Constraint Relaxation in Approximate Linear Programs

Marek Petrik, Shlomo Zilberstein
University of Massachusetts Amherst

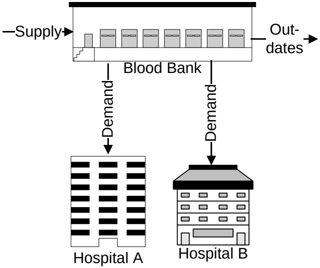
June 16, 2009

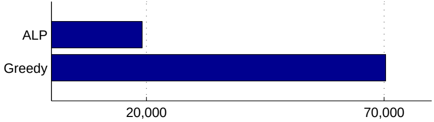
Approximate Linear Programming

- Value function approximation in large Markov decision problems
- **Properties**:
  - \+ Better convergence properties than other algorithms
  - \+ Easier to analyze
  - − Inferior empirical performance
- **Goals:**
  1. Identify why ALP under-performs
  2. Automatically improve the performance

Blood Inventory Management Problem

- Managing inventory of blood
- Objectives:
  - Minimize **shortage** – demand that is not satisfied
  - Maximize **utilization** – amount of blood used before it perishes
- Challenging optimization problem:
  - Continuous action space
  - 48-dimensional continuous state space
  - High level of stochasticity





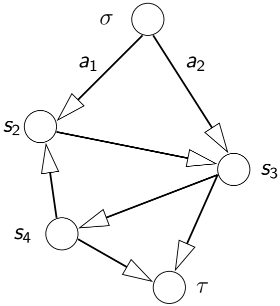
Problem Framework

Markov decision process:

- States: $\mathcal{S}$, including goal state
- Actions: $\mathcal{A}$
- **Transition function**: $p(s_2 \mid s_1, a)$ – probability of transition from s_1 to s_2 with action a
- **Reward function**: $r(s, a)$ for state s and action a

Objective:

- Start with an initial state σ
- **Maximize** discounted reward:



Blood Management

$$\mathbf{E}_{s_0}\left[\sum_{i=0}^{\infty} \gamma^i R_i\right] = \mathbf{E}_{s_0}\left[R_0 + 0.9 R_1 + 0.9^2 R_2 + 0.9^3 R_3 + \ldots\right]$$

Linear Program Formulation

- Linear program:

$$\begin{aligned} \min_{v} \quad & c^{\mathsf{T}} v \\ \text{s.t.} \quad & Av \geq b \end{aligned}$$

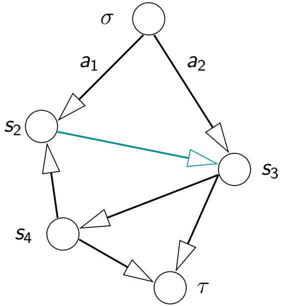
- Constraints:

$$v(s') \geq \gamma \sum_{s \in \mathcal{S}} p\left(s \,\middle|\, s', {\color{blue}a_1}\right) v(s) + r(s', {\color{blue}a_1})$$

$$v(s') \geq \gamma \sum_{s \in \mathcal{S}} p\left(s \,\middle|\, s', {\color{blue}a_2}\right) v(s) + r(s', {\color{blue}a_2})$$

- *Example:*

$${\color{green}v(s_2) \geq \gamma v(s_3) + r(s_2, a_1)}$$



Approximate Linear Program Formulation

- Linear program:

$$\begin{aligned} \min_{v} \quad & c^{\mathsf{T}} v \\ \text{s.t.} \quad & Av \geq b \end{aligned}$$

- Reduce the number of variables in the LP
  - Consider an **approximation basis**: M, as a matrix Example
  - Value function from span(M): $v = Mx$
  - Columns represent features
- Approximate linear program: Example

