Approximate Dynamic Programming for the Merchant Operations of Commodity and Energy Conversion Assets

Selvaprabu (Selva) Nadarajah,
(Joint work with François Margot and Nicola Secomandi)
Tepper School of Business, Carnegie Mellon University

Enterprise Wide Optimization Seminar
February 19th 2013

(Research supported by NSF grant CMMI 1129163)

Commodity Conversion Assets: Real Options

- Refineries: Real option to convert a set of inputs into a different set of outputs

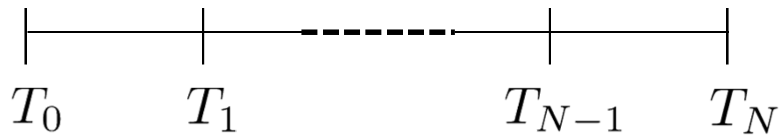
- Natural gas storage: Real option to convert natural gas at the injection time to natural gas at the withdrawal time

How do we optimally manage the available real optionality?

Key Problem Features

Dynamic Decisions

Decisions can be taken over a set of discrete times (stages)



Operational Constraints

Decisions must satisfy operational constraints

These constraints couple decisions over time

Uncertainty

Decisions depend on the evolution of uncertain information

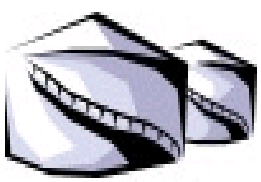
Examples: Commodity forward curve or demand forecast

Decision Making Process

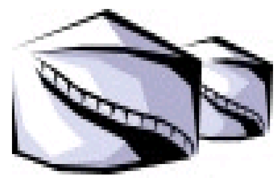
Current Time Period

Forward/Demand Curve

Inventory



Make a decision and receive a reward



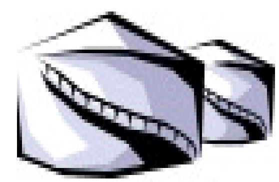
New Inventory

Observe new forward curve

Next Time Period

New Forward/Demand Curve

New Inventory



Elements of Markov Decision Process

- N stages $i \in \mathcal{I} := \{0, \ldots, N-1\}$
- Endogenous state: $x_i := \{x_{i,1}, \ldots, x_{i,L}\} \in \mathcal{X}$ (inventory)
- Exogenous state: $F_i := \{F_{i,i}, F_{i,i+1}, \ldots, F_{i,N-1}\}$ (forward/demand curve)
- Action $a_i(x_i, F_i)$ belongs to the discrete set $\mathcal{A}_i(x_i)$
- Reward function: $r_i : (a_i, x_i, F_i) \mapsto \mathbb{R}$
- Transition rule:
  - Endogenous: $x_{i+1} = x_i - a_i$
  - Exogenous: transition from F_i to F_{i+1} determined by the model of uncertainty (price model or demand model)

The above collection of elements is referred to as a Markov decision process (Puterman 1994)

Markov Decision Problem (MDP)

- Decision function of a policy π: $A_i^\pi : (x_i, F_i) \mapsto \mathcal{A}_i(x_i)$
  - Prescribes a feasible action at every state
- A policy π: Collection of decision functions $\{A_0^\pi, \ldots, A_{N-1}^\pi\}$

Find a policy π that maximizes the sum of discounted expected rewards starting from initial state (x_0, F_0):

$$V_0(x_0, F_0) := \max_{\pi \in \Pi} \sum_{i \in \mathcal{I}} \delta^i \mathbb{E}\left[r(A_i^\pi(\tilde{x}_i^\pi, \tilde{F}_i), \tilde{x}_i^\pi, \tilde{F}_i) | x_0, F_0 \right]$$

Discount factor

(Puterman 1994, Bertsekas 2005)

Stochastic Dynamic Programs (SDPs)

