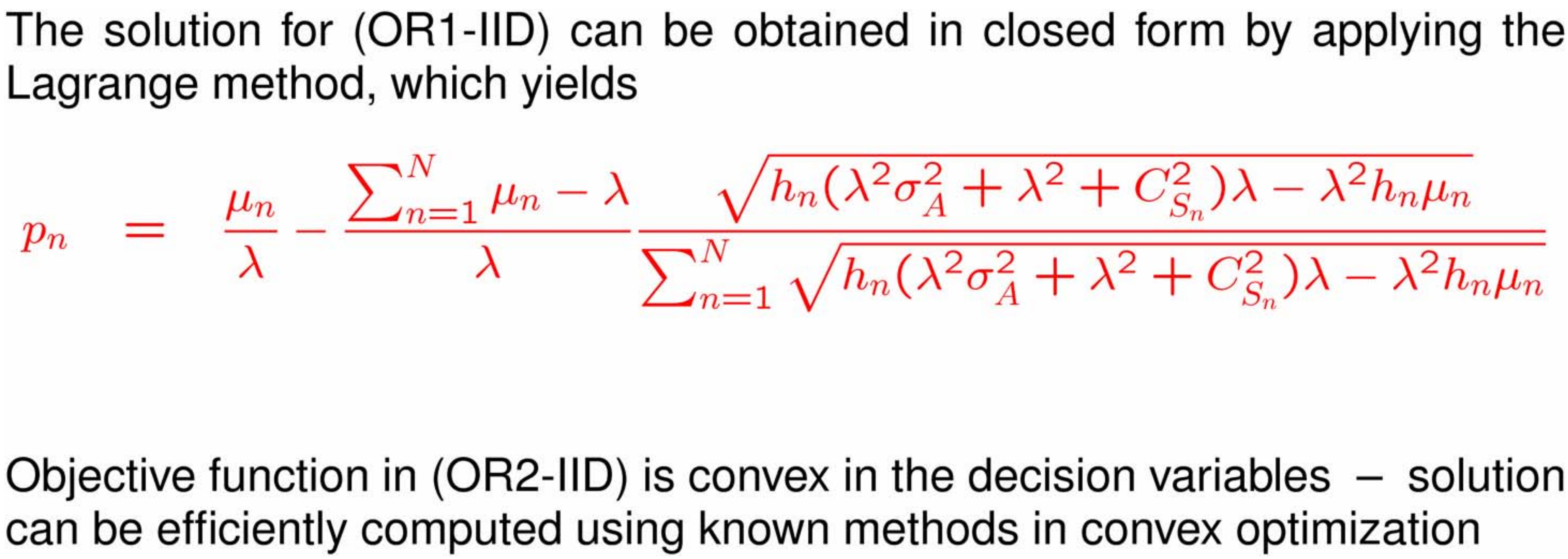
$$
\begin{aligned}
\mathsf{E}[\mathbf{Q}_n] &\approx \rho_n + \frac{\lambda_n(\mathcal{C}^2_{A_n} + \mathcal{C}^2_{S_n})}{2(\mu_n - \lambda_n)}, \\
\mathsf{E}[\mathcal{T}_n] &\approx \frac{1}{\mu_n} + \frac{\lambda^2\sigma^2_A p_n + 1 - p_n + \mathcal{C}^2_{S_n}}{2(\mu_n - \lambda p_n)}.
\end{aligned}
$$

Substituting into (OR1) and (OR2) respectively yields

$$
\min \quad \sum_{n=1}^{N} h_n \left(\frac{1}{\mu_n} + \frac{\lambda^2\sigma^2_A p_n + 1 - p_n + \mathcal{C}^2_{S_n}}{2(\mu_n - \lambda p_n)} \right);
$$

$$
\min \quad \sum_{n=1}^{N} h_n \left(\frac{1}{\mu_n} + \frac{\lambda^2\sigma^2_A p_n + 1 - p_n + \mathcal{C}^2_{S_n}}{2(\mu_n - \lambda p_n)} \right) p_n.
$$

Stochastic-Process Limits: Renewal Arrivals



The solution for (OR1-IID) can be obtained in closed form by applying the Lagrange method, which yields

$$p_n \quad = \quad \frac{\mu_n}{\lambda} - \frac{\sum_{n=1}^{N} \mu_n - \lambda}{\lambda} \frac{\sqrt{h_n(\lambda^2\sigma_A^2 + \lambda^2 + C_{S_n}^2)\lambda - \lambda^2 h_n \mu_n}}{\sum_{n=1}^{N} \sqrt{h_n(\lambda^2\sigma_A^2 + \lambda^2 + C_{S_n}^2)\lambda - \lambda^2 h_n \mu_n}}$$

Objective function in (OR2-IID) is convex in the decision variables – solution can be efficiently computed using known methods in convex optimization

Stochastic-Process Limits: Variance Bound

Consider for each queue n a generic RBM $\mathbf{R}_n$ having drift $\zeta_n < 0$ and variance ω_n

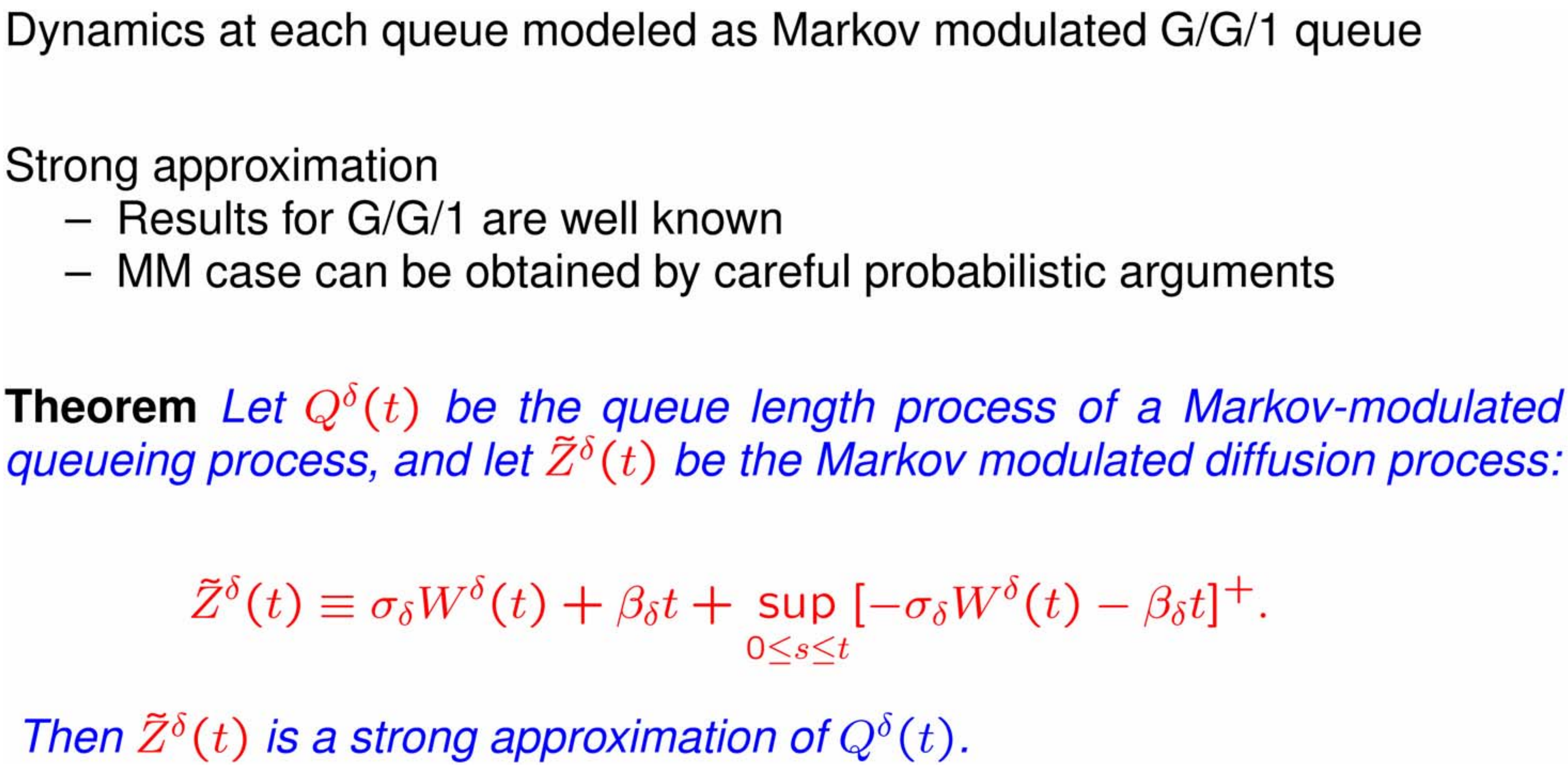
Derive an upper bound on the variance:

$$\text{Var}\left[\sum_{n=1}^{N} p_n \mathbf{R}_n\right] \le 2N \sum_{n=1}^{N} \frac{p_n^2 \omega_n^2}{(-2\zeta_n)^2} - \left(\sum_{n=1}^{N} \frac{p_n \omega_n}{-2\zeta_n}\right)^2$$

Include following side constraint in (all) optimization problems

$$2N \sum_{n=1}^{N} \frac{p_n^2 \omega_n^2}{(-2\zeta_n)^2} - \left(\sum_{n=1}^{N} \frac{p_n \omega_n}{-2\zeta_n}\right)^2 \le \alpha.$$

Stochastic-Process Limits: Correlated Arrivals



Dynamics at each queue modeled as Markov modulated G/G/1 queue

Strong approximation

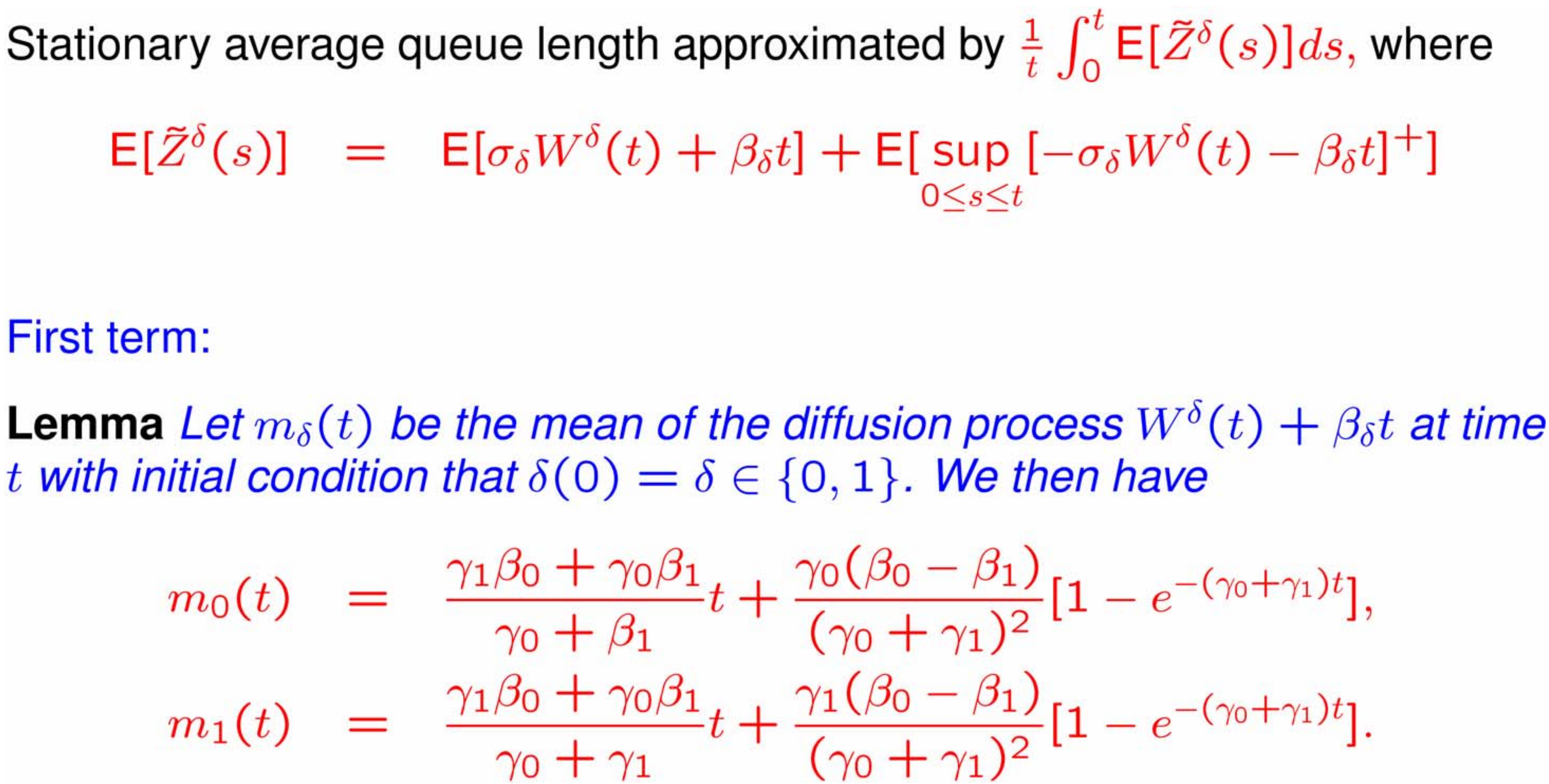
- Results for G/G/1 are well known
- MM case can be obtained by careful probabilistic arguments

Theorem *Let $Q^\delta(t)$ be the queue length process of a Markov-modulated queueing process, and let $\tilde{Z}^\delta(t)$ be the Markov modulated diffusion process:*

$$\tilde{Z}^\delta(t) \equiv \sigma_\delta W^\delta(t) + \beta_\delta t + \sup_{0\leq s\leq t} [-\sigma_\delta W^\delta(t) - \beta_\delta t]^+.$$

Then $\tilde{Z}^\delta(t)$ is a strong approximation of $Q^\delta(t)$.