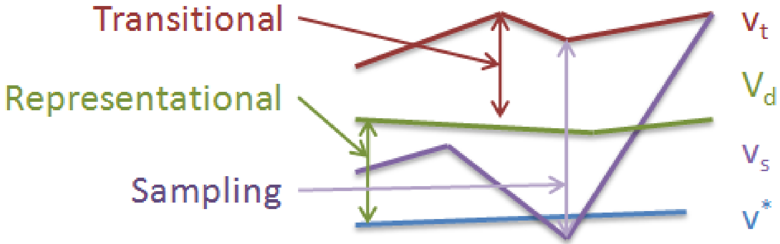
$$\begin{aligned} \min_{x} \quad & c^{\mathsf{T}} M x \\ \text{s.t.} \quad & AMx \geq b \end{aligned}$$

- Many constraints – reduce by *sampling*

Approximation Error

Approximation error:

1. *Representational* – Limited approximation features (basis) M
2. *Transitional* – Limitation of ALP formulation
3. *Sampling* – Limited number of sampled constraints



Formulations

Transitional Error Bounds

- ALP bounds in theory better than other algorithms
- Typical ADP Algorithms:

$$\limsup_{k\to\infty} \|v^* - v_k\|_\infty \leq \limsup_{k\to\infty} \frac{2}{(1-\gamma)^2} \|\tilde{v}_k - v_k\|_\infty$$

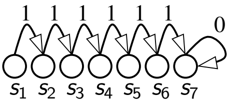
- ALP converges:

$$\|v^* - \tilde{v}\|_1 \leq \frac{2}{1-\gamma} \min_x \|v^* - Mx\|_\infty$$

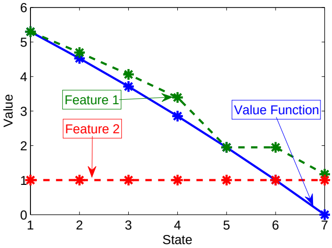
- The error may be too large anyway – high discount factor
- When $\gamma \to 1$ then $\frac{2}{1-\gamma} \to \infty$
- Better bounds with structure, but hard to guarantee Structure

Chain Problem

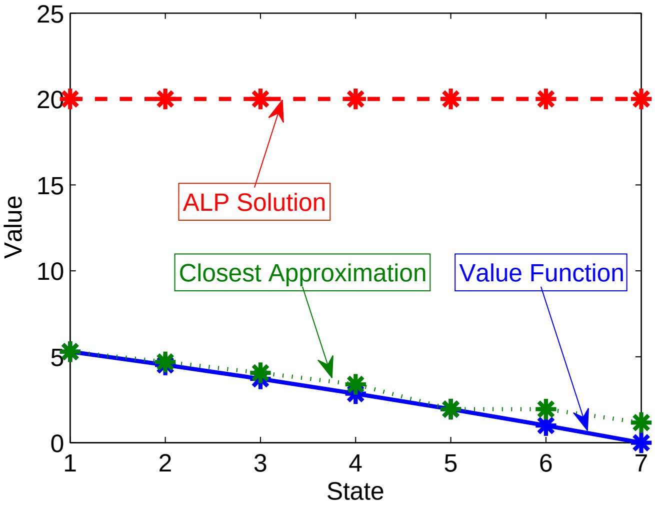
- Chain problem:



- Approximation basis:



Chain Problem: ALP Result



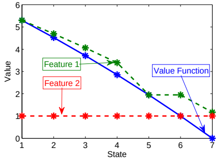
Causes of Large Transitive Error

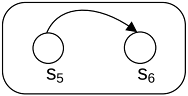
- Presence of a **virtual loop**
  - No loop in original problem
  - Loop when approximated
- Assume $v(s_6) = 0$
- Precise LP constraints:

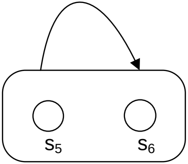
$$v(s_5) \geq \gamma v(s_6) + r$$
$$v(s_5) = r$$

- In the approximation: $v(s_5) = v(s_6)$
- **Approximate** LP constraints:

$$x \geq \gamma x + r$$
$$v(s_5) = x \geq \frac{1}{1-\gamma} r$$







Loops and Dual Variables

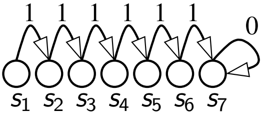
Primal:

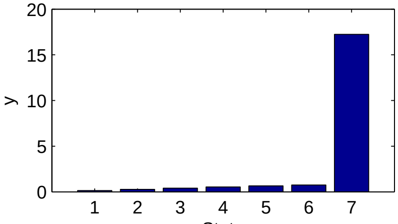
$$\begin{aligned} \min_{v} \quad & c^{\mathsf{T}} v \\ \text{s.t.} \quad & A v \geq b \end{aligned}$$

Dual:

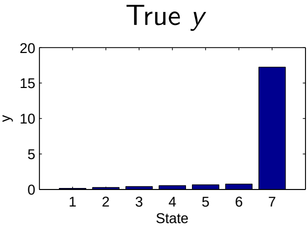
$$\begin{aligned} \max_{y} \quad & b^{\mathsf{T}} y \\ \text{s.t.} \quad & A^{\mathsf{T}} y = c \\ & y \geq 0 \end{aligned}$$

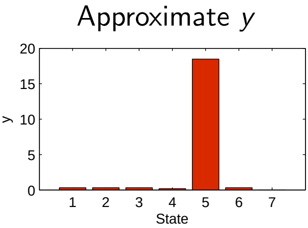
- Dual variable y corresponds to "discounted visitation frequencies"
- Chain example:

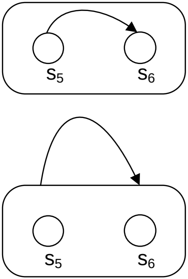


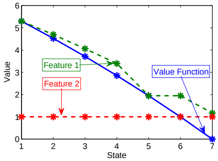


Virtual Loops and Dual Variables





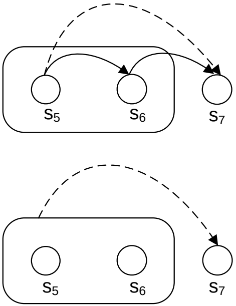


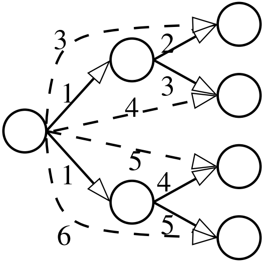


Use dual variables to eliminate virtual loops

Expanding Constraints

- Roll out constraints
- Can “break” virtual loops





Error Bounds

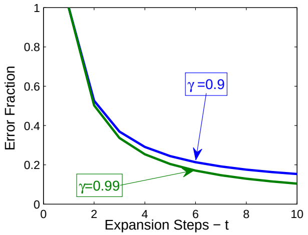
- Assume that $\mathbf{1} \in \operatorname{span} M$
- Constraint expansion lowers the discount factor

Theorem

Let $\tilde{v}_t$ be a solution of a t-step ALP:

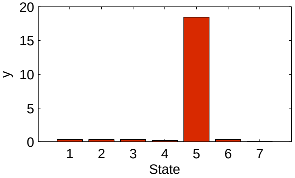
$$\|\tilde{v}_t - v^*\|_{1,c} \leq \frac{2}{1-\gamma^t} \min_x \|v^* - Mx\|_\infty$$

Error reduction with t:



Adaptive Constraint Expansion

- Too many constraints to expand:
  1. Computational problem
  2. Number of samples to bound the approximation error
- Expand only some constraints using y
- Solution of ALP: v
- Solution of **expanded** ALP: $\bar{v}$



Theorem

Improvement from constraint expansion is at most:

$$\|v - v^*\|_{1,c} - \|\bar{v} - v^*\|_{1,c} \le \frac{\| [Av - b]_+ \|_\infty}{1-\gamma} \|y^\mathsf{T} A\|_1$$

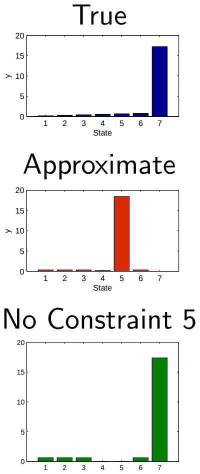
Relaxed Approximate Linear Program

- A few constraints may cause large error
- **Allow limited constraint violation**
- Original linear program:

$$\begin{aligned} \min_{v} \quad & c^{\mathsf{T}} v \\ \text{s.t.} \quad & Av \geq b \end{aligned}$$