Value function SDP:

$$V_i(x_i, F_i) = \max_{a \in \mathcal{A}_i(x_i)} r_i(a, x_i, F_i) + \delta \mathbb{E}\left[V_{i+1}\left(x_i - a, \tilde{F}_{i+1}\right) | F_i\right]$$

$$\forall i \in \mathcal{I}, (x_i, F_i) \in \mathcal{X}_i \times \mathbb{R}^{N-i}$$

Continuation function SDP:

$$C_i(x_{i+1}, F_i) = \delta \mathbb{E}\left[\max_{a \in \mathcal{A}_{i+1}(x_{i+1})} r_{i+1}(a, x_{i+1}, \tilde{F}_{i+1}) + C_{i+1}\left(x_{i+1} - a, \tilde{F}_{i+1}\right) | F_i\right]$$

- Value function at state (x_i, F_i): The sum of discounted expected rewards from following an optimal policy starting from (x_i, F_i)
- If we can solve any one of these formulations, we have an optimal policy!

Curses Of Dimensionality

$$V_i(x_i, F_i) = \max_{a \in \mathcal{A}_i(x_i)} r_i(a, x_i, F_i) + \delta \mathbb{E}\left[V_{i+1}\left(x_i - a, \tilde{F}_{i+1}\right) | F_i\right]$$

$$C_i(x_{i+1}, F_i) = \delta \mathbb{E}\left[\max_{a \in \mathcal{A}_{i+1}(x_{i+1})} r_{i+1}(a, x_{i+1}, \tilde{F}_{i+1}) + C_{i+1}\left(x_{i+1} - a, \tilde{F}_{i+1}\right) | F_i\right]$$

- High dimensional exogenous information state (e.g. 12 months, 365 days)
    1. Exact value/continuation function is high dimensional
    2. Expectations are high dimensional
- We need to solve these intractable SDPs approximately

Approximate Dynamic Programming (ADP) Template

1. Compute a value function approximation or continuation value function approximation

2. Estimate lower bounds by simulating the induced heuristic policy in Monte Carlo simulation

3. Estimate upper bounds using the information relaxation and duality approach

Optimality gap provides a guarantee on the policy quality

1) *Functional Approximations*

- Fundamental idea: Approximate $V_i(x_i, F_i)$ or $C_i(x_{i+1}, F_i)$ by a *low dimensional* function

- In many practical applications it is typically possible to find good lower dimensional approximations

- $\hat{V}_i(x_i, F_i)$ value function approximation

- $\hat{C}_i(x_{i+1}, F_i)$ continuation function approximation

- Many different ways of obtaining these approximations (Bertsekas 2005, Powell 2011)

2) Lower Bounds: Online Heuristic Policies

- Forgo trying to find an explicit policy over entire state

- Instead, given a state (x_i, F_i), solve a math program in an online fashion to compute actions

$$\underbrace{\max_{a \in \mathcal{A}_i(x_i)} r_i(a, F_i) + \delta \mathbb{E}\left[\hat{V}_{i+1}\left(x_i - a, \tilde{F}_{i+1}\right) | F_i\right]}$$

Stochastic optimization problem.

$$\underbrace{\max_{a \in \mathcal{A}_i(x_i)} r_i(a, F_i) + \delta \hat{C}_i(x_i - a, F_i)}$$

Deterministic optimization problem. No expectation!

Simulating this online policy in Monte Carlo simulation gives a lower bound estimate

When these approximations are exact the online actions match the actions from an optimal policy

3) Upper Bounds

Intuition: Allow the decision maker to use future information and then penalize this future knowledge [Rogers (2002), Haugh and Kogan (2004), Brown et al.(2010)]

- Upper bound estimation involves solving a collection of deterministic dynamic programs in Monte Carlo simulation

- Value/continuation function approximations can be used in this procedure to define penalties

- If the value/continuation function approximations are exact then the upper bound is equal to the value of an optimal policy

