Evaluation Relaxation as an Efficiency-Enhancement Technique: Handling Variance and Bias

Kumara Sastry, and David E. Goldberg

{ksastry,deg}@uiuc.edu

Illinois Genetic Algorithms Laboratory
University of Illinois at Urbana-Champaign
Urbana, IL 61801
http://www-illigal.ge.uiuc.edu

Genetic and Evolutionary Computation COnference 2002

July 9–13, 2002. New York, NY

Background

- *Competent GAs* (Goldberg, 1999)
  - Solve hard problems quickly, reliably, and accurately
  - Subquadratic number of function evaluations
- Costly function evaluations
  - Subquadratic function evaluations can be high
  - $\ell = 100$, Time/Eval. $= 1$ s. **Need 2.5 hours**
- *Efficiency-enhancement* techniques (EETs)
  - *Evaluation relaxation* is one such EET

Motivation

- *Evaluation Relaxation*
  - Replace accurate, but costly function evaluation
  - Use approximate, but cheap function evaluation
  - Approximation introduces error
- Practical guidelines are lacking
  - Which, When and How?
  - What's the speed-up?

Overview

- Evaluation relaxation
- Decomposition of evaluation relaxation
- Problem statement
- Key assumptions
- Part I: Handling Variance
- Part II: Handling Bias
- Conclusions

Evaluation Relaxation

- Use approximate, but low-cost fitness function
- Approximation introduces error
- Previous work:
  - Grefenstette & Fitzpatrick (1985)
  - Aizawa & Wah (1993, 1994)
  - Miller & Goldberg (1995, 1996, 1997)
  - Jin, Olhofer, & Sendhoff (2001)
  - Smith, Dike, & Stegmann (1996)
  - Sastry, Goldberg, & Pelikan (2001)
  - Albert & Goldberg (2001, 2002)

Decomposition of Evaluation Relaxation

- Fitness function approximation
  - *Endogenous:* Error introduced by the GA
    - Eg., fitness inheritance
  - *Exogenous:* Error in fitness functions
    - Eg., response surface methodology
- Components of error: (Geman *et al*, 1992)
  - *Variance:* Changes the fitness value
  - *Bias:* Changes the optimal solution

Problem Statement

- Two fitness functions
  - High cost, but low error
  - Low cost, but high error
- Which fitness function to use?
  - Spatial vs. temporal decision making
  - Strategy that yields maximum speed-up
  - Utilize facetwise models
    * Convergence-time and population-sizing models

Assumptions

- Non-overlapping population of fixed size
- Generationwise selectorecombinative GAs
  - Uniform crossover, tournament selection
- Binary encoding and fixed string length
- Stationary fitness functions
- models for OneMax domain
- Tight linkage or usage of a competent GA

Part I: Handling Variance

- Noisy OneMax function
  - OneMax function + Gaussian noise
- Two fitness functions: f_1, and f_2
  - f_1: High cost (c_1), and low noise variance ($\sigma^2_{N_1}$).
  - f_2: Low cost c_2, and high noise variance ($\sigma^2_{N_2}$).

$$c_1 > c_2, \quad \sigma^2_{N_1} < \sigma^2_{N_2}$$

  - Higher the accuracy, costlier the evaluation.
- Averaging eliminates variance effects

Convergence-Time Model - I

- *Selection-intensity* based model (Bulmer, 1980)
  - Mühlenbein & Schlierkamp-Voosen (1993); Thierens & Goldberg (1993); Bäck (1994); Miller & Goldberg (1996)
- Gaussian fitness distribution: $F_i \sim \mathcal{N}\left(\mu_t, \sigma_t^2 + \sigma_{N_i}^2\right)$
- Prop. of correct BBs: (Miller & Goldberg, 1996)

$$p_{t+1} - p_t = \frac{I}{\rho_e \sqrt{\ell}} \sqrt{p_t \left(1 - p_t\right)}$$

  - *Elongation factor*, $\rho_e(t) = \sqrt{1 + \left(\sigma_{N_i}^2 / \sigma_t^2\right)}$
- Exact solution is complex

Convergence-Time Model - II

- Prop. of correct BBs: (Miller & Goldberg, 1996)

$$p_{t+1} - p_t = \frac{I}{\rho_e \sqrt{\ell}} \sqrt{p_t \left(1 - p_t\right)}$$

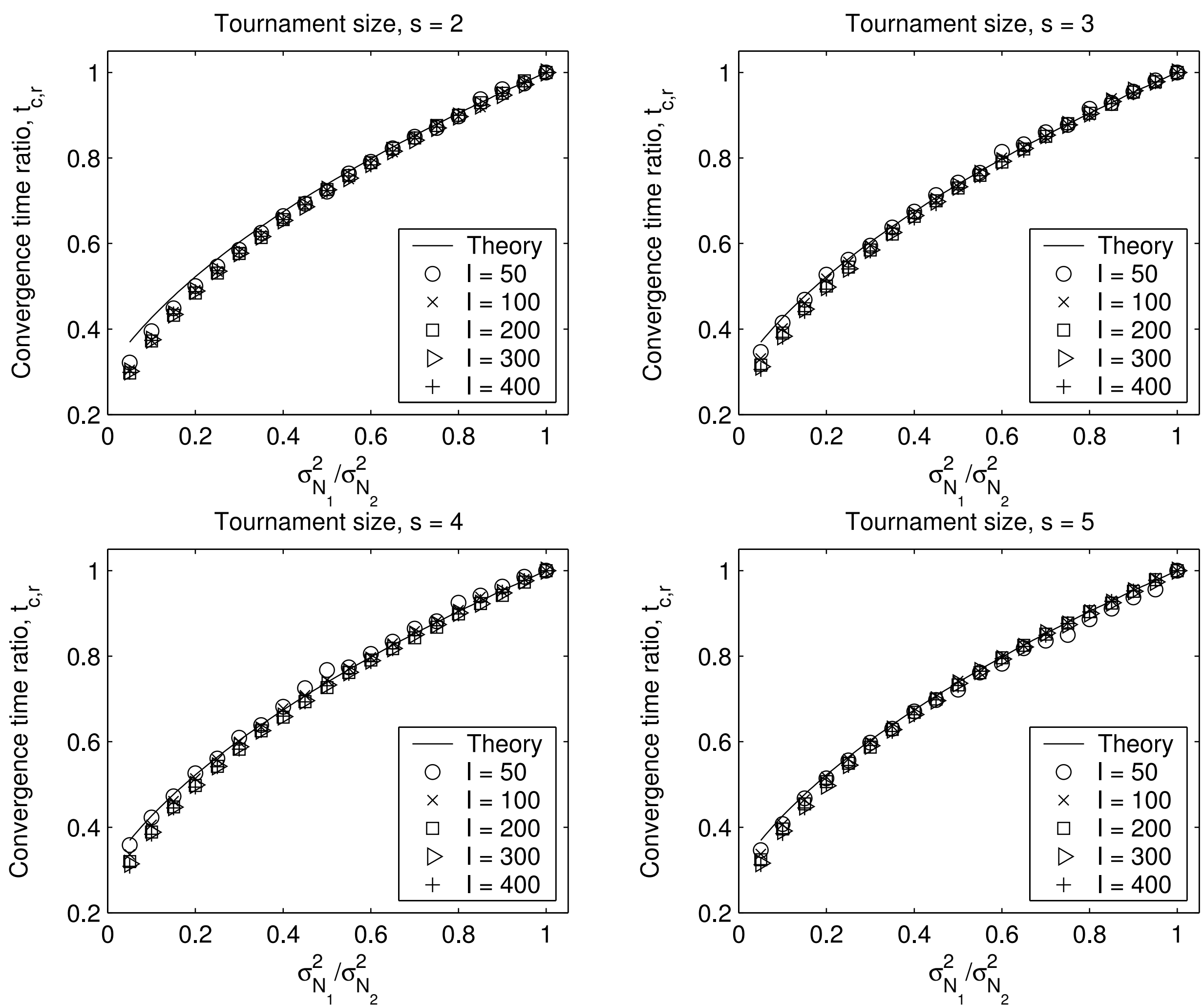
- Constant elongation factor: $\rho_e(t) = \rho_e(t = 0)$
- Convergence time