- Penalty for constraint violation: d

$$\min_{v} \quad c^{\mathsf{T}} v + d^{\mathsf{T}} [b - Av]_{+}$$



Dual Motivation

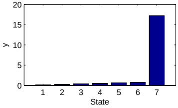
- Offending constraints indicated by large y
- Relaxed ALP:

$$\min_{v} \quad c^{\mathsf{T}} v + d^{\mathsf{T}} [b - Av]_{+}$$

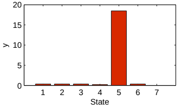
- Dual of relaxed ALP:

$$\begin{aligned} \max_{y} \quad & b^{\mathsf{T}} y \\ \text{s.t.} \quad & A^{\mathsf{T}} y = c \\ & y \geq \mathbf{0} \\ & y \leq d \end{aligned}$$

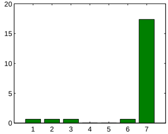
True



Approximate



No Constraint 5



Number of Violated Constraints

- Assume that $\mathbf{1} \in \operatorname{span} M$
- Violated constraints: I_V
- Active constraints: I_A

Theorem

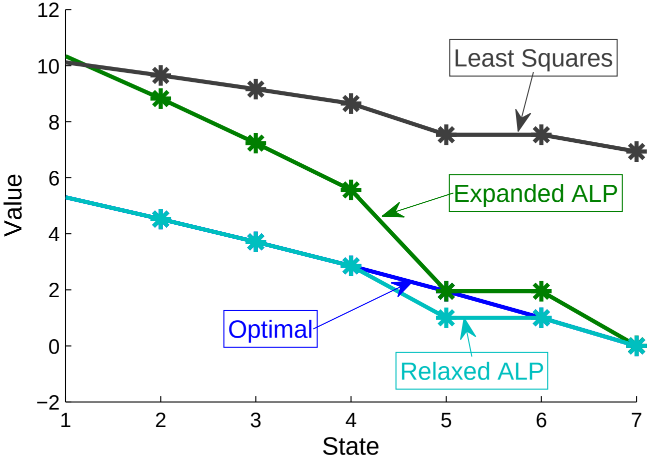
Let $d(\cdot)$ denotes the sum of the weights on the set of constraints:

$$d(I_V) \leq \frac{1}{1-\gamma}$$
$$d(I_A) + d(I_V) \geq \frac{1}{1-\gamma}$$

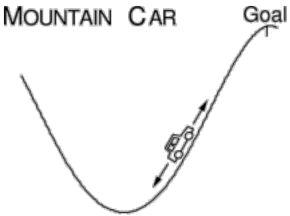
- Guarantee that at most k constraints are violated More Bounds

$$d > \frac{1}{(k+1)(1-\gamma)}\mathbf{1}$$

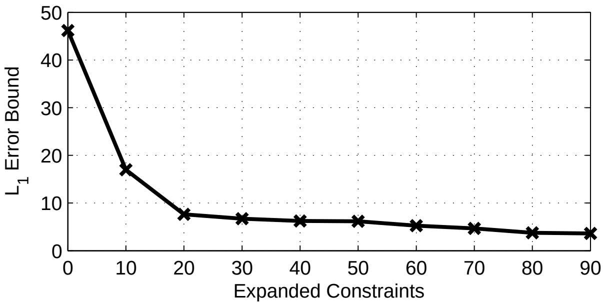
Chain


Least Squares
Expanded ALP
Optimal
Relaxed ALP
Value
State
12
10
8
6
4
2
0
−2
1
2
3
4
5
6
7

Mountain Car

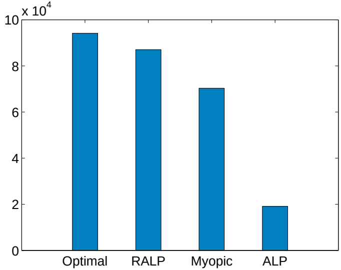


- Underpowered car must climb a hill
- 2-dimensional state space
- Total constraints: 9000