Stochastic-Process Limits: Correlated Arrivals



Stationary average queue length approximated by $\frac{1}{t}\int_0^t \mathsf{E}[\tilde{Z}^\delta(s)]ds$, where

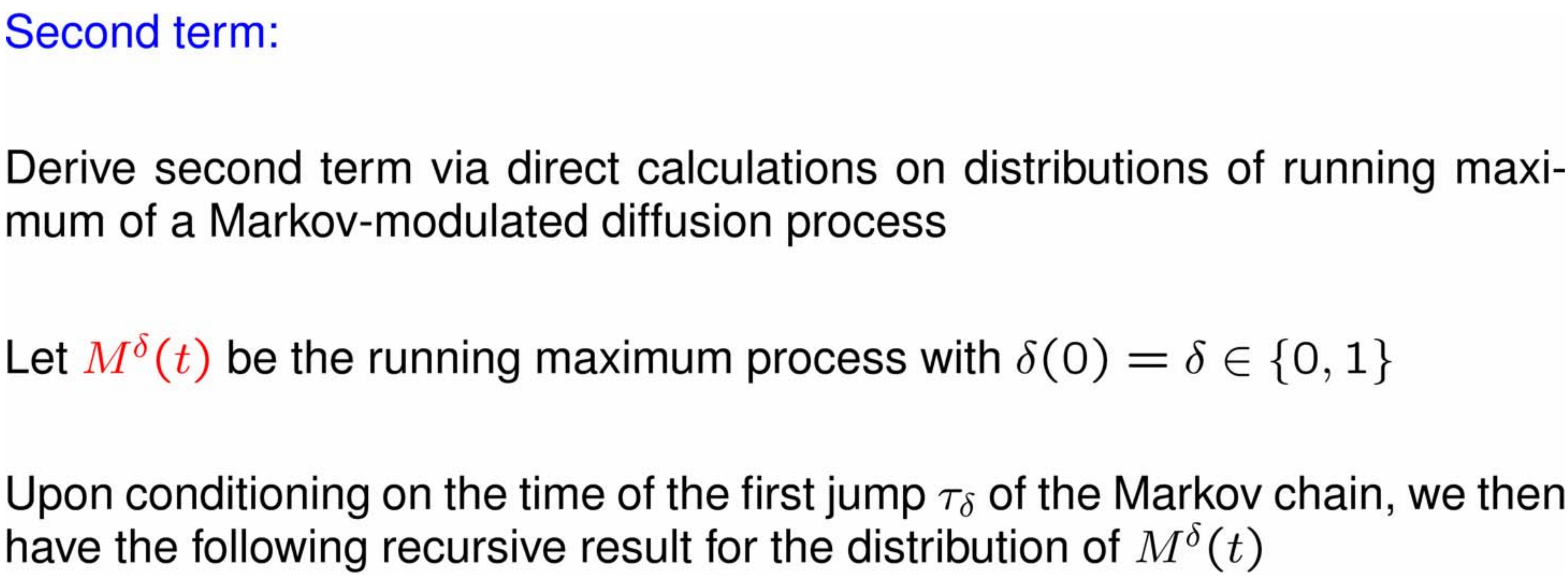
$$\mathsf{E}[\tilde{Z}^\delta(s)] \quad = \quad \mathsf{E}[\sigma_\delta W^\delta(t) + \beta_\delta t] + \mathsf{E}[\sup_{0\le s\le t}[-\sigma_\delta W^\delta(t) - \beta_\delta t]^+]$$

First term:

Lemma *Let* $m_\delta(t)$ *be the mean of the diffusion process* $W^\delta(t) + \beta_\delta t$ *at time* t *with initial condition that* $\delta(0) = \delta \in \{0, 1\}$. *We then have*

$$\begin{aligned} m_0(t) &= \frac{\gamma_1\beta_0 + \gamma_0\beta_1}{\gamma_0 + \beta_1} t + \frac{\gamma_0(\beta_0 - \beta_1)}{(\gamma_0 + \gamma_1)^2}[1 - e^{-(\gamma_0+\gamma_1)t}], \\ m_1(t) &= \frac{\gamma_1\beta_0 + \gamma_0\beta_1}{\gamma_0 + \gamma_1} t + \frac{\gamma_1(\beta_0 - \beta_1)}{(\gamma_0 + \gamma_1)^2}[1 - e^{-(\gamma_0+\gamma_1)t}]. \end{aligned}$$

Stochastic-Process Limits: Correlated Arrivals



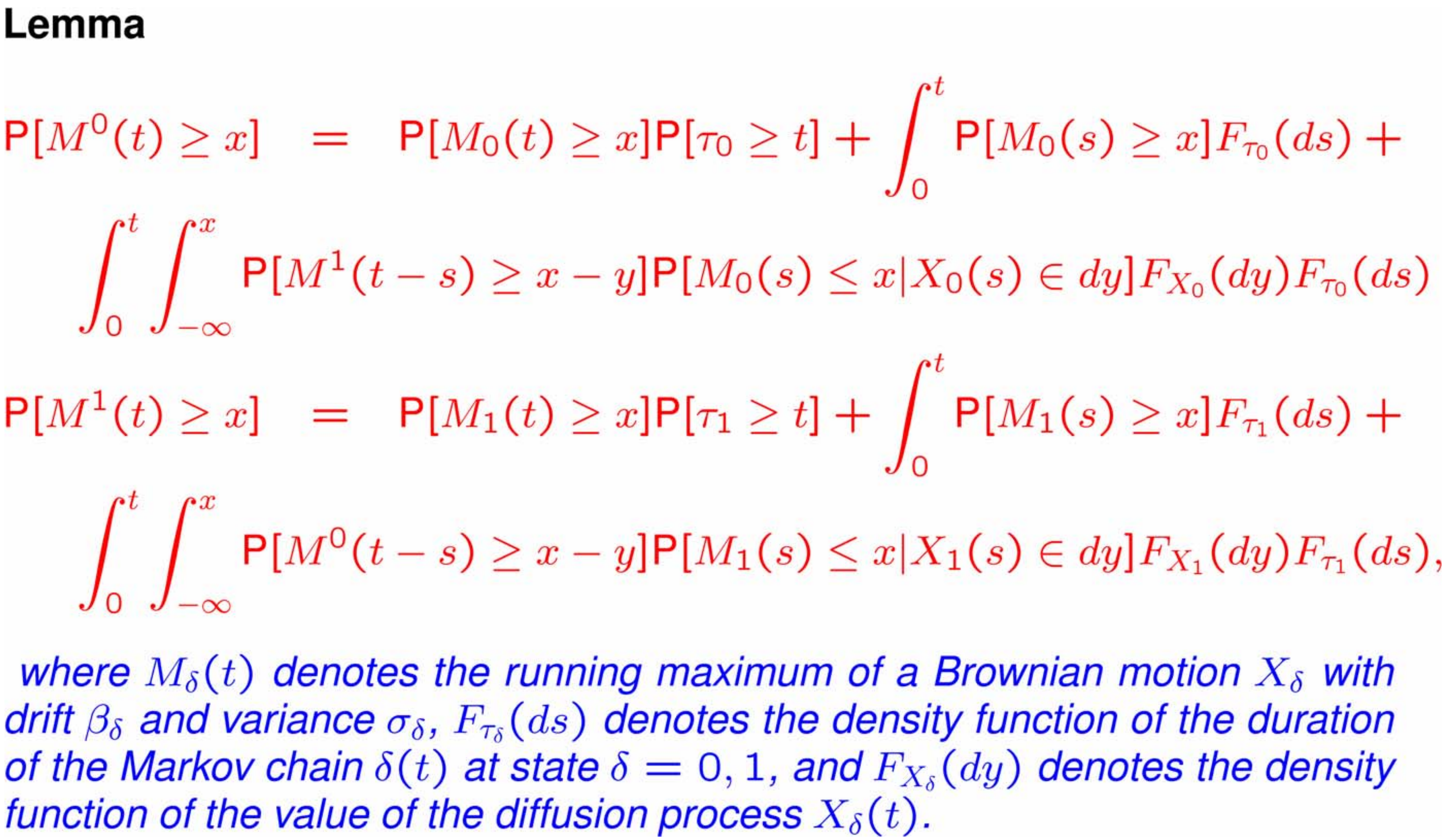
Second term:

Derive second term via direct calculations on distributions of running maximum of a Markov-modulated diffusion process

Let $M^{\delta}(t)$ be the running maximum process with $\delta(0) = \delta \in \{0, 1\}$

Upon conditioning on the time of the first jump τ_{δ} of the Markov chain, we then have the following recursive result for the distribution of $M^{\delta}(t)$

Stochastic-Process Limits: Correlated Arrivals



Lemma

$$
\begin{aligned}
\mathsf{P}[M^0(t) \geq x] \quad = \quad & \mathsf{P}[M_0(t) \geq x]\mathsf{P}[\tau_0 \geq t] + \int_0^t \mathsf{P}[M_0(s) \geq x] F_{\tau_0}(ds) + \\
& \int_0^t \int_{-\infty}^x \mathsf{P}[M^1(t-s) \geq x-y]\mathsf{P}[M_0(s) \leq x | X_0(s) \in dy] F_{X_0}(dy) F_{\tau_0}(ds) \\
\mathsf{P}[M^1(t) \geq x] \quad = \quad & \mathsf{P}[M_1(t) \geq x]\mathsf{P}[\tau_1 \geq t] + \int_0^t \mathsf{P}[M_1(s) \geq x] F_{\tau_1}(ds) + \\
& \int_0^t \int_{-\infty}^x \mathsf{P}[M^0(t-s) \geq x-y]\mathsf{P}[M_1(s) \leq x | X_1(s) \in dy] F_{X_1}(dy) F_{\tau_1}(ds),
\end{aligned}
$$

where $M_\delta(t)$ denotes the running maximum of a Brownian motion X_δ with drift β_δ and variance σ_δ, $F_{\tau_\delta}(ds)$ denotes the density function of the duration of the Markov chain $\delta(t)$ at state $\delta = 0, 1$, and $F_{X_\delta}(dy)$ denotes the density function of the value of the diffusion process $X_\delta(t)$.

Stochastic-Process Limits: Correlated Arrivals

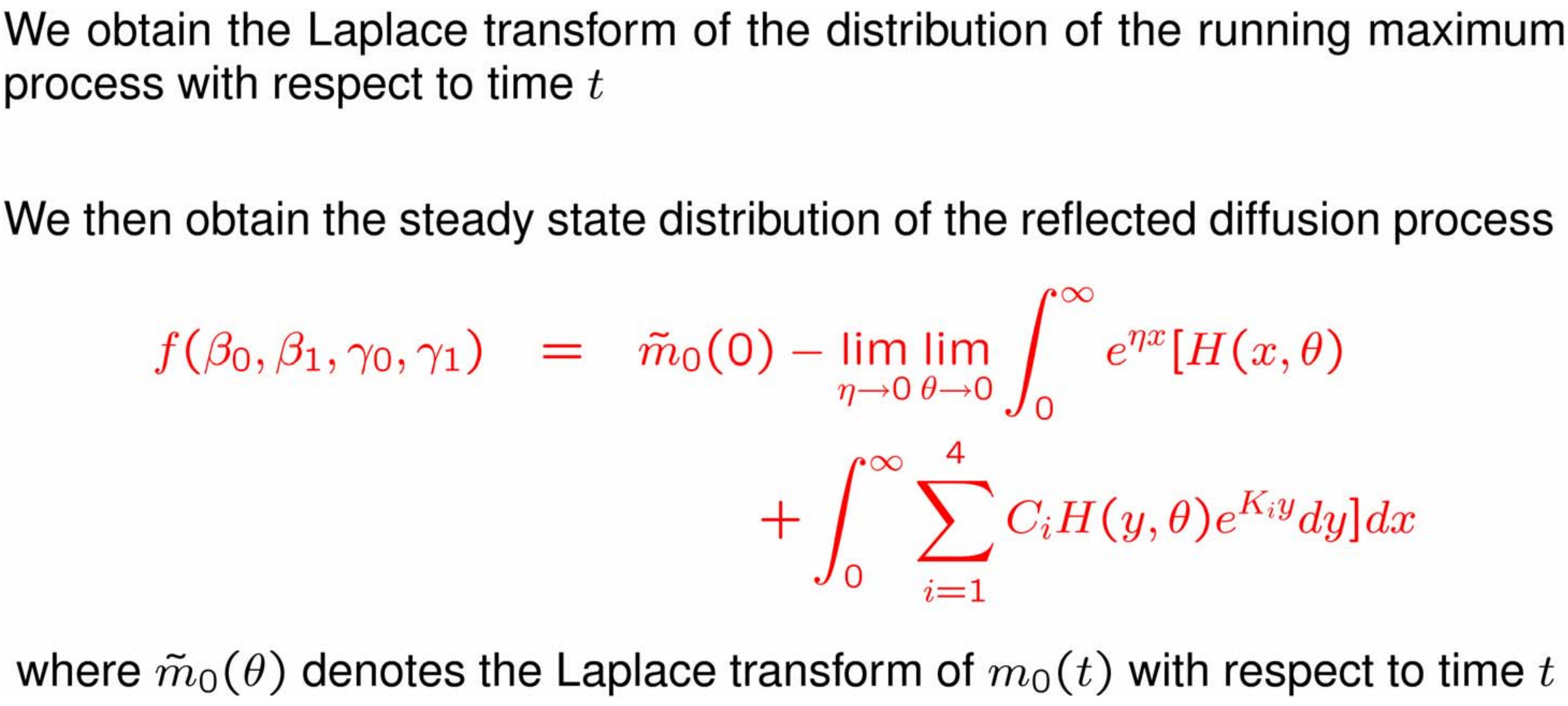
Upon taking the Laplace transform on both sides of the equations in lemma and expressing $\int_0^\infty P(M^\delta(t) \geq x)e^{-\theta t}dt = \tilde{G}^\delta(x, \theta)$, $\theta > 0$, we obtain

Theorem

$$\begin{aligned}
\tilde{G}^0(x, \theta) &= H_1(x, \theta) + \sum_{i=1}^{4} \int_0^\infty C_i \tilde{G}^0(y, \theta) e^{K_i y} dy \\
\tilde{G}^1(x, \theta) &= H_2(x, \theta) + \sum_{i=1}^{4} \int_0^\infty D_i \tilde{G}^1(y, \theta) e^{L_i y} dy
\end{aligned}$$

where . . .