$$t_{\text{conv}}\left(\sigma^2_{N_i}\right) = \frac{\pi\sqrt{\ell}}{2I}\sqrt{1 + \frac{\sigma^2_{N_i}}{\sigma^2_0}}$$

- Convergence-time ratio

$$t_{c,r} = \frac{t_{\text{conv}}\left(\sigma^2_{N_1}\right)}{t_{\text{conv}}\left(\sigma^2_{N_2}\right)} = \left(\frac{\sigma^2_0 + \sigma^2_{N_1}}{\sigma^2_0 + \sigma^2_{N_2}}\right)^{\frac{1}{2}}$$

Convergence-Time Model Verification



Population-Sizing Model

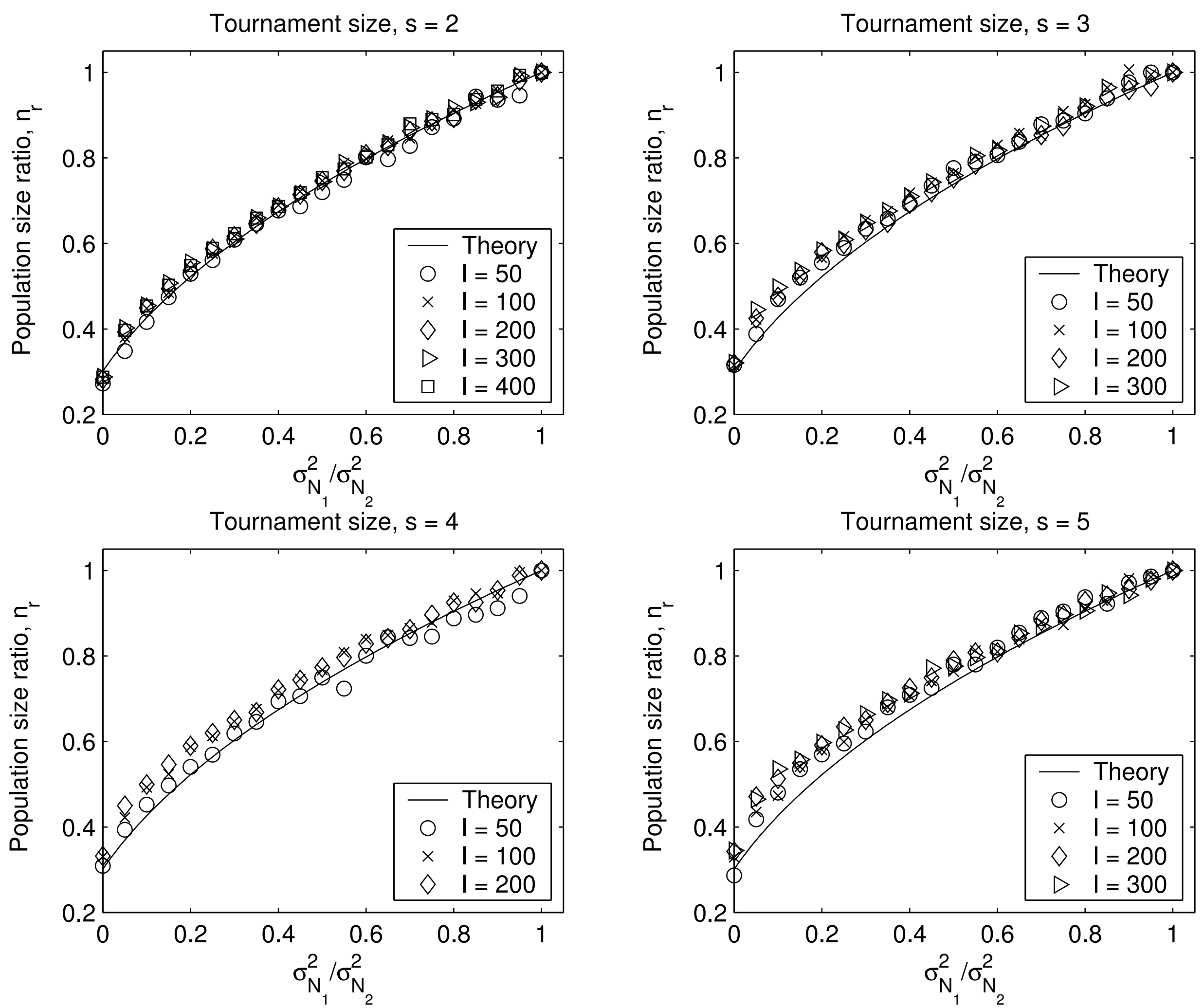
- Gambler's ruin population-sizing model
  - Harik, Cantú-Paz, Goldberg & Miller (1997)
  - Building-block supply + Decision-making model
- External noise (Miller & Goldberg, 1997)

$$n\left(\sigma_{N_i}^2\right) = -\frac{\sqrt{\pi}}{d} 2^{k-1} \ln(\alpha) \sqrt{\sigma_0^2 + \sigma_{N_i}^2}$$

- Population-size ratio

$$n_r = \frac{n\left(\sigma_{N_1}^2\right)}{n\left(\sigma_{N_2}^2\right)} = \left(\frac{\sigma_0^2 + \sigma_{N_1}^2}{\sigma_0^2 + \sigma_{N_2}^2}\right)^{\frac{1}{2}}$$

Population-Sizing Model Verification



Number of Function Evaluations

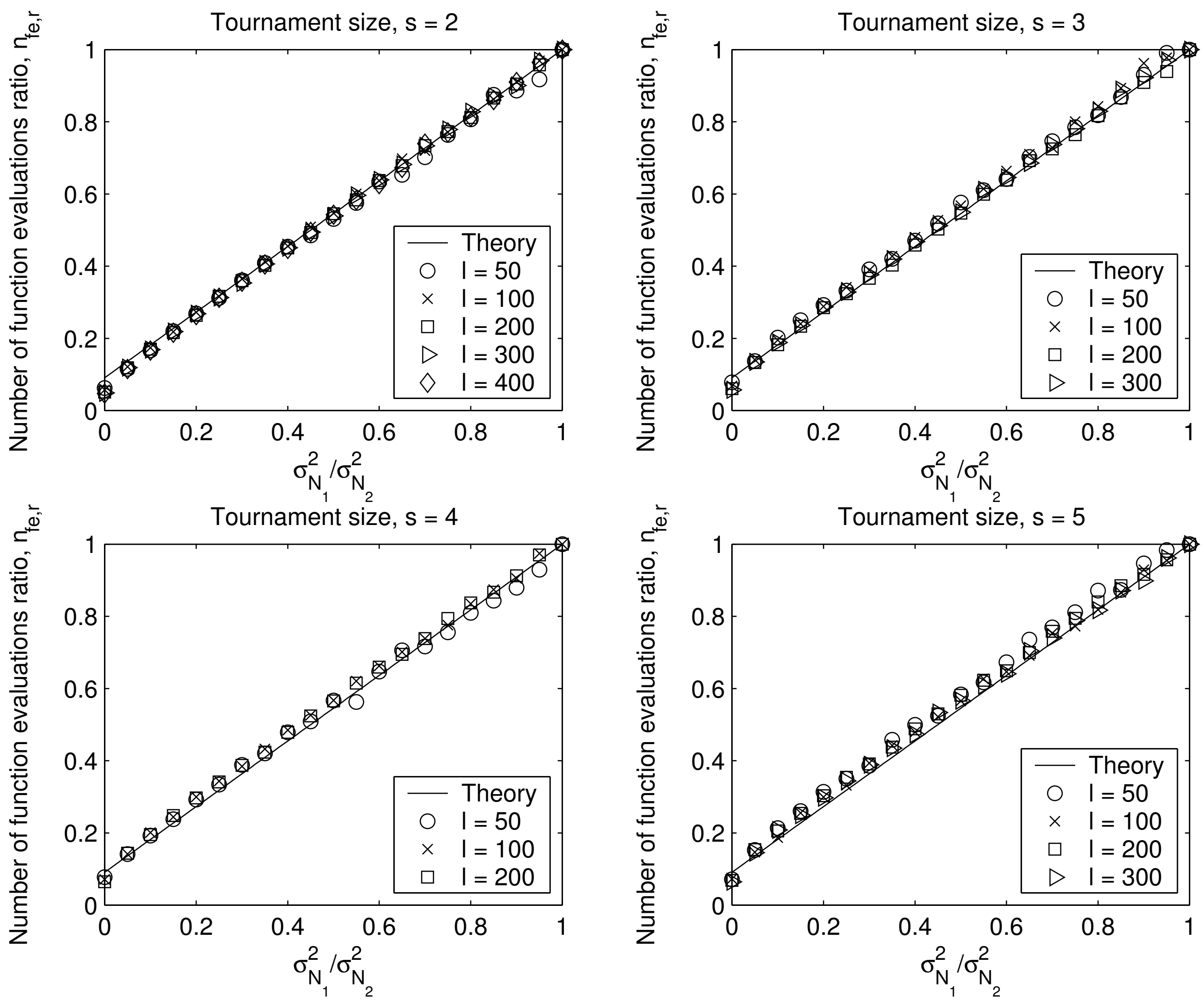
- Convergence-time, and population-size ratios

$$t_{c,r} = n_r = \left(\frac{\sigma_0^2 + \sigma_{N_1}^2}{\sigma_0^2 + \sigma_{N_2}^2}\right)^{\frac{1}{2}}$$

- No. of function evaluations ratio

$$n_{fe,r} = n_r t_{c,r} = \left(\frac{\sigma_0^2 + \sigma_{N_1}^2}{\sigma_0^2 + \sigma_{N_2}^2}\right)$$

Number of Function Evaluations



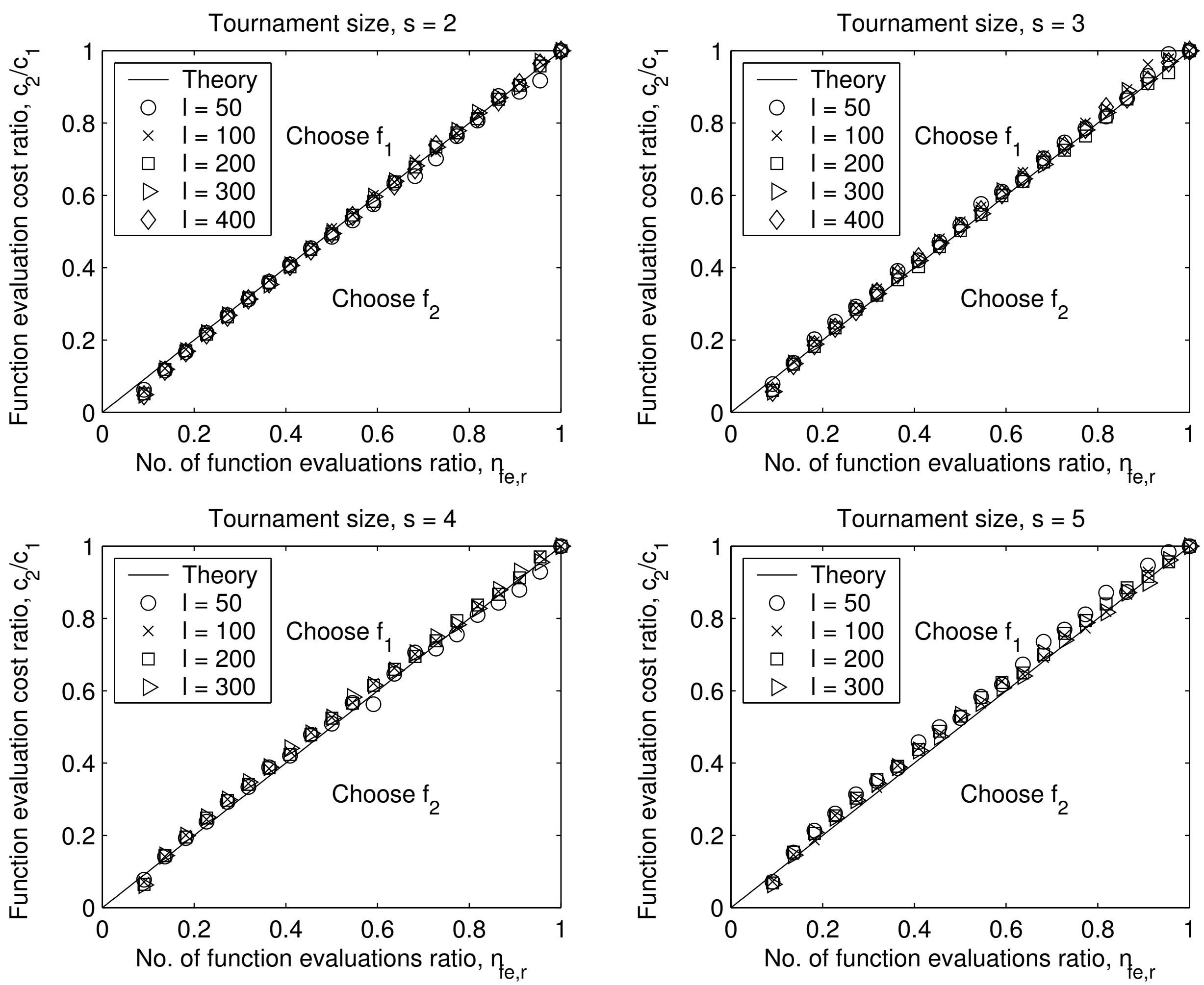
Optimal Decision for Handling Variance

- Total computational cost ratio

$$c_{\text{tot},r} = \frac{c_1}{c_2} n_{fe,r} = \frac{c_1}{c_2} \left(\frac{\sigma_0^2 + \sigma_{N_1}^2}{\sigma_0^2 + \sigma_{N_2}^2} \right)$$

- If $c_2/c_1 > (\sigma_0^2 + \sigma_{N_1}^2)/(\sigma_0^2 + \sigma_{N_2}^2)$, then use f_1
- If $c_2/c_1 < (\sigma_0^2 + \sigma_{N_1}^2)/(\sigma_0^2 + \sigma_{N_2}^2)$, then use f_2
- If $c_2/c_1 = (\sigma_0^2 + \sigma_{N_1}^2)/(\sigma_0^2 + \sigma_{N_2}^2)$, then use f_1 or f_2

Verification of Optimal Decision

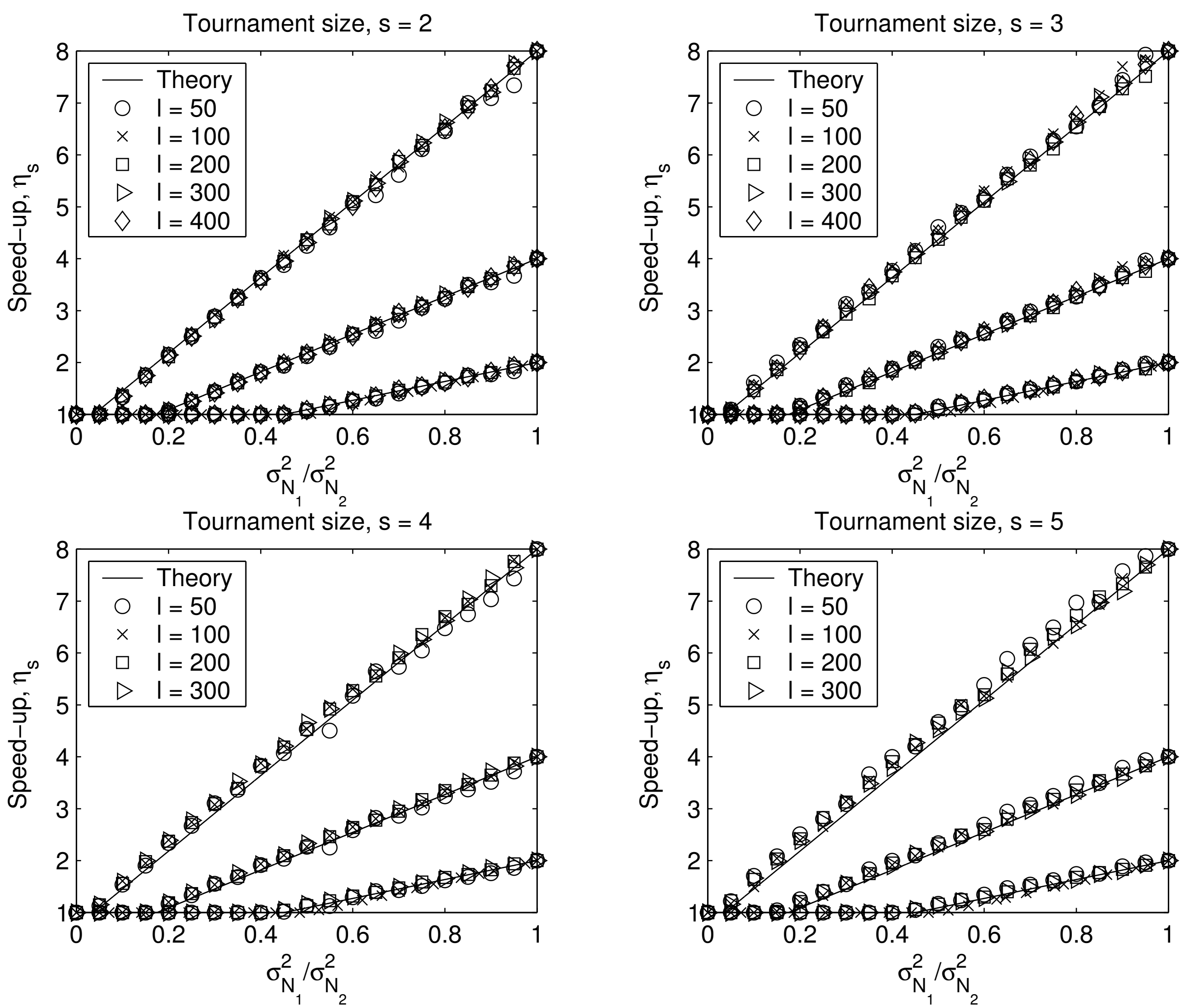


Model to Predict Speed-Up

- Optimal vs. Naïve decision
  - Use low-variance, high-cost fitness function

$$\eta_s = \begin{cases} c_{\mathrm{tot},r} & \frac{c_2}{c_1} < \frac{\sigma_0^2+\sigma_{N_1}^2}{\sigma_0^2+\sigma_{N_2}^2} \\ 1 & \text{otherwise} \end{cases}$$

Verification of Speed-Up



Aside: Sampling Fitness Functions

- Noisy fitness function

$$f' = f + \mathcal{N}\left(0, \sigma_N^2\right)$$

- Sampling fitness function

$$f_{n_s}(S) = \frac{1}{n_s} \sum_{j=1}^{n_s} f'(S)$$

- Effect of sampling
  - Variance: σ_N^2 / n_s
  - Cost: $\alpha + n_s \beta$

Optimal sampling

- Total computational cost

$$c_{\text{tot}} \propto (\alpha + n_s \beta)\left(\sigma_0^2 + \sigma_N^2 / n_s\right)$$

- Optimal sampling rate

$$\frac{\partial c_{\text{tot}}}{\partial n_s} = 0$$

$$n_s^* = \left(\frac{\alpha}{\beta}\frac{\sigma_N^2}{\sigma_0^2}\right)^{\frac{1}{2}}$$

- Same result as Miller & Goldberg (1997)
  - Does not account for convergence-time
  - Uses BB decision-making model

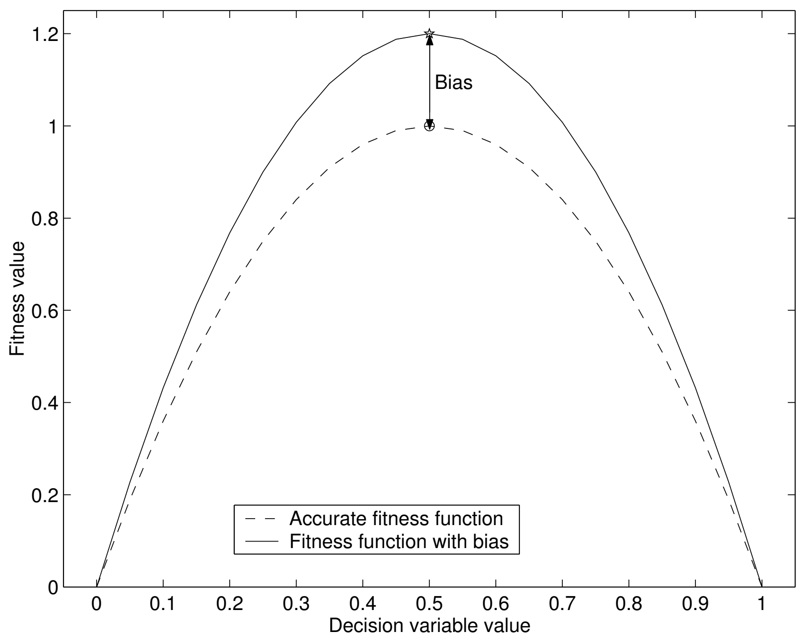
Handling Variance: Summary

- Decision-making strategy:
  - Fitness functions with differing variances
  - Utilized facetwise models
  - Provides maximum speed-up
- Key parameters:
  - Fitness-variance and fitness-cost ratios
- Variance can be handled spatially

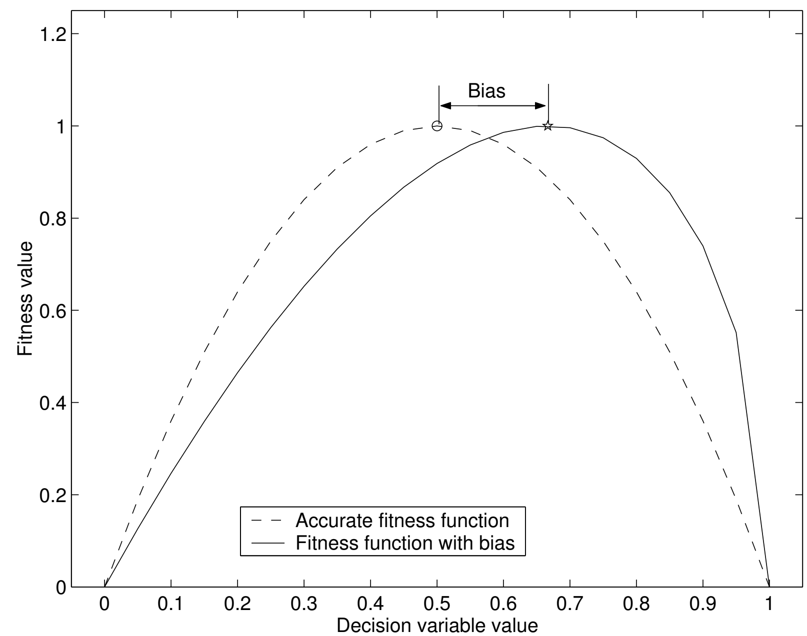
Ingredients of Error

- *Variance:* Changes the fitness value
  - Sampling eliminates effects of variance
- *Bias:* Changes the optimal solution
  - Bias cannot be averaged out
  - Can be handled temporally
    - Use high-bias function initially
    - Then switch to low-bias function

Bias in Fitness Functions



(a) Bias in optimal value



(b) Bias in optimal solution

Part II: Handling Bias

- Two fitness functions: f_1, and f_2
  - f_1: High cost (c_1), and low bias (b_1)
  - f_2: Low cost (c_2), and high bias (b_2).

$$c_1 > c_2, \quad b_2 > b_1$$

  - Higher the accuracy, costlier the evaluation.
- Use f_2 initially
- Switch to f_1 later
- Key parameter: *Switching time*

Weighted OneMax function

- Weights associated with each bit

$$f(\mathbf{X}) = \sum_{i=1}^{\ell} w_i x_i \qquad w_i \in \{-1, +1\}$$

- Different bias values = Different weights
- OneMax properties
  - Unimodal, uniformly scaled
  - Analytical tractability
- Bias does not affect population sizing

GA Run: Initial Stages ($t < t_s$)

- High-bias, low-cost function f_2
- Gaussian fitness distribution

$$\begin{aligned} \mu_{f_2,t} &= \ell p_t - (\ell - \ell_1 - b) \\ \sigma^2_{f_2,t} &= \ell p_t \left(1 - p_t\right) \end{aligned}$$

- *Selection-intensity* based model

$$p_t = \frac{1}{2}\left[1 + \sin\left(\frac{It}{\sqrt{\ell}}\right)\right]$$

GA Run: At Switching Time ($t = t_s$)

- Proportion of correct building blocks

$$p_{t_s} = \frac{1}{2}\left[1 + \sin\left(\frac{I t_s}{\sqrt{\ell}}\right)\right]$$

- Switch from f_2 to f_1
  - $\ell - b$ BBs: p_{t_s}
  - b BBs: $1 - p_{t_s}$
- Corrected proportion of correct BBs

$$p'_{t_s} = \left(1 - 2\frac{b}{\ell}\right) p_{t_s} + \frac{b}{\ell}$$

GA Run: Later Stages ($t > t_s$)

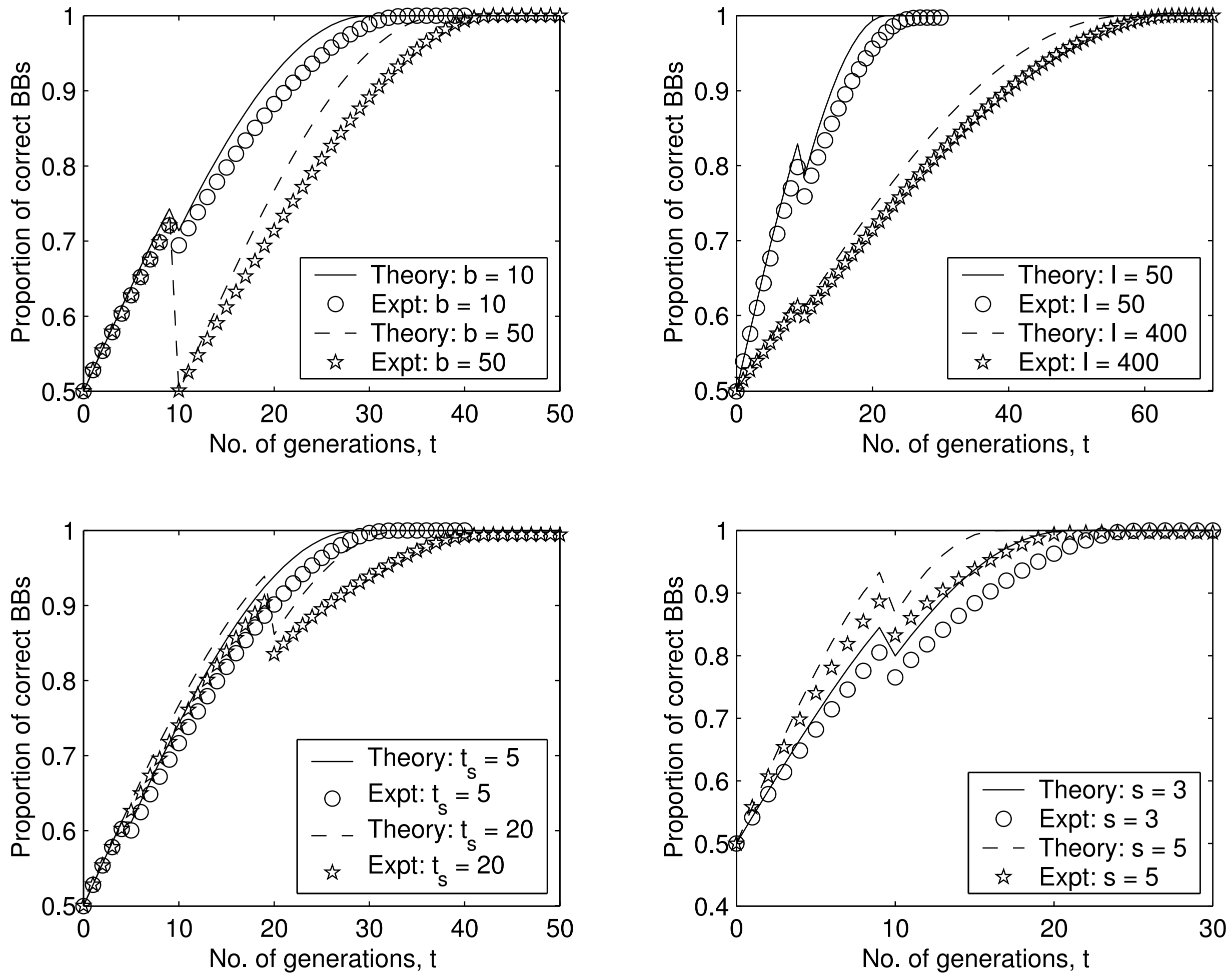
- Proportion of correct BBs
  - Unbiased and biased portions are normally distributed
  - Overall fitness distribution is normal

$$p'_t = \frac{1}{2}\left[1 - \cos\left(\frac{I(t-t_s)}{\sqrt{\ell}} + 2\sin^{-1}\left(\sqrt{p'_{t_s}}\right)\right)\right]$$

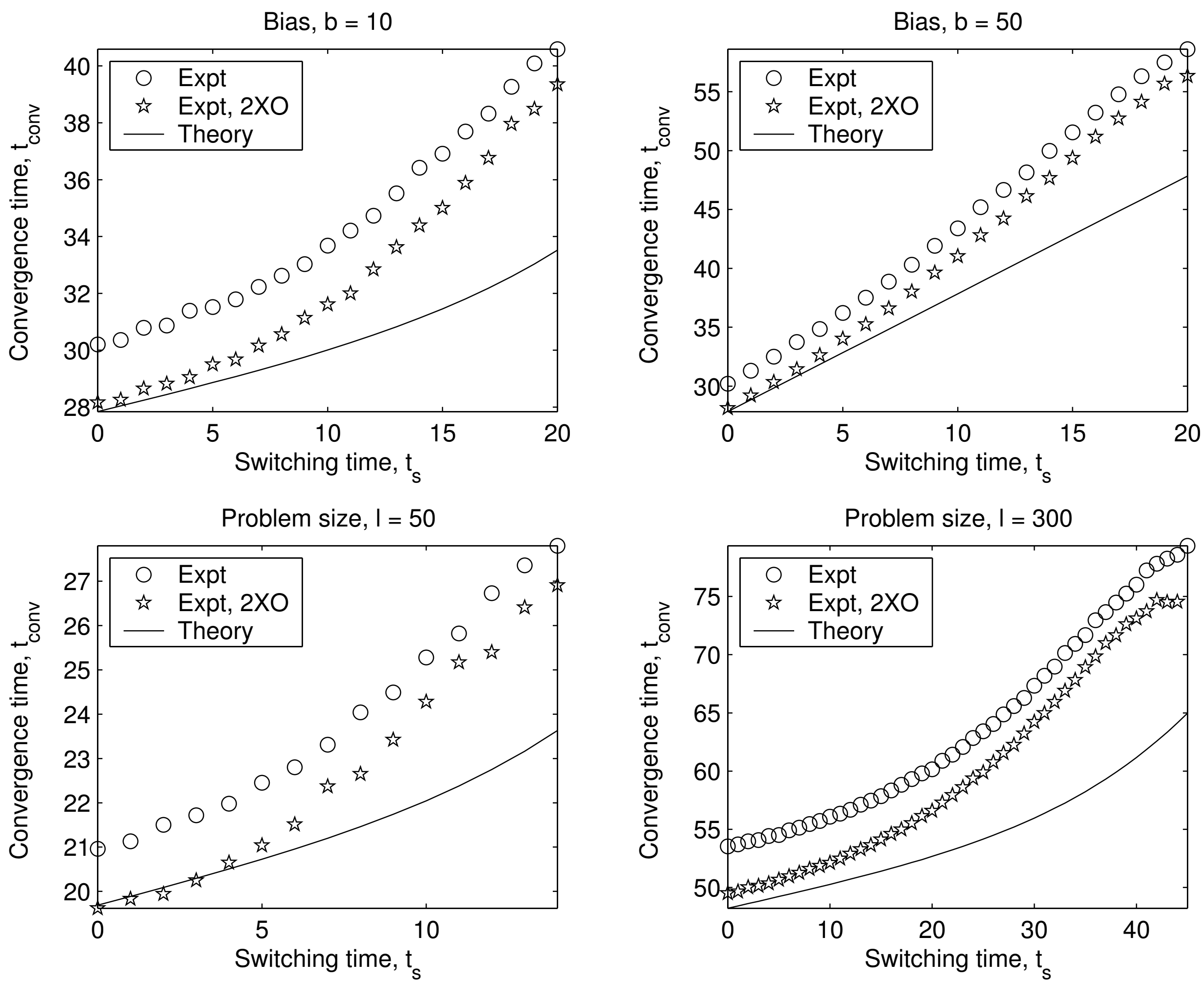
- Convergence time

$$t_{\mathrm{conv}} = t_s + \frac{\sqrt{\ell}}{I}\left[\pi - 2\sin^{-1}\left(\sqrt{p'_{t_s}}\right)\right]$$

Run-Duration Model Verification - I



Run-Duration Model Verification - II



When Should We Switch?

- Total evaluation cost

$$c_{\text{tot}} = n\left[c_2 t_s + c_1\left(t_{\text{conv}} - t_s\right)\right]$$

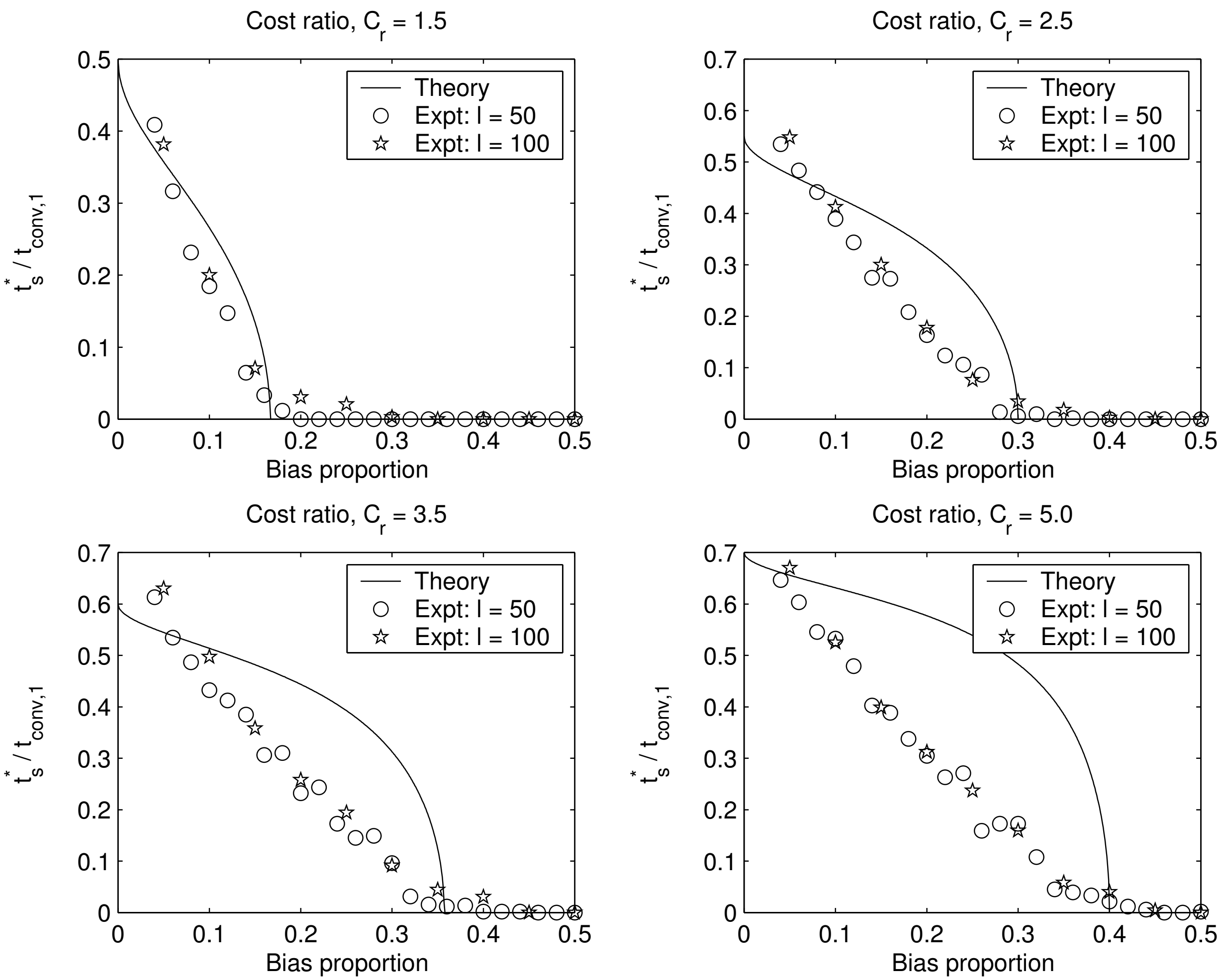
- Optimal switching time

$$\begin{aligned}\frac{\partial c_{\text{tot}}}{\partial t_s} &= 0\\ \tau_s^* &= \frac{t_s^*}{t_{\text{conv},1}} \approx 1 - \frac{4}{\pi}\frac{\sqrt{\beta(1-\beta)}}{(1-2\beta)\sqrt{c_r^2 - 1}}\end{aligned}$$

  – Bias proportion $\beta = b/\ell$

  – f_1 is used alone: $t_{\text{conv},1} = \pi\sqrt{\ell}/(2I)$

Optimal Switching Time



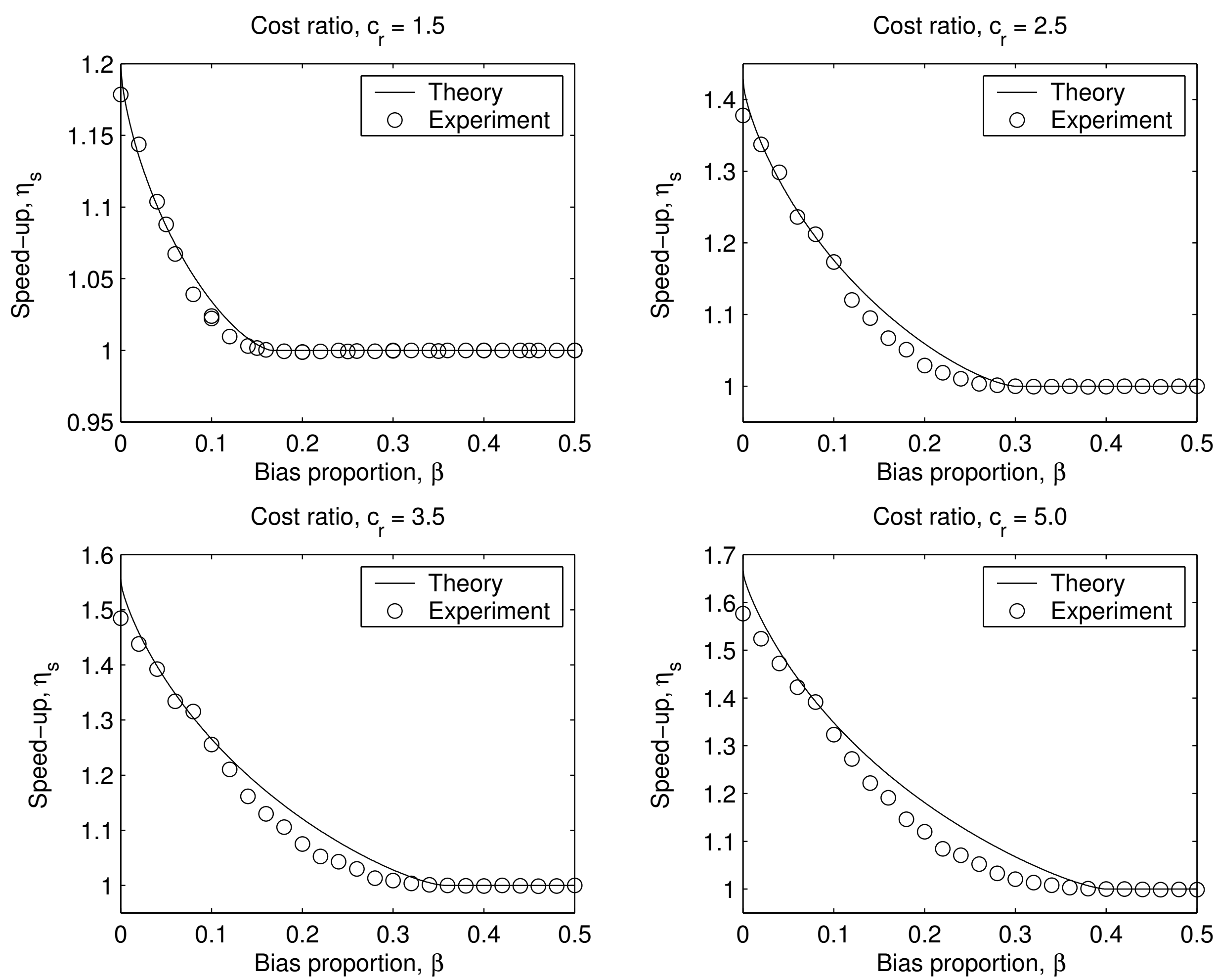
Model for Predicting Speed-Up

- Optimal vs. Naïve decision
  - Use high-cost, low-bias function

$$\eta_s = \frac{c_r}{\left[\frac{t_{\text{conv}}}{t_{\text{conv},1}} - (c_r - 1)\tau_s^*\right]}$$

  - Cost proportion: $c_r = c_1/c_2$
  - Valid when $c_r > (1 - 2\beta)^{-1}$

Verification of Speed-Up



Handling Bias: Summary

- Decision-making strategy:
  - Fitness functions with differing bias values
  - Utilized facetwise models
  - Provides maximum speed-up
- Key parameters:
  - Bias proportion and fitness-cost ratios
- Bias can be handled temporally

Conclusions

- Evaluation-relaxation can yield good speed-up
- Error has two parts: Variance, and bias
- Handle variance spatially
  - Fitness-variance and fitness-cost ratios
- Handle bias temporally
  - Use high-bias, low-cost function initially
  - Switch to low-bias, high-cost function later
  - Bias proportion and fitness-cost ratios

Acknowledgments

- Air Force Office of Scientific Research, Air Force Materiel Command, USAF, F49620-00-1-0163.
- National Science Foundation, DMI-9908252.