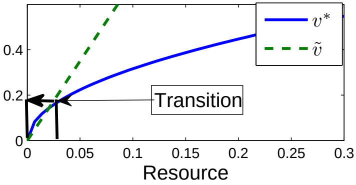


Relaxed ALP: Blood Inventory Management

Results:



- Concave value function
- Piece-wise linear approximation
- ALP is an upper bound on the derivative of the value function



Conclusion

- Approximation error in ALP
  - Representational error
  - Transitional error
  - Sampling error
- Reduction of the transitional error:
  - Constraint expansion
  - Relaxed linear program formulation
- Can significantly improve the ALP performance

Domain Samples

Solution is based on samples of the domain

- Arbitrary goal-terminated paths:

$$(\sigma, a_1), (s_2, a_1), (s_3, a_2), \tau$$

- Optimal goal-terminated paths:

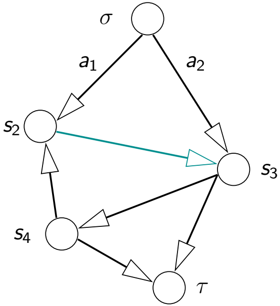
$$(\sigma, a_2), (s_3, a_2), \tau$$

- Transitional samples:

$$(s_2, a_1, s_2)$$

- Expected transitional samples (model) :

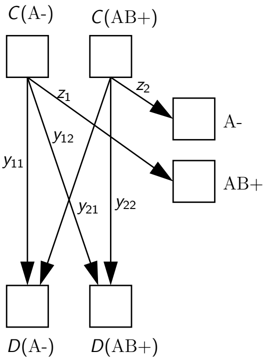
$$(s_2, a_1, \mathbf{E}[s_2])$$



Blood Inventory Management: Greedy Solution

- Finding the best way of using a given inventory – single step
- Actions:
  - y_{ij} – Type i used to satisfy demand for type j
  - z_i – Type i that is retained in inventory
- Solved as a simple flow problem:

$$
\begin{aligned}
\max_{y,z} \quad & \sum_{ij} c_{ij} y_{ij} \\
\text{s.t.} \quad & \sum_{j \in \mathcal{T}} y_{ij} + z_k \leq C(i) \quad \forall i \in \mathcal{T} b \\
& \sum_{i} y_{ij} \leq D(j) \quad \forall j \in \mathcal{T} \\
& y_{ij}, z_i \geq 0 \quad \forall i, j \in \mathcal{T}
\end{aligned}
$$



Lyapunov Hierarchy [?]

Definition

Let $u^1 \dots u^k \geq 0$ be a set of vectors, and A and b be partitioned into A_i and b_i respectively. This set of vectors is called a Lyapunov vector hierarchy if there exist $\beta_i < 1$ such that:

$$\begin{aligned} A_i u^i &\leq \beta_i u^i \\ A_j u^i &\leq 0 \quad \forall j < i \end{aligned}$$

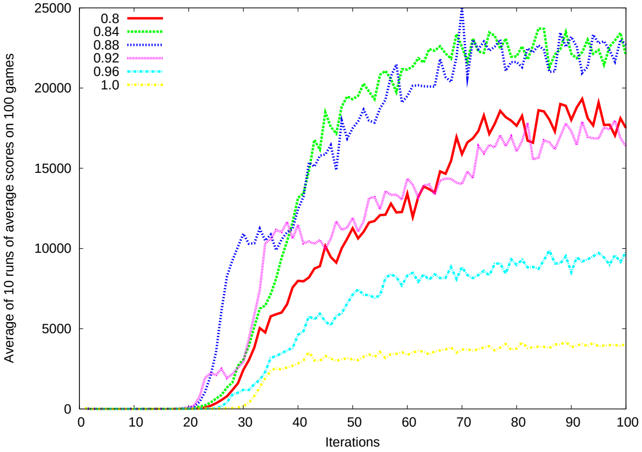
Theorem

Assume that there exists a Lyapunov hierarchy $u^1 \dots u^l \in \mathrm{span}(M)$. *Then:*

$$\|\tilde{v} - v^*\|_\infty \leq \left(1 + \prod_{i=1}^{l} \frac{(1+\alpha\gamma) \max_k u^i(k)}{(1-\gamma\beta_i) \min_k u^i_i(k)}\right) 2 \min_x \|v^* - Mx\|_\infty .$$