Stochastic-Process Limits: Correlated Arrivals



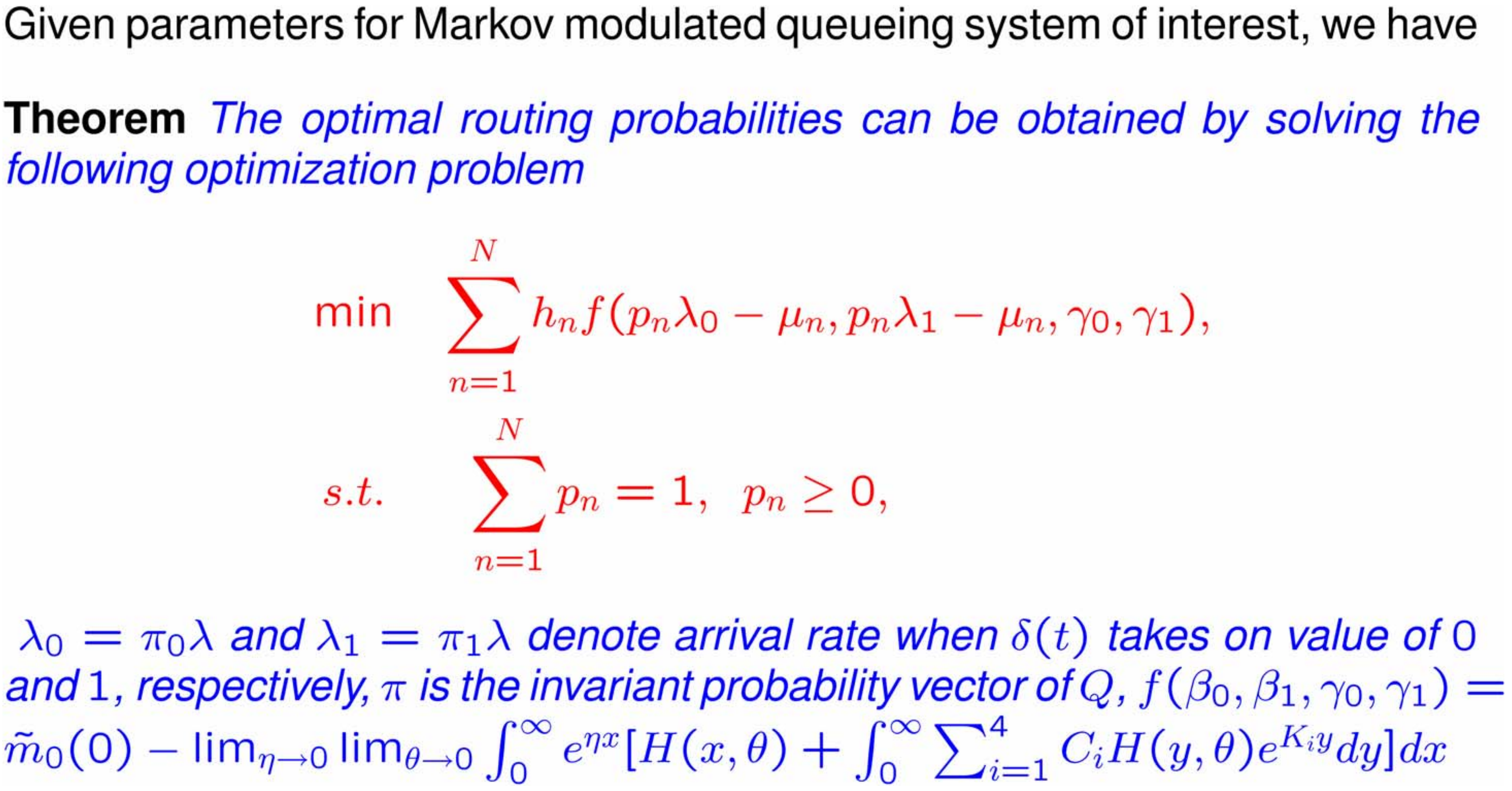
We obtain the Laplace transform of the distribution of the running maximum process with respect to time t

We then obtain the steady state distribution of the reflected diffusion process

$$\begin{aligned} f(\beta_0, \beta_1, \gamma_0, \gamma_1) \quad = \quad & \tilde{m}_0(0) - \lim_{\eta \to 0} \lim_{\theta \to 0} \int_0^\infty e^{\eta x} [H(x, \theta) \\ & + \int_0^\infty \sum_{i=1}^{4} C_i H(y, \theta) e^{K_i y} dy] dx \end{aligned}$$

where $\tilde{m}_0(\theta)$ denotes the Laplace transform of $m_0(t)$ with respect to time t

Stochastic-Process Limits: Correlated Arrivals



Given parameters for Markov modulated queueing system of interest, we have

Theorem *The optimal routing probabilities can be obtained by solving the following optimization problem*

$$\min \quad \sum_{n=1}^{N} h_n f(p_n\lambda_0 - \mu_n, p_n\lambda_1 - \mu_n, \gamma_0, \gamma_1),$$

$$s.t. \quad \sum_{n=1}^{N} p_n = 1, \;\; p_n \geq 0,$$

$\lambda_0 = \pi_0\lambda$ *and* $\lambda_1 = \pi_1\lambda$ *denote arrival rate when* $\delta(t)$ *takes on value of* 0 *and* 1, *respectively,* π *is the invariant probability vector of* Q, $f(\beta_0, \beta_1, \gamma_0, \gamma_1) = \tilde{m}_0(0) - \lim_{\eta\to 0}\lim_{\theta\to 0}\int_0^\infty e^{\eta x}[H(x,\theta) + \int_0^\infty \sum_{i=1}^4 C_i H(y,\theta)e^{K_i y}dy]dx$

Stochastic-Process Limits: Correlated Arrivals

- We have extended analysis to establish corresponding weak convergence and strong approximation results for the semi-Markov modulated case

- System modulated by chains with general state space
  - The Laplace transform can then be obtained by solving a differential-functional equation, extending scheme developed by M. Jacobsen

- Then the corresponding optimization problem can be efficiently calculated based on these results