ADP Template

1. Compute/Estimate a value function approximation or continuation value function approximation

2. Estimate lower bounds by simulating the induced heuristic policy in Monte Carlo simulation ✓

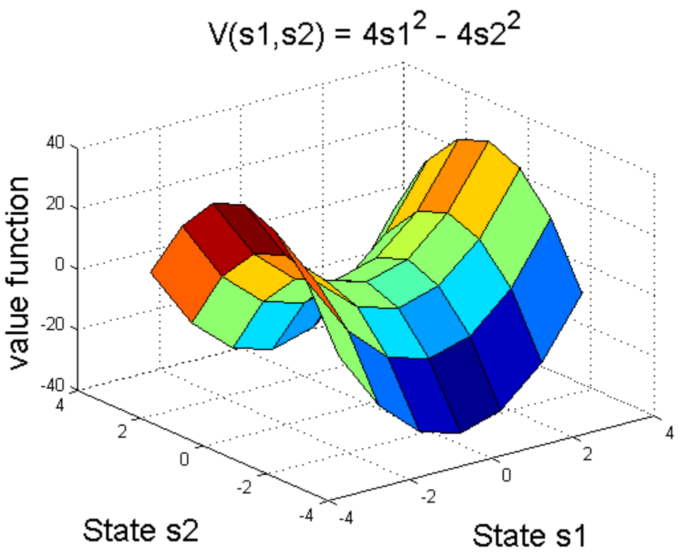
3. Estimate upper bounds using the information relaxation and duality approach ✓

Optimality gap provides a guarantee on the policy quality

Basis Function Approximations

- Express approximation as a linear combination of known functions referred to as *basis functions*

- Basis functions: Maps from the state space to the real line (Bellman and Dreyfus 1956, Bertsekas 2005, Powell 2011)



- Choose basis function as $s1^2 - s2^2$
- Write $\hat{V}(s1, s2) := (s1^2 - s2^2)\beta$
- In practice, the value function is unknown
- It is typically possible to obtain some information about the function's structure

- Basis functions are typically a user input to an ADP method

Basis Function Approximations contd

- Value function approximation

$$\hat{V}_i(x_i, F_i; \beta_i) := \sum_{b=1}^{B_i} \phi_{i,b}(x_i, F_i)\beta_{i,b}$$

Basis functions

- Continuation function approximation

$$\hat{C}_i(x_{i+1}, F_i; \theta_i) := \sum_{b=1}^{B_i} \psi_{i,b}(x_{i+1}, F_i)\theta_{i,b}$$

- How do we compute the weights β_i or θ_i ?

ADP Approximation Methods

This talk:

1. Monte Carlo based regression methods

2. Approximate linear programming

Other methods

3. Reinforcement learning

Regression Methods

1. Simple endogenous state and high dimensional exogenous state
   - Endogenous state is typically one dimensional
   - Exogenous state is a forward curve or demand curve
   - Pioneered by Carriere 1996 (250+ citations), Longstaff and Schwartz 2001 (1650+ citations) and Tsitsiklis and Van Roy 2001 (300+ citations) for pricing American options

2. High dimensional endogenous state and no exogenous state
   - Endogenous state is high dimensional
   - Uncertainty is iid and thus does not appear in the MDP state
   - see Powell (2011) for more details

3. High dimensional endogenous and exogenous state
   - Largely unexplored by the OR community

Regression Methods: Real Options

- Compute a continuation function approximation $\hat{C}_i(x_{i+1}, F_i; \theta_i)$ using extensions of the Longstaff and Schwartz (2001) approach for American options
- Combine Monte Carlo simulation and least squares regression in a recursive procedure to compute the basis function weights θ_i
- Standard for real option pricing in practice and academia
  - Switching options (Cortazar 2008)
  - Gas storage (Boogert and De Jong 2008)

Elegant Idea: Point Estimate of Expectation

- Suppose we have a continuation function approximation $\hat{C}_{i+1}(x_{i+2}, F_{i+1}; \bar{\theta}_{i+1})$ at stage $i+1$ and want to find θ_i

- Sample P forward curve paths $\{F_i^p, p = 1, \ldots, P\}$

- For each sample compute the stage i continuation function estimate

$$\delta \max_{a \in \mathcal{A}_{i+1}(x_{i+1})} r_{i+1}(a, x_i, F_{i+1}^p) + \hat{C}_{i+1}\left(x_{i+1} - a, F_{i+1}^p; \bar{\theta}_{i+1}\right)$$

- Regress over estimates to compute stage i continuation function approximation weights θ_i

Regression Methods: Value Function

- N. et al. (2012a): Wouldn't it be nice if we could compute expectations exactly?

- Possible when using a *value function approximation* for:
  1. a class of basis functions and
  2. a rich class of forward curve evolution models that is popular among practitioners

- Value function approach outperforms the continuation function approach on our numerical experiments on swing option and commodity storage instances

- We also provide some theoretical support for this numerical performance

ADP Approximation Methods

This talk:

1. Monte Carlo based regression methods ✓

2. Approximate linear programming

Approximate Linear Programming

- Computes the weights of a value function approximation by solving a linear program

  (Schweitzer and Seidman 1985, deFarias and Van Roy 2003)

- Popular in the operations research literature:
  - Economics: Trick and Zin (1997)
  - Inventory control: Adelman (2004) and Adelman and Klabjan (2011)
  - Revenue Management: Adelman (2007), Farias and Van Roy (2007), Zhang and Adelman (2009)
  - Queueing: Morrison and Kumar (1999), de Farias and Van Roy (2001,2003), Moallemi et al. (2008), and Vaetch (2010).

- A large exogenous information vector is absent in the state of most SDPs considered in the approximate LP literature

Exact Primal and Dual Linear Programs

- LP reformulation of the value function SDP (Manne 1960)

$$
\begin{aligned}
&\min_{V} V_0(x_0, F_0) \\
&\text{s.t. } V_N(x_N, F_N) = 0, \ \forall x_N \in \mathcal{X}_N. \\
&\quad V_i(x_i, F_i) \geq r(a_i, s_i) + \delta \mathbb{E}\left[V_{i+1}\left(x_i - a_i, \tilde{F}_{i+1}\right) | F_i \right], \\
&\qquad \forall i \in \mathcal{I}, (x_i, F_i) \in \mathcal{X}_i \times \mathcal{F}_i, a_i \in \mathcal{A}_i(x_i).
\end{aligned}
$$

Intractable!

- Computes the value function at all states visited by an optimal policy starting from the initial state.

- Dual variables $w_i(x_i, F_i, a_i)$ can be interpreted as (discounted) probabilities and are in one-one correspondence with feasible policies (Puterman 1994)

- The exact dual finds an optimal policy

Approximate Primal and Dual Linear Programs

$$\text{ALP:} \min_{\beta} \beta_{0,1}$$

$$\text{s.t.} \quad \beta_N = 0,$$

$$\sum_{b=1}^{B_i} \phi_{i,b}(x_i, F_i)\beta_{i,b} \geq r(a_i, s_i) + \delta \mathbb{E}\left[\sum_{b=1}^{B_{i+1}} \phi_{i+1,b}(x_i - a_i, \tilde{F}_{i+1})\beta_{i+1,b} | F_i\right],$$

$$\forall i \in \mathcal{I}, (x_i, F_i) \in \mathcal{X}_i \times \mathcal{F}_i, a_i \in \mathcal{A}_i(x_i).$$

- Apply value function approximation on the exact primal variables
- Tractable number of variables but large number of constraints
- Solve ALP to compute weights β_i
- Dual variables $w_i(x_i, F_i, a_i)$ can be still interpreted as (discounted) probabilities
- ALP has theoretical guarantees (deFarias and Van Roy 2003)

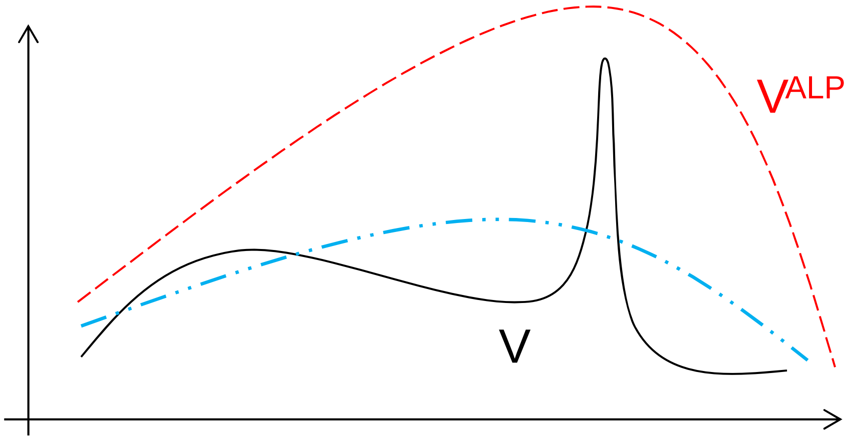
Solving ALP

- Constraint sampling
  - A small number of constraints are sufficient to determine the optimal solution to ALP
  - Theoretical sampling guarantees (de Farias and Van Roy 2004)
  - Standard approach for solving an ALP

- Column generation
  - Solve the ALP dual using column generation
  - Revenue management (Adelman 2004)

Is ALP the Correct Math Program?

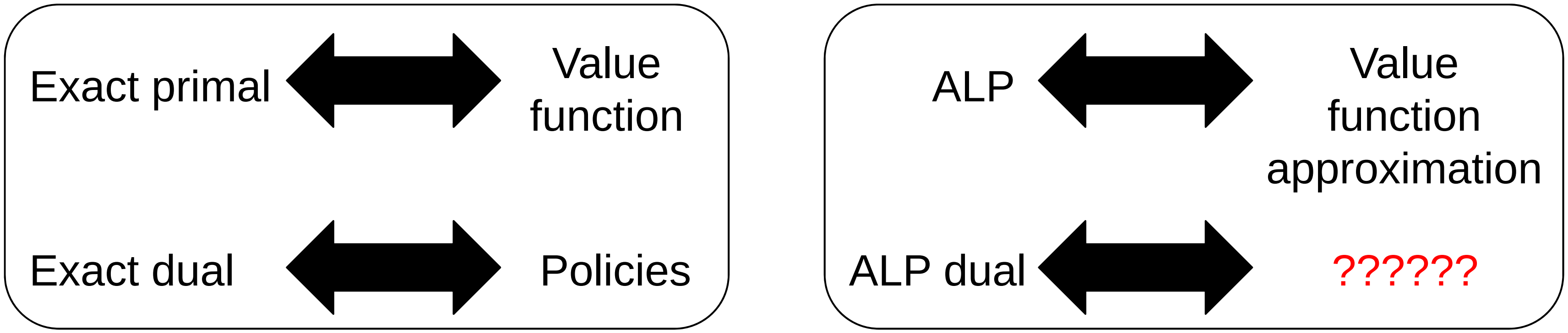
- The ALP constraints require the value function approximation to be an upper bound on the exact value function at every state



- Petrik and Zilberstein (2009) proposed a relaxation of ALP to overcome this issue

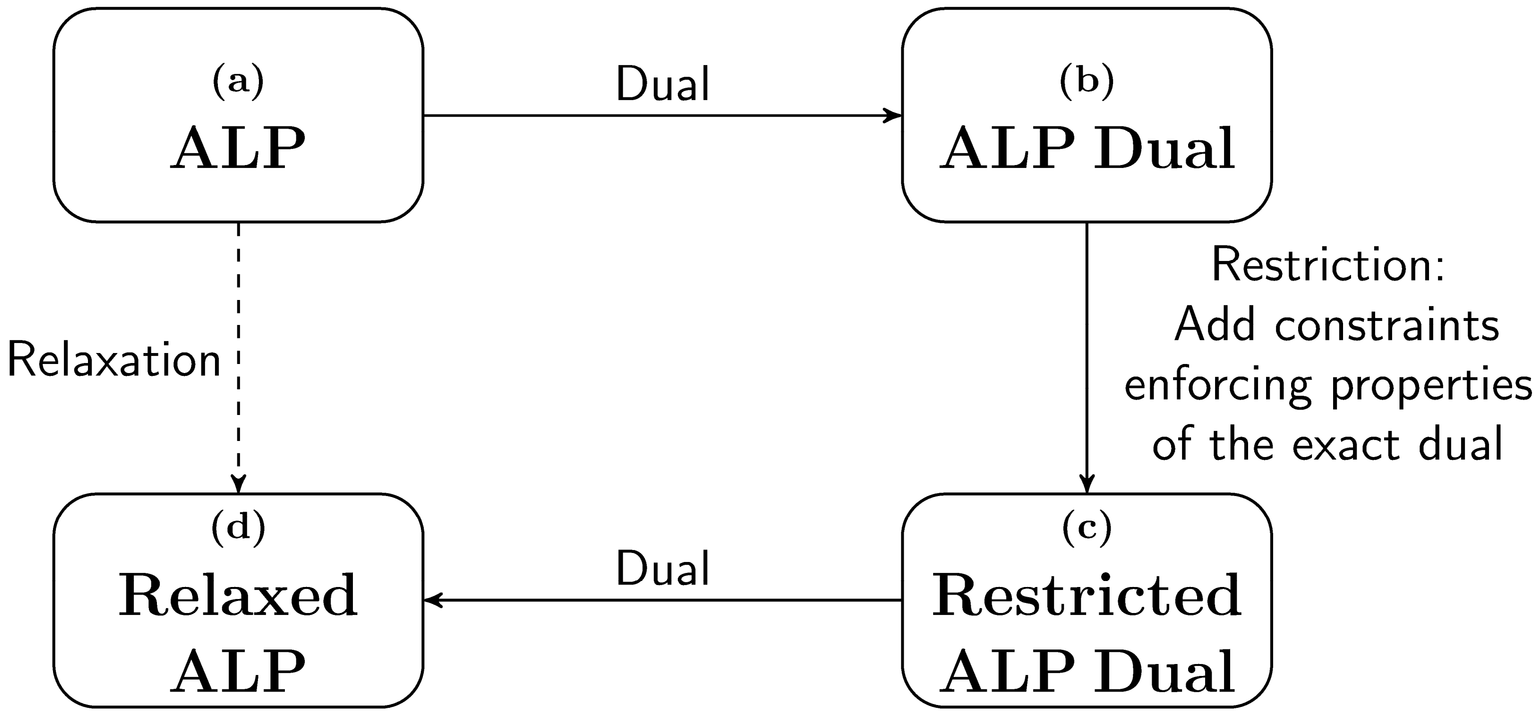
- Desai et al. (2012) provide strong theoretical guarantees and practical implementation rules for this ALP relaxation

Probability Distortions and Pathologies



- N. et al. (2012b): Is the optimal solution set of the ALP dual related to optimal policies?
- Not necessarily! The optimal solution set of the ALP dual can have large distortions from the probability distributions of optimal policies.
- These large distortions can lead to pathological scenarios

A New ADP Approach



- General framework to derive ALP relaxations (N. et al. 2012b)
- Solve relaxed ALP to obtain a value function approximation

Are ALP relaxations useful?

- We apply ALP relaxations to commodity storage (N. et al 2012b)
- Lower and upper bound improvements over ALP as a percentage of best upper bound
  - Lower bound improvements as large as 99%
  - Upper bound improvements as large as 600%
- Policies from an ALP relaxation were near optimal on our commodity storage instances

Summary

- The merchant operations of commodity and energy conversion assets is a practically important area of research that give rises to intractable SDPs.
- Approximate dynamic programming provides a rich set of tools to heuristically solve intractable SDPs
- Problems with large (correlated) exogenous information variables in the state lead to new challenges that require new ADP methodology

Ongoing Work

- Methodology:
  – Exploring other math programming approaches for obtaining value function approximations
  – ADP methods for real options problems where the endogenous state is also a vector

- Applications:
  – Integrated management of commodity storage and transport on a pipeline system
  – Many more…..

Thank you!

(snadarajah@cmu.edu)

References

- D. Adelman. A price-directed approach to stochastic inventory/routing. Operations Research, 52(4):499-514,2004.
- D. Adelman. Dynamic bid prices in revenue management. Operations Research, 55(4):647-661, 2007.
- D. Adelman and D. Klabjan. Computing near optimal policies in generalized joint replenishment. INFORMS Journal on Computing, Forthcoming, 2011.
- Boogert, A., C. De Jong. Gas storage valuation using a Monte Carlo method. The Journal of Derivatives15(3) 81-98, 2008.
- Bellman, R., S. Dreyfus. 1959. Functional approximations and dynamic programming. Mathematical Tables and Other Aids to Computation 13(68) 247{251.
- Bertsekas, P. B. 2007. Dynamic Programming and Optimal Control , vol. 2. 3rd ed. Athena Scientic, Nashua, New Hampshire, USA.
- Carriere, J.F.Valuation of the early-exercise price for options using simulations and nonparametric regression. Insurance: Mathematics and Economics 19(1) 19-30, 1996.
- Cortazar, G., M. Gravet, J. Urzua.The valuation of multidimensional American real options using the LSM simulation method. Computers & Operations Research 35(1) 113-129, 2008.
- V. F. Farias and B. Van Roy. An approximate dynamic programming approach to network revenue management.Working paper, Stanford Univ., 2007.

References Contd

- Glasserman, P. Monte Carlo Methods in Financial Engineering. Springer, New York, NY, USA, 2004.
- F. A. Longstaff and E. S. Schwartz. Valuing American options by simulation: A simple least-squares approach. Review of Financial Studies, 14(1):113-147, 2001.
- A. S. Manne. Linear programming and sequential decisions. Management Science, 60(3):259-267, 1960.
- C. C. Moallemi, S. Kumar, and B. Van Roy. Approximate and data-driven dynamic programming for queuing networks. Working paper, Stanford Univ., 2008.
- J. R. Morrison and P. R. Kumar. New linear program performance bounds for queuing networks. Journal of Optimization Theory and Applications, 100(3):575-597, 1999.
- S. Nadarajah, F. Margot, N. Secomandi, Relaxations of Approximate Linear Programs for the Real Option Management of Commodity Storage, Working paper, Carnegie Mellon Univ., 2012
- S. Nadarajah, F. Margot, N. Secomandi, Valuation of Multiple Exercise Options with Energy Applications, Working paper, Carnegie Mellon Univ., 2012
- M. Petrik and S. Zilberstein. Constraint relaxation in approximate linear programs. In Proceedings of the Twenty-Sixth International Conference on Machine Learning, pages 809-816, Montreal, Canada, 2009.
- W. B. Powell. Approximate Dynamic Programming: Solving the Curses of Dimensionality, 2nd Edition. John Wiley & Sons, Hoboken, New Jersey, USA, 2011.

References Contd

- M.L. Puterman. Markov Decision Processes: Discrete Stochastic Dynamic Programming. John Wiley & Sons, Inc., New York, NY, USA, 1994.
- L. C. G. Rogers. Monte Carlo valuation of American options. Mathematical Finance, 12(3):271-286, 2002.
- P. J. Schweitzer and A. Seidmann. Generalized polynomial approximations in Markovian decision processes. Journal of Mathematical Analysis and Applications, 110(2):568-582, 1985.
- M. A. Trick and S. E. Zin. Spline approximations to value functions. Macroeconomic Dynamics, 1(1):255-277,1997.
- J.N. Tsitsiklis and B. Van Roy. Regression methods for pricing complex American-style options. IEEE Transactions on Neural Networks, 12(4):694-703, 2001.
- M. H. Veatch. Approximate linear programming for networks: Average cost bounds. Working paper, Gordon College, 2010.
- D. Zhang and D. Adelman. An approximate dynamic programming approach to network revenue management with customer choice. Transportation Science, 43(3):381-394, 2009.