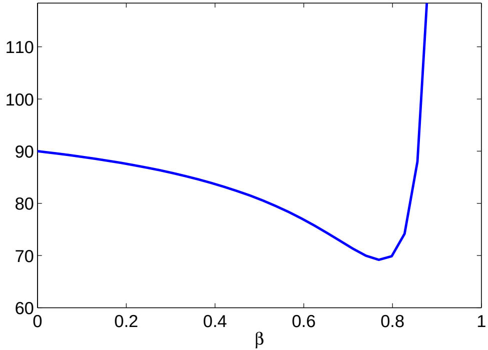
Hard to ensure the hierarchy

Tetris: Effect of Discount Factor [?]



Discount Factor Biasing

Works in problems with sparse rewards



Back

Constraint Formulation Properties

Direct Formulation:

$$v(s) \geq v^*(s)$$

- − Impractical in stochastic problems
- − Many constraints per state: $|\mathcal{A}|^h$
- − Large sampling error
- + Small transitional error

A hybrid approach?

Transitional Formulation:

$$v(s') \geq \gamma \sum_{s \in \mathcal{S}} p(s \mid s', a)\, v(s) + r(s', a)$$

- + Practical in stochastic problems
- + Constraints per state: $|\mathcal{A}|$
- + Small sampling error
- − Large transitional error

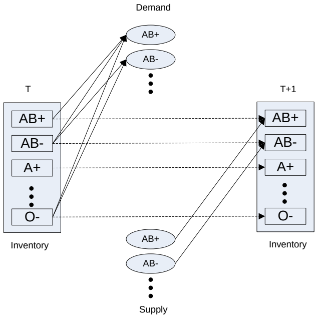
Online Solution Methods

Use value function v to act:

1. Greedy
   - One step lookahead
   - Fixed solution time
   - Solution quality depends on value function v
2. A*
   - Only Deterministic problems
   - Fixed solution quality (optimal if v is admissible)
   - Solution time depends on value function v
3. LAO*
   - Extends A* to stochastic problems
4. Tradeoff
   - Minimize time complexity, satisfying time bound

Blood Inventory Management: MDP Formulation

- Stage = week
- **State:** = (Inventory, Demand)
- **Actions:** How to satisfy supply with
  - Blood type
  - Blood amount
- **Transition function:**
  1. Old blood discarded
  2. New stochastic demand
  3. Stochastic supply added to inventory
- **Reward function:**
  - **Linear** contribution per unit of satisfied blood demand
  - Multiple levels of demand priority



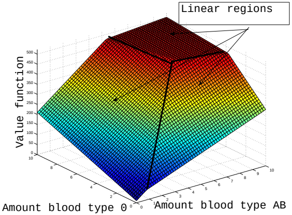
Approximation Basis in Blood Inventory Management

- Defines a set of values for each post-decision state – inventory.
- Structure:
  - Piece-wise linear
  - Fixed regions of linearity
- $M =$

	Feature A	Feature B
A=0, B=1	0	1
A=0, B=2	0	2
A=1, B=0	1	0
A=2, B=0	2	0
A=1, B=1	1	1

- Greedy step be formulated as a flow problem LP

Example value function:



Blood Inventory Management: ALP

- ALP Constraints:

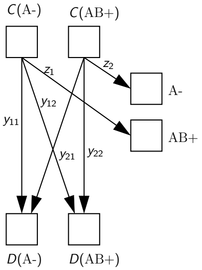
$$v(s') \geq \gamma \sum_{s \in \mathcal{S}} p\left(s \,\middle|\, s', a_1\right) v(s) + r(s', a_1)$$

$$v(s') \geq \gamma \sum_{s \in \mathcal{S}} p\left(s \,\middle|\, s', a_2\right) v(s) + r(s', a_2)$$

- But $|\mathcal{A}| = \infty$; use:

$$v(s_1) \geq \max_{a \in \mathcal{A}} \sum_{s \in \mathcal{S}} p\left(s \,\middle|\, s', a\right) v(s) + r(s', a)$$