System Dynamics: Application

- Workloads u_i can be modeled as stochastic processes that vary over time
- Given nonstationary behavior, allocation decisions made periodically at time t_k
  - Time scale depends upon the delays, overheads and constraints involved in changing variables, the QoS requirements, the properties of underlying stochastic processes
- Decentralized optimization problem solved at each scheduling epoch t_k
  - Based on measurements collected during scheduling intervals $\tau_j=[t_{j+1},t_j)$, $j=0,\ldots,k-1$
  - Determine optimal variables x_i^*, r_i^* to be deployed during next scheduling interval τ_k
  - Assume intervals τ_k are sufficiently long for each Env to reach steady state within τ_k
- Focus on typical scenario in which QoS requirements based on response times
  - $f_i(x_i,r_i,u_i) = f_i(\ \mathbf{E}[T_i(x_i,r_i,u_i)]\ )$, where $\mathbf{E}[T_i(x_i,r_i,u_i)]$ is expected response time for E_i
  - Aggregate cost function is given by weighted sum of $g_i(r_i,u_i)$

System Dynamics: Application

Consider each E_i during any scheduling interval τ_k in which the workload processes u_i are stationary

AM_i: determine optimal routing variable $x_i^* \in C_i(r_i, u_i)$

$$g_i(r_i, u_i) \;=\; \min_{x_i} \sum_{S_j \in r_i} H_j \, f_i(\mathbf{E}[T_i(x_i, S_j, u_i)])$$

CM: determine optimal allocation $(r_1^*, \ldots, r_N^*) \in R$

$$h_d \;=\; \min_{r_i} \sum_{i=1}^{N} \widehat{H}_i \, g_i(r_i, u_i)$$

System Dynamics: Application

$$\mathbf{AM_i}: \; g_i(r_i,u_i) \;=\; \min_{x_i} \sum_{S_j \in r_i} H_j \left(\frac{1}{\mu_{i,j}} + \frac{x_{i,j}\alpha_i + \beta_i}{\mu_{i,j} - \lambda_i x_{i,j})} \right) x_{i,j},$$

$$\text{s.t.} \quad \sum_{S_j \in r_i} x_{i,j} = 1, \;\; x_{i,j} \geq 0, \;\; \lambda x_{i,j} < \mu_{i,j}$$

$$\mathbf{CM}: \; h_d \;=\; \min_{(r_1,\ldots,r_N)} \sum_{i=1}^{N} \widehat{H}_i \min_{x_i} \sum_{S_j \in r_i} H_j \left(\frac{1}{\mu_{i,j}} + \frac{x_{i,j}\alpha_i + \beta_i}{\mu_{i,j} - \lambda_i x_{i,j})} \right) x_{i,j},$$

$$\text{s.t.} \quad \sum_{S_j \in r_i} x_{i,j} = 1, \;\; x_{i,j} \geq 0, \;\; \lambda x_{i,j} < \mu_{i,j}$$

where $\alpha_i = (\mathcal{C}^2_{A_i} - 1)/2$ and $\beta_i = (\mathcal{C}^2_{B_i} + 1)/2$

System Dynamics: Application

- When $\lambda_i\ x_{i,j}$ ¸ $\mu_{i,j}$, the response time process for E_i on server S_j 2 r_i blows up
  - Value of $ET_{i,j}$ within interval τ_k increases with length of τ_k s.t. $ET_{i,j}$! 1 as τ_k ! 1

- Time delays can cause this situation to occur as we will demonstrate

- The smaller the length of τ_k, the smaller the explosion in value of **E**$T_{i,j}$ during τ_k

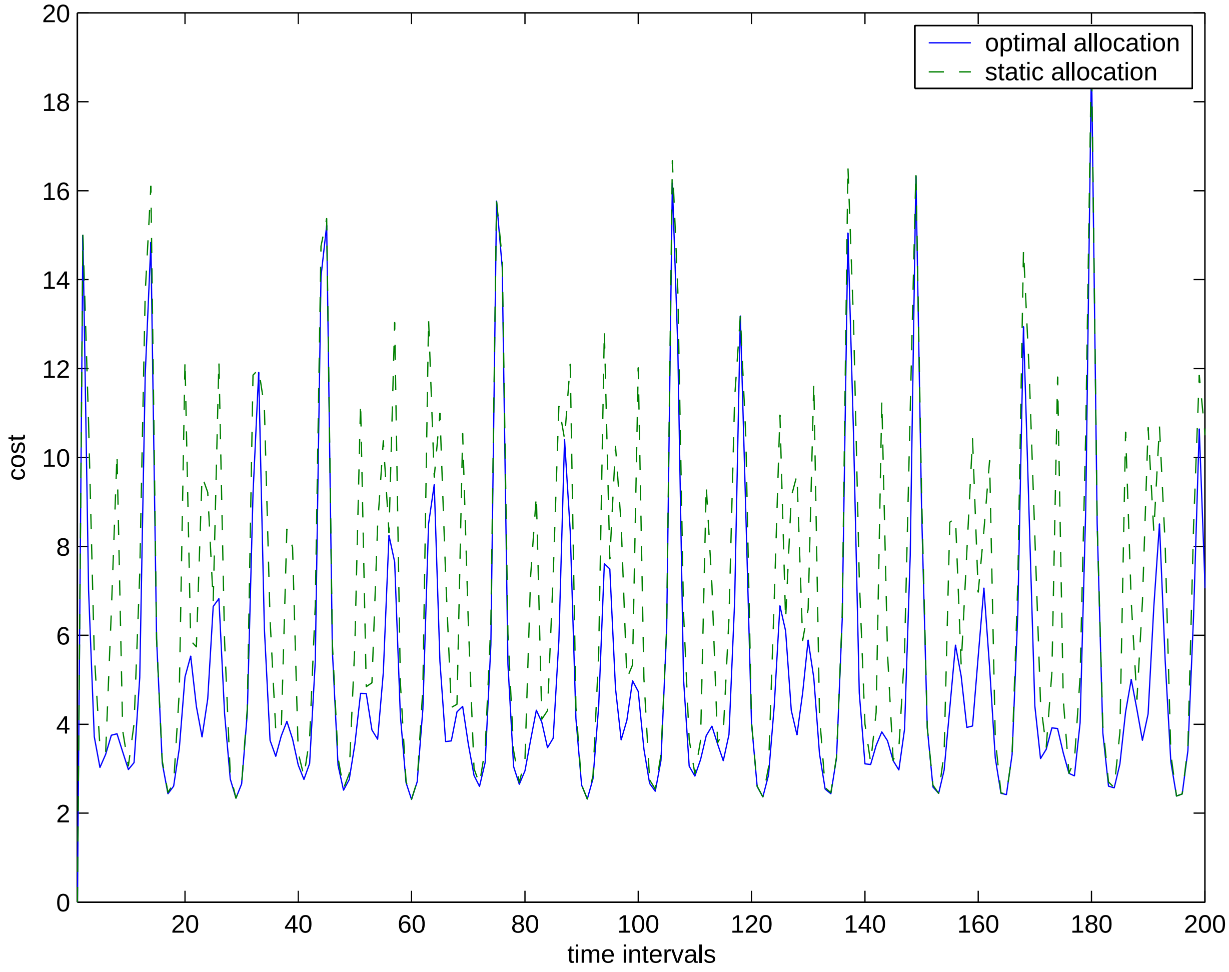
- The smaller the length of τ_k, the larger the delay in the dynamical system (due to fairly consistent overheads and communication delays)

- The smaller the length of τ_k, the more likely it is that a backlog of customers from interval τ_k are not served within this interval and spill over into intervals τ_{k+m}

- Consider numerical experiments with our results to illustrate and quantify some of these issues for simplified case where all holding costs are 1

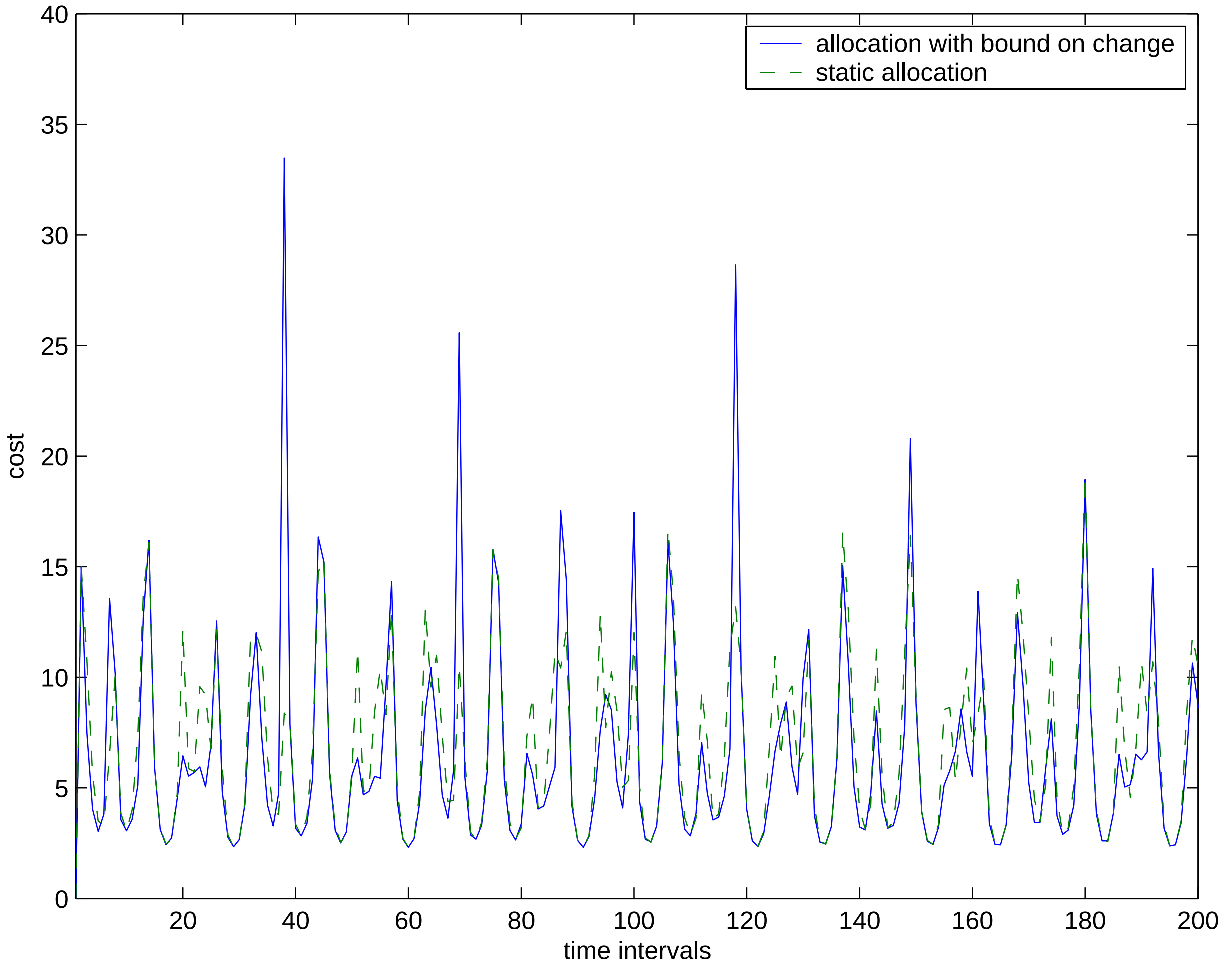
- The two time-varying arrival rates are modeled as sinusoidal functions of time

System Dynamics: Numerical Results



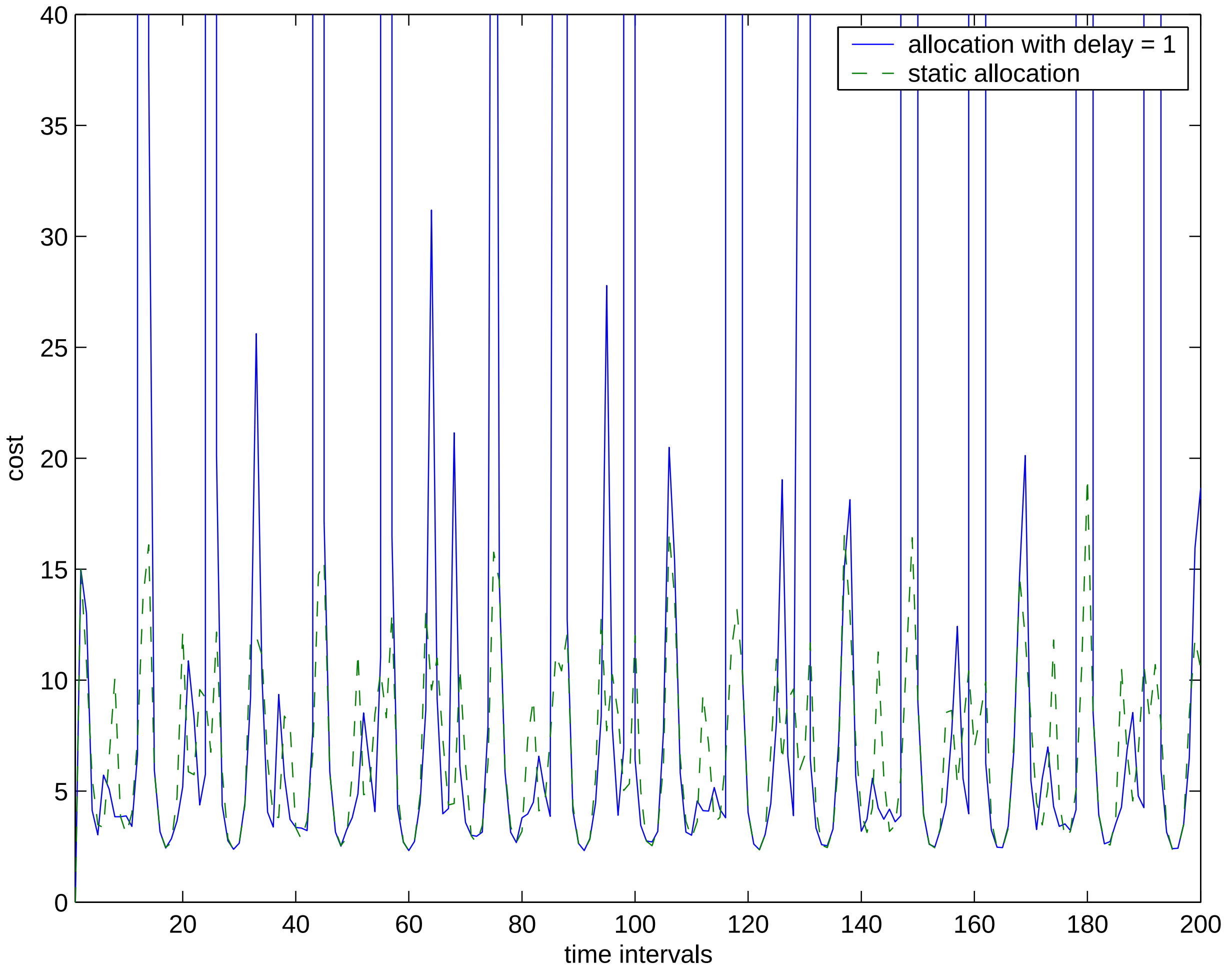
Total response time for optimal allocation and static allocation, when no delay

System Dynamics: Numerical Results



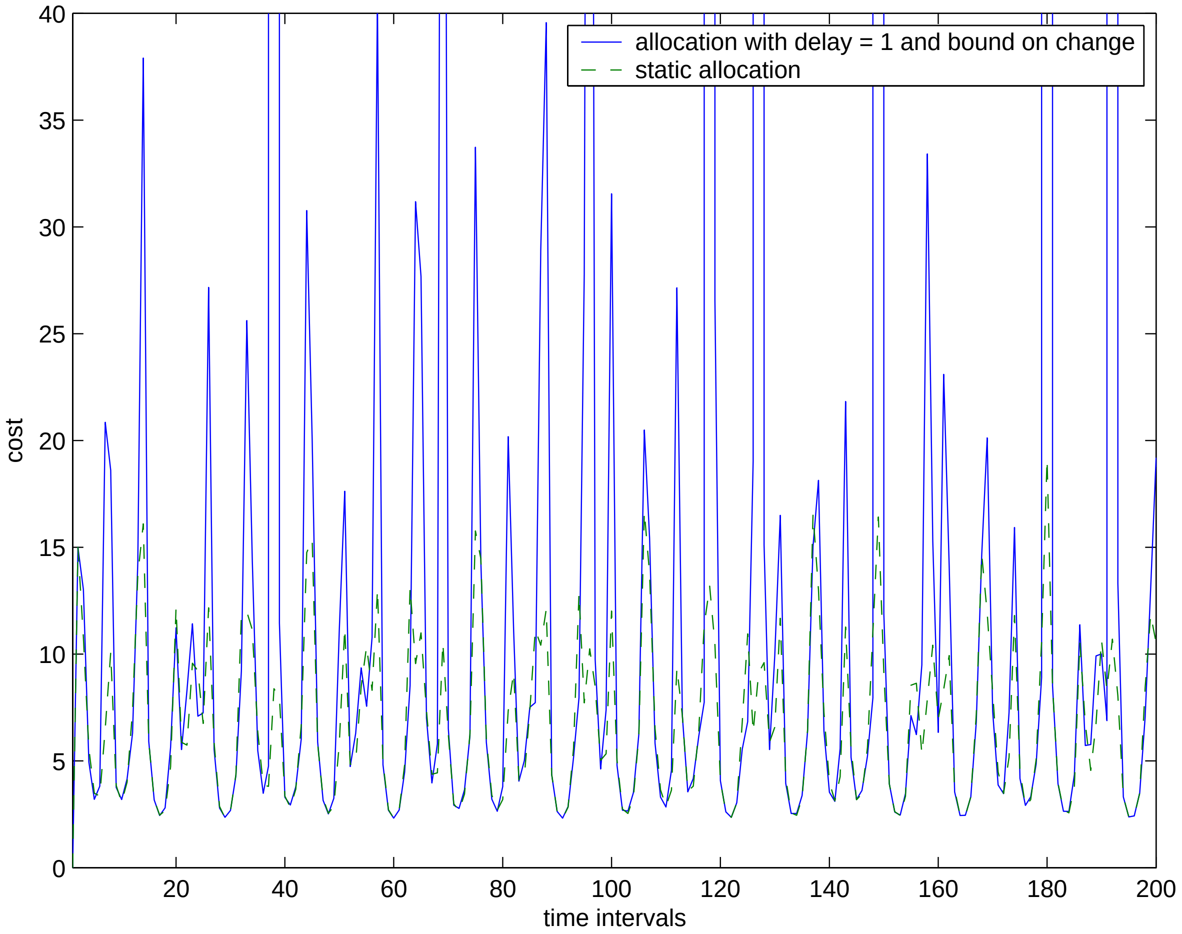
Total response time for static allocation and allocation w/ bound on change, when no delay

System Dynamics: Numerical Example



Total response time for optimal allocation and static allocation, when delay = 1

System Dynamics: Numerical Example

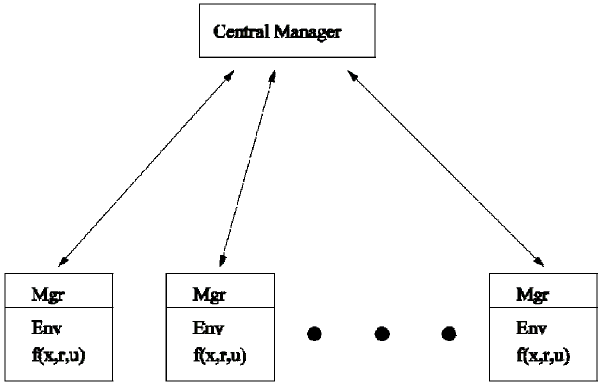


Total response time for static allocation and allocation w/ bound on change, when delay=1

General Overview

- Decentralized Optimization
  - Conditions for same quality of solution under decentralized as centralized
  - Hierarchical algorithmic issues
  - Representative application: Central Manager (CM) with multiple Envs
- Stochastic-Process Limits
  - Optimal routing problem for each Env (E_i) under renewal arrivals
  - Optimal routing problem for each Env (E_i) under correlated arrivals
- System Dynamics
  - Dynamics of optimal solutions for both

CM and E_i under time-varying workloads


Central Manager
Mgr
Env
f(x,r,u)
Mgr
Env
f(x,r,u)
Mgr
Env
f(x,r,u)