- Solutions:
  1. Use flow LP
  2. Use constraint generation – LP to find the most violated constraint

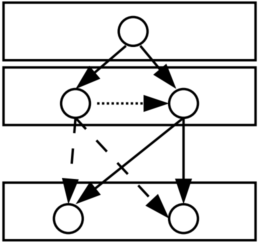


State Of the Art in Solution Techniques

- Operations research:
  - Mature field
  - Focus on specialized problems
  - Mathematical optimization
- Reinforcement learning:
  - Many successful applications
  - Approximate dynamic programming
  - Often need extensive tweaking
- Planning:
  - Branch and bound
  - Heuristic search
- Solved approximately
- **Research Objectives:**
  1. Better understand the tradeoffs involved in the approximation
  2. Develop general methods
  3. Develop robust methods that rely on little tuning

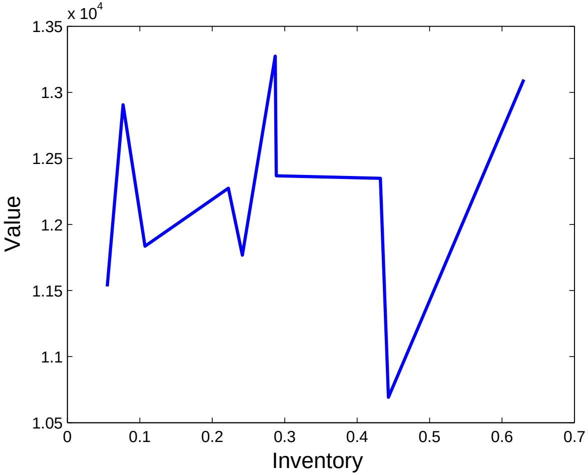
Approximation Basis Structure

- May guarantee that the the transitive error is small
- Examples:
  1. Simple structure: $\mathbf{1} \in \operatorname{span} M$
  2. Smoothness structure: *Lyapunov hierarchy* [**?**] Formal
- Structure hard to guarantee in complex problems
- Solutions
  1. Expand/roll-out selected constraints
  2. Solve a relaxed linear program



Constraint Estimation: Blood Inventory Management

40 samples per constraint



Synchronized Sampling

- Reduce constraint estimation error
- Exploit:
  - Inventory influence mostly independent of the demand and supply
- Use ω to denote the stochastic supply/demand
- $f(s, \omega) =$ the state that follows from s given action a and demand/supply ω

Synchronized Sampling

- Sampled supply/demand: $\omega_1^1, \omega_2^1, \ldots, \omega_1^2, \omega_2^2, \ldots$
- Standard constraint sampling

$$A = \begin{pmatrix} 1 & 0 & 0 & \ldots \\ 0 & 1 & 0 & \ldots \\ & & \vdots & \\ 0 & 0 & 0 & \ldots 1 \end{pmatrix} - \gamma \frac{1}{n} \begin{pmatrix} - & \sum_{j=1}^{n} v(f(s_1, \omega_j^1)) & - \\ - & \sum_{j=1}^{n} v(f(s_2, \omega_j^2)) & - \\ - & \vdots & - \end{pmatrix}$$

- Synchronized constraint sampling

$$A = \begin{pmatrix} 1 & 0 & 0 & \ldots \\ 0 & 1 & 0 & \ldots \\ & & \vdots & \\ 0 & 0 & 0 & \ldots 1 \end{pmatrix} - \gamma \frac{1}{n} \sum_{j=1}^{n} \begin{pmatrix} - & v(f(s_1, \omega_j)) & - \\ - & v(f(s_2, \omega_j)) & - \\ - & \vdots & - \end{pmatrix}$$

Back

ALP Solution Robustness

1. $ALP_1 = (c, A_1, b_1, M)$ with optimal solution v_1
2. $ALP_2 = (c, A_2, b_2, M)$ with optimal solution v_2

Theorem

Also let $\epsilon_a = \|A_1 M - A_2 M\|_{1,\infty}$ *and* $\epsilon_b = \|b_1 - b_2\|_\infty$. *Assuming that* $A_1\mathbf{1} = A_2\mathbf{1} = (1-\gamma)\mathbf{1}$ *then:*

$$\|\tilde{v}_1 - \tilde{v}_2\| \leq \frac{\epsilon_a \hat{x}}{1-\gamma} + \frac{\epsilon_b}{1-\gamma}.$$

- Omitting constraints that are similar does not change the solution
- May use similarity of the transitions

Constraint Estimation

- Constraints in ALP:

$$v(s') \geq \gamma \sum_{s \in \mathcal{S}} p\left(s \,\middle|\, s', a_1\right) v(s) + r(s', a_1) \quad \forall s \in \mathcal{S}$$

- Sample states from the transition probability $s \rightarrow s_1, s_2, \ldots, s_n$
- Constraint:

$$\begin{aligned} v(s) &\geq \gamma P_a v + r_a = \gamma \mathrm{E}_S \left[v(S)\right] + r_a \\ &\approx \gamma \frac{1}{n} \sum_{j=1}^{n} v(s_j) + r_a \end{aligned}$$

- For sufficiently large n, the error is sufficiently small
- The number of samples depends on the number of features in the ALP

Constraint Estimation Error

Theorem

Let v_1 be the solution of the true ALP_1 and let v_2 be the solution of the sampled ALP_q. Then:

$$\mathbf{P}\left[\|v_1 - v_2\|_{1,c} \geq \epsilon\right] \leq nm \exp\left(-\frac{2q\epsilon^2 m^2 (1-\gamma)^2}{\hat{x}^2}\right) + \\ + n \exp\left(-\frac{2q\epsilon^2 (1-\gamma)^2}{\|r\|_\infty^2}\right),$$

where $\hat{x} \geq |x(i)|$ for all i assuming that $\|M\|_\infty = 1$.

Total Constraint Violation

- Let

$$\min_{v \in \operatorname{span} M} \|v - v^*\|_\infty \leq \epsilon$$

- Minimizer $\hat{v}$
- Constraint violation penalty:

$$d = y^* + \Delta d$$

Theorem

Let $\tilde{v}$ be the optimal solution of the relaxed ALP, then:

$$\|[b - A\tilde{v}]_+\|_{1,\Delta d} \leq (2 + \Delta d^{\mathsf{T}} \mathbf{1})\epsilon.$$

- If $v^* \in \operatorname{span} M$ then $\tilde{v} = v^*$
- Proof differs from other ALP bounds
- Cannot use that $\tilde{v}$ is an upper bound on v^*

Objective Function

- L_1 minimization:
  - Problem with the nonlinearity of the absolute value
  - Possible when $v \geq v^*$:

$$\|v - v^*\|_1 = \sum_{s \in \mathcal{S}} |v(s) - v^*(s)| = \sum_{s \in \mathcal{S}} v(s) - v^*(s)$$

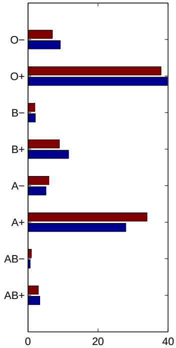
  - Constants can be ignored:

$$\arg\min_v \sum_{s \in \mathcal{S}} v(s) - v^*(s) = \arg\min_v \sum_{s \in \mathcal{S}} v(s)$$

- Possible to bound the policy error

Considerations

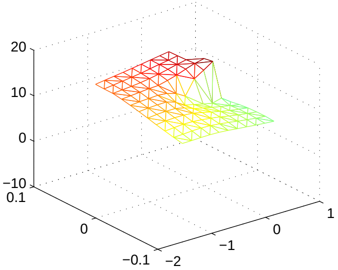
- Demand and supply of blood are **stochastic**
- Blood is perishable
- Multiple blood types are compatible
- Blood type distribution: Supply $\neq$ Demand
- Manage how much of which blood is:
  1. Used to satisfy the demand
  2. Retained in inventory
- Challenging optimization problem:
  - Continuous action space
  - 48-dimensional continuous state space
  - High level of stochasticity



Solution Methods

Mountain Car Value Function

Unexpanded:



Expanded 10 steps: