Dynamic programming is widely applicable in principle, but often difficult to apply in practice.

- Dynamic programming is widely applicable because the world is full of sequential decision problems with uncertainty
- Dynamic programming is often difficult to apply because of:
  - The “curse of dimensionality” – the state space grows exponentially in the number of variables
  - The challenge of modeling the gradual resolution of uncertainty
- Monte Carlo simulation (MCS) is easy to apply with large problems:
  - Can simulate to find value with a given policy
  - Hard to find an optimal policy or know how a given policy compares to the optimal policy
Our goal is to develop tools for studying large-scale DPs.

Modern approaches to large-scale DPs include:

- **Neurodynamic programming**: Use simulation methods and linear approximations of value functions
  - Bertsekas and Tsitsiklis (1996); Powell (2007) - General theory
  - Van Roy and Tsitsiklis (2001), Longstaff and Schwartz (2001) - Option valuation

- **Linear programming techniques with constraint generation**:
  - de Farias and B. Van Roy (2003, 2004, ...) - General theory
  - Adelman (forthcoming) - Revenue management

We consider a dual approach based on information relaxations.

- A flexible framework that allows use of MCS with DPs:
  - Provides upper bounds on DP values
  - Users can control the computational effort vs. accuracy tradeoff

  Use with lower bounds from approximate policies to bracket optimal solution

- Inspired by Haugh and Kogan (2004) who developed a similar dual approach for option pricing problems; see also Rogers (2002) and Anderson and Broadie (2004).

- We generalize to consider:
  - General DPs
  - Arbitrary information relaxations
  - Linear programming duality rather than martingales
The Basic Idea (1):

- □ A good structure for MCS:

```
<table>
<thead>
<tr>
<th>Decision 1</th>
<th>Chance 1</th>
<th>Chance 2</th>
<th>Chance 3</th>
<th>Chance 4</th>
<th>Chance 5</th>
<th>Chance 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alt A</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Alt B</td>
<td>Nominal</td>
<td>Nominal</td>
<td>Nominal</td>
<td>Nominal</td>
<td>Nominal</td>
<td>Nominal</td>
</tr>
<tr>
<td>Alt C</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>
```

- □ A dynamic program, not a good structure for MCS:

```
<table>
<thead>
<tr>
<th>Decision 1</th>
<th>Chance 1</th>
<th>Decision 2</th>
<th>Chance 2</th>
<th>Decision 3</th>
<th>Chance 3</th>
<th>Decision 4</th>
<th>Chance 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alt A</td>
<td>Low</td>
<td>Alt A</td>
<td>Low</td>
<td>Alt A</td>
<td>Low</td>
<td>Alt A</td>
<td>Low</td>
</tr>
<tr>
<td>Alt B</td>
<td>Nominal</td>
<td>Nominal</td>
<td>Nominal</td>
<td>Nominal</td>
<td>Nominal</td>
<td>Nominal</td>
<td>Nominal</td>
</tr>
<tr>
<td>Alt C</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>
```

To determine the optimal policy, you need to calculate expected values conditional on the outcomes of earlier decisions and uncertainties.

The Basic Idea (2):

- □ The DP with additional information and a penalty:

```
<table>
<thead>
<tr>
<th>Chance 1</th>
<th>Chance 2</th>
<th>Chance 3</th>
<th>Chance 4</th>
<th>Decision 1</th>
<th>Decision 2</th>
<th>Decision 3</th>
<th>Decision 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Alt A</td>
<td>Alt A</td>
<td>Alt A</td>
<td>Alt A</td>
</tr>
<tr>
<td>Nominal</td>
<td>Nominal</td>
<td>Nominal</td>
<td>Nominal</td>
<td>Alt B</td>
<td>Alt B</td>
<td>Alt B</td>
<td>Alt B</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Alt C</td>
<td>Alt C</td>
<td>Alt C</td>
<td>Alt C</td>
</tr>
</tbody>
</table>
```

- Additional information adds value; penalties take it away.
- With the right penalty, we can solve the DP.
- With a feasible penalty, we get an upper bound on the DP value.

⇒ A good structure for simulation!
Related Literature:

- Relaxations of temporal feasibility constraints have been considered in the stochastic programming literature:
  - E.g., Rockafellar and Wets (1991); Shapiro and Ruszczyński (2007)
  - Requires concave “reward” functions and convex action spaces
  - Considers perfect info relaxations and linear penalties
  - Uses sophisticated convex duality arguments

- Rogers (2007) "Pathwise Stochastic Optimal Control"
  - Perfect information relaxation
  - Restricted forms for penalties

- Our results are simpler and more general
  - General rewards and penalties
  - Simple direct proofs

Outline:

- Introduction
  ⇒ Basic setup: The primal DP

- Duality results:
  - Penalties and relaxations
  - Linear programming duality results
  - “Good” penalties

- Example: Inventory control

- Example: Option pricing with stochastic volatility

- Example: Sequential exploration

- Conclusions
Basic Setup

- Discrete time; finite horizon: $t = 0, \ldots, T$

- Uncertainty is described by a probability space $(\Omega, \mathcal{F}, \mathbb{P})$
  - DM's information described by a filtration $\mathcal{F} = \{\mathcal{F}_t\}$ where $\mathcal{F}_t$ describes the DM's information at the beginning of period $t$
  - No forgetting: $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}_T \subseteq \mathcal{F}$

- DM will choose an action $a_t$ in period $t$ from $A_t$;
  $A \subseteq A_0 \times \cdots \times A_T$ = the set of all possible action vectors $a$

- Rewards $\{r_t(a, \omega)\}$ depend on the actions $a \in A$ and state $\omega \in \Omega$;
  Total reward: $r(a, \omega) = \sum_{t=0}^{T} r_t(a, \omega)$ ($r(a)$ is a random variable.)

The Primal DP

- A policy $\alpha$ in $A$ selects an action vector $a \in A$ for each state $\omega \in \Omega$.

- In the primal DP, policies must be temporally feasible or $\mathbb{F}$-adapted:
  - Choice of action in period $t$ can only depend on what is known then
  - Let $A_{\mathbb{F}} \subseteq A$ denote the set of $\mathbb{F}$-adapted policies

- The primal DP:

$$\sup_{\alpha \in A_{\mathbb{F}}} \mathbb{E}[r(\alpha)] = \sup_{\alpha \in A_{\mathbb{F}}} \mathbb{E}[r(\alpha(\tilde{\omega}), \tilde{\omega})]$$

*If we allow mixed policies, the primal DP is a linear program with mixing probabilities as decision variables.*
The Dual Approach: Information Relaxations

- A filtration $\mathcal{G} = \{G_t\}$ is a relaxation of filtration $\mathcal{F} = \{F_t\}$ if, for each $t$, $F_t \subseteq G_t \subseteq \mathcal{F}$.

The DM “knows more” in every period under $\mathcal{G}$ than under $\mathcal{F}$.

- Perfect information relaxation: $\mathcal{I} = \{I_t\}$ where $I_t = \mathcal{F}$

- Let $\mathcal{A}_G$ = the set of $\mathcal{G}$-adapted policies. For any relaxation $\mathcal{G}$ of $\mathcal{F}$,

$$\mathcal{A}_F \subseteq \mathcal{A}_G \subseteq \mathcal{A}_I = \mathcal{A}$$

⇒ As we relax the filtration, we expand the set of feasible policies.

Example: Standard Inventory Control Model

- Discrete time: $t = 0, \ldots, T$
  - Order $a_t$ units of good in period $t$; $a_t \geq 0$
  - Demand in period $t$, $d_t$, is uncertain; revealed after ordering
  - Inventory at the end of period $t$ given by $x_{t+1} = x_t + a_t - d_t$
    (Assumes unmet demand is backordered.)

- Costs: $r_t(a, d) = c_t(a_t) + f_t(x_{t+1})$

- Primal DP: $\inf_{\alpha_F \in \mathcal{A}_F} \mathbb{E} \left[ \sum_{t=0}^{T} r_t(\alpha_F, d) \right]$

- Perfect Information Dual DP: $\mathbb{E} \left[ \inf_{\alpha \in \mathcal{A}} \left\{ \sum_{t=0}^{T} (r_t(a, d) - z_t(a, d)) \right\} \right]$. 
The Dual Approach: Penalties

- Penalties $z(a, \omega)$, like rewards, depend on the choice of actions $a$ and the state $\omega$.

- Penalties will be used to “punish” policies that use the additional information in $G$. Penalized objective function: $\mathbb{E} [r(\alpha_G) - z(\alpha_G)]$

- A penalty $z$ is dual feasible if
  $$\mathbb{E} [z(\alpha_F)] \leq 0 \text{ for all } \alpha_F \text{ in } A_F.$$  

  A dual feasible penalty assigns zero (or negative) expected penalty to any temporally feasible policy.

- Let $Z_F = \text{set of all dual feasible penalties}$

The Dual Approach: Weak Duality

- Our main result can be viewed as a version of the “weak duality lemma” of linear programming.

  **Lemma.** If $\alpha_F$ and $z$ are primal and dual feasible respectively and $G$ is a relaxation of $F$, then
  $$\mathbb{E} [r(\alpha_F)] \leq \sup_{\alpha_G \in A_G} \mathbb{E} [r(\alpha_G) - z(\alpha_G)].$$
The Dual Approach: Weak Duality

□ Our main result can be viewed as a version of the “weak duality lemma” of linear programming.

Lemma. If $\alpha_F$ and $z$ are primal and dual feasible respectively and $G$ is a relaxation of $F$, then

$$\mathbb{E} [r(\alpha_F)] \leq \sup_{\alpha_G \in \mathcal{A}_G} \mathbb{E} [r(\alpha_G) - z(\alpha_G)].$$

Proof: $\mathbb{E} [r(\alpha_F)] \leq \mathbb{E} [r(\alpha_F) - z(\alpha_F)]$

($\mathbb{E} [z(\alpha_F)] \leq 0$ because $z \in \mathcal{Z}_F$ and $\alpha_F \in \mathcal{A}_F$)
The Dual Approach: Weak Duality

□ With the perfect information relaxation, weak duality implies that for any \( \alpha_F \) in \( A_F \) and \( z \in Z_F \),

\[
\mathbb{E} \left[ r(\alpha_F) \right] \leq \mathbb{E} \left[ \sup_{a \in A} \{ r(a) - z(a) \} \right].
\]

This form is convenient for Monte Carlo simulation: In each trial,

- Randomly generate state \( \omega \)
- Solve a deterministic “inner problem” of choosing a vector of actions \( a \) to maximize \( r(a, \omega) - z(a, \omega) \)

□ With imperfect information relaxations, we can often use MCS but the “inner problems” may be stochastic

The Basic Idea (revisited):

□ Primal DP:

□ Dual DP with perfect information relaxation:

Imperfect information relaxations can be viewed as moving some uncertainties to the left.
The Dual Approach: Strong Duality

In principle, we can find the optimal value by choosing the right penalty.

**Theorem.** Let $G$ be a relaxation of $F$. Then

$$\sup_{\alpha_F \in A_F} \mathbb{E}[r(\alpha_F)] = \inf_{z \in Z_F} \left\{ \sup_{\alpha_G \in A_G} \mathbb{E}[r(\alpha_G) - z(\alpha_G)] \right\}.$$ 

If the primal problem on the left is bounded, the dual problem on the right has an optimal solution $z^* \in Z_F$ that achieves this bound.

**Proof:** Let $z^*(a) = r(a) - v^*$ where $v^* = \sup_{\alpha_F \in A_F} \mathbb{E}[r(\alpha_F)]$.

Then $z^*$ is dual feasible and $r(a) - z^*(a) = v^*$.

---

The Dual Approach: Complementary Slackness

The complementary slackness condition further characterizes optimal policies and penalties.

**Theorem.** Let $\alpha^*_F$ and $z^*$ be feasible solutions for the primal and dual problems respectively with information relaxation $G$. A necessary and sufficient condition for these to be optimal solutions for their respective problems is:

$$\mathbb{E}[z^*(\alpha^*_F)] = 0 \quad \text{and} \quad \mathbb{E}[r(\alpha^*_F) - z^*(\alpha^*_F)] = \sup_{\alpha_G \in A_G} \{\mathbb{E}[r(\alpha_G) - z^*(\alpha_G)]\}$$

$\Rightarrow$ With an optimal penalty $z^*$, $\alpha^*_F$ has zero expected penalty and solves the dual problem.
The Dual Approach: Structured Policies

□ If the optimal policies in the primal problem have some particular structure (thresholds, Markovian, stationary, ...), we can focus on policies with this structure in the dual.

**Proposition.** Let $S \subseteq A$ and suppose $\alpha^*_F \in A_F \cap S$ is a primal optimal solution. Then weak and strong duality hold within this restricted policy set $S$. Moreover, for a fixed penalty $z$ in

$$Z^S_F = \{ z \in Z : \mathbb{E}[z(\alpha_F)] \leq 0 \quad \forall \alpha_F \in A_F \cap S \}$$

the bound with the policy restriction to $S$ is weakly tighter than the corresponding bound without the restriction.

⇒ Structural constraints on policies lead to tighter bounds and more efficient computation.

How to Create “Good” Penalties:

□ Let $G$ be a relaxation of $F$; let \{w_t(a, \omega)\} be a sequence of “generating functions” where each $w_t$ depends only on the first $t + 1$ actions $(a_0, ..., a_t)$ of $a$.

□ Define $z_t(a) = \mathbb{E}[w_t(a)|G_t] - \mathbb{E}[w_t(a)|F_t]$ and $z(a) = \sum_{t=0}^{T} z_t(a)$.

**Proposition.** For these penalties $z$:

(i) $\mathbb{E}[z(\alpha_F)] = 0$ for all $\alpha_F$ in $A_F$ (Dual feasible; potentially optimal);

(ii) \{z_t(a)\} is adapted to $G$ and $z_t$ depends only on the first $t + 1$ actions $(a_0, ..., a_t)$ of $a$ (Leads to DP structure on $G$); and

(iii) $z$ is orthogonal to $F$ in that $\mathbb{E}[z_t(a)|H] = 0$ whenever $H$ is in $F_t$. (Depends only on the additional information in $G$)
Examples of “Good” Penalties:

- Define the DP value function:
  - Let $A^t_F(a)$ be the $a$ in $A^t_F$ with the first $t$ actions $(a_0, \ldots, a_{t-1})$ constrained to match the first $t$ actions in $a$
  - Define $V_t(a) = \sup_{\alpha \in A^t_F(a)} \mathbb{E}[r(\alpha)|F_t]$

- An optimal penalty: Take $z^*_t(a) = \mathbb{E}[V_{t+1}(a)|G_t] - \mathbb{E}[V_{t+1}(a)|F_t]$
  $\Rightarrow z^*_t(a)$ cancels the benefit of the knowledge in $G$ in each period

- Alternatively, replace $V$ with:
  - An approximate value function (Haugh and Kogan)
  - A value function based on a given policy (Anderson and Broadie)
  - Some simple approximation

Example: Inventory Control – Penalties

- Perfect information bound: $z_t = 0$
  Inner problem is a deterministic dynamic lot sizing problem:
  $$\inf_{a \in A} \sum_{t=0}^{T} (c_t(a_t) + f_t(x_{t+1}))$$
  Easy to solve, esp. with linear ordering, holding and shortage costs

- Smooth bound: $z_t(a, d) = r_t(a, d) - \mathbb{E}[r_t(a, d)|F_t]$
  Inner problem as before but with a "smoothed" cost function:
  $$\inf_{a \in A} \sum_{t=0}^{T} (c_t(a_t) + \mathbb{E}[f_t(x_{t+1})|F_t])$$
  With linear costs, if we drop $a_t \geq 0$ constraints, we get a weaker bound that decomposes into a series of simple newsvendor problems.
Properties of Penalties:

- Penalties that are close to each other (or close to optimal) give bounds that are close to each other (or close to optimal):

\[
\sup_{\alpha G \in A_G} \mathbb{E}[r(\alpha G) - z_1(\alpha G)] \leq \sup_{\alpha G \in A_G} \mathbb{E}[r(\alpha G) - z_2(\alpha G)] + \sup_{\alpha G \in A_G} \mathbb{E}[z_2(\alpha G) - z_1(\alpha G)]
\]

- In practice, there will typically be a tradeoff between:
  - The accuracy of the bound and
  - The computational effort required to compute it.

We can control this tradeoff through our choice of information relaxation $G$ and penalty $z$.

Outline:

- Introduction
- Basic setup: The primal DP
- Duality results:
  - Penalties and relaxations
  - Linear programming duality results
  - “Good” penalties
- Example: Inventory control
- Example: Option pricing with stochastic volatility
- Example: Sequential exploration
- Conclusions
Example: Inventory Control – Numerical Example; Assumptions

- From Zipkin (2000)
- Linear costs; 5 Periods
- Detailed Assumptions:

<table>
<thead>
<tr>
<th>Period</th>
<th>Nonstationary Demand and Stationary Costs</th>
<th>Stationary Demand and Nonstationary Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ordering costs</td>
<td>$k_t$</td>
<td>2</td>
</tr>
<tr>
<td>Holding costs</td>
<td>$h_t$</td>
<td>1</td>
</tr>
<tr>
<td>Backorder costs</td>
<td>$p_t$</td>
<td>9</td>
</tr>
<tr>
<td>Mean demand</td>
<td>$E(d_t)$</td>
<td>40</td>
</tr>
</tbody>
</table>

- We consider Poisson and geometric demand distributions

Example: Inventory Control – Numerical Example; Results

<table>
<thead>
<tr>
<th>Poisson Demand Distributions</th>
<th>Nonstationary Demand and Stationary Costs</th>
<th>Stationary Demand and Nonstationary Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal value</td>
<td>293.94</td>
<td>-</td>
</tr>
<tr>
<td>Smooth bound</td>
<td>292.58</td>
<td>0.20</td>
</tr>
<tr>
<td>Newsvendor bound</td>
<td>287.84</td>
<td>-</td>
</tr>
<tr>
<td>Perfect info. bound</td>
<td>248.00</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Geometric Demand Distributions</th>
<th>Nonstationary Demand and Stationary Costs</th>
<th>Stationary Demand and Nonstationary Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal value</td>
<td>607.14</td>
<td>-</td>
</tr>
<tr>
<td>Smooth bound</td>
<td>599.85</td>
<td>1.22</td>
</tr>
<tr>
<td>Newsvendor bound</td>
<td>538.99</td>
<td>-</td>
</tr>
<tr>
<td>Perfect info. bound</td>
<td>248.00</td>
<td>-</td>
</tr>
</tbody>
</table>
Example: Inventory Control – Dual Policies (Poisson)

Y-axis = Period-\(t\) order quantity, X-axis = Period-\(t\) inventory level; Red line = Optimal Policy

<table>
<thead>
<tr>
<th>Period</th>
<th>( t )</th>
<th>Nonstationary Demand and Stationary Costs</th>
<th>Stationary Demand and Nonstationary Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Ordering costs ( k_t )</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Holding costs ( h_t )</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Backorder costs ( p_t )</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Mean demand ( \mathbb{E}[d_t] )</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
</tbody>
</table>

Example: Inventory Control – Dual Policies (Geometric)

Y-axis = Period-\(t\) order quantity, X-axis = Period-\(t\) inventory level; Red line = Optimal Policy

<table>
<thead>
<tr>
<th>Period</th>
<th>( t )</th>
<th>Nonstationary Demand and Stationary Costs</th>
<th>Stationary Demand and Nonstationary Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Ordering costs ( k_t )</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Holding costs ( h_t )</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Backorder costs ( p_t )</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Mean demand ( \mathbb{E}[d_t] )</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
</tbody>
</table>
Example: Inventory Control – Nonstationary Demand and Partial Information

- Suppose ordering costs and demand distribution evolve stochastically following exogenous Markov chain.

- Ordering costs are observed before ordering.

- Demand distribution is not directly observable.

\[
\begin{array}{ccc}
20 & 0.3 & 30 \\
0.7 & 0.2 & 0.6 \\
2.0 & & 0.7
\end{array}
\]

Mean Demands $E[\text{d}]$

\[
\begin{array}{ccc}
1.2 & 0.3 & 2.0 \\
0.7 & 0.2 & 0.6 \\
2.8 & & 0.7
\end{array}
\]

Ordering Costs $k_i$

Prior belief on potential distributions.
Bayesian updates after observe demand

Example: Inventory Control – Nonstationary Demand and Partial Information

- Exponential growth of state space: Under $\mathbb{F}$, prior belief distribution is in the state information.

- Imperfect information relaxation: Under $\mathbb{G}$, DM knows the whole demand sequence and current ordering cost.

  Inner problem $\Rightarrow$ a simpler stochastic DP.

- Penalty through generating function $w$:
  - Smooth bound
  - Short horizon value function

- Upper bound (evaluation purpose): various heuristic (Treharne and Sox, 2002).
Example: Valuing an Option with Stochastic Volatility

- We will value put and call options on a dividend-paying stock
- Heston's (1993) stochastic volatility model, plus dividends:

\[
\begin{align*}
\text{Stock Price } S_t : & \quad dS_t = \mu S_t dt + \sqrt{v_t} S_t d\xi_t \\
\text{Instant. Var. } v_t : & \quad dv_t = \kappa (\bar{v} - v_t) dt + \sigma \sqrt{v_t} d\nu_t \\
\text{Dividends } D_{t_i} : & \quad S_{t_i} = S_{t_{i-1}} - D_{t_i}
\end{align*}
\]

where \( d\xi_t \) and \( d\nu_t \) are increments of a Brownian motion process with correlation \( \rho \) (Risk-neutral process)

- Base filtration \( \mathbb{F} \) assumes:
  - \( S_t \) and \( v_t \) are known at time \( t \) (Could also take \( v_t \) to be unknown)
  - Dividend times \( (t_1, \ldots, t_m) \) and amounts \( D_{t_i} \) are given

---

Example: Valuing an Option with Stochastic Volatility

- Actions: \( a_t = \text{exercise} \) or \( \text{not} \); can exercise at most once
- Call option rewards:

\[
E(t, a_t, S_t, v_t) = \begin{cases} 
  e^{-rt} (S_t - K) & \text{if } a_t = \text{exercise} \\
  0 & \text{otherwise}
\end{cases}
\]

where \( K \) is the strike price and \( r \) the risk-free interest rate.
(Rewards for a put option are the negative of this.)

- Primal DP: Fix a finite set of possible exercise dates \( \{t_0, t_1, \ldots, t_T\} \):

\[
\sup_{\alpha_F \in A_F} \mathbb{E} \left[ \sum_{i=0}^{T} r_{t_i}(\alpha_{t_i}, S_{t_i}, v_{t_i}) \right]
\]
Example: Valuing an Option with Stochastic Volatility

- We consider a relaxation $\mathbb{G}$ where all volatilities are known in advance. Solved by simulation:
  - Generate all volatilities $\mathbf{v} = \{v_t\}_{t=0}^T$
  - Given $\mathbf{v}$, future stock prices are log-normally distributed with
    \[
    \mathbb{E}[S_{t_2}|S_{t_1}, \mathbf{v}] = S_{t_1}e^{\mu(t_2-t_1)} \exp \left( \rho \int_{t=t_1}^{t_2} \sqrt{v_t} d\nu_t - \frac{\rho^2}{2} \int_{t=t_1}^{t_2} v_t dt \right).
    \]
  - Solve inner problem: One-dimensional option valuation problem with time-varying volatility; we used a trinomial lattice.

- Structural restriction for call options:
  - Under $\mathbb{F}$, early exercise occurs only before dividends.
  - Not true under $\mathbb{G}$, but we can require this in the dual.

Example: Valuing an Option with Stochastic Volatility – Penalties

- Generating function for period $i$ (time $t_1$ to $t_2$)
  \[
  w_i(a) = \begin{cases} 
  0 & \text{if } a_i = \text{exercise} \\
  \Delta(S_{t_1}, t_1) \left( e^{\nu(t_2-t_1)} S_{t_2} \right) & \text{if } a_i = \text{do not exercise}
  \end{cases}
  \]
  - $\Delta(S_t, t)$ is the “delta” for an option with constant volatility ($\bar{v}$)
  - Easy to compute in the trinomial lattice; calculate once.

- When $a_i = \text{exercise}$, $z_i(a) = 0$; When $a_i = \text{do not exercise}$,
  \[
  z_i(a) = \mathbb{E}[w_i(a)|\mathcal{G}_{t_i}] - \mathbb{E}[w_i(a)|\mathcal{F}_{t_i}]
  = \Delta(S_{t_1}, t_1) \left( e^{\nu(t_2-t_1)} \mathbb{E}[S_{t_2}|S_{t_1}, \mathbf{v}] - S_{t_1} \right)
  \]

- Approximately cancels the effect of knowing future volatilities on the stock price expectations.
Example: Valuing an Option with Stochastic Volatility – Results

- Put and call options, expire in 1 year
- $S_0 = K = $100; \kappa = 0.2; \sigma = 10\%; r = \mu = 0\%; v_0 = \bar{v} = 25\%$
- One $2.50$ dividend after 6 months – sometimes optimal to exercise

Lower bounds found by simulating with a fixed exercise threshold

<table>
<thead>
<tr>
<th></th>
<th>$\rho = 0$ Mean</th>
<th>$\rho = 0.25$ Mean</th>
<th>$\rho = 0.5$ Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Std Error</td>
<td>Std Error</td>
<td>Std Error</td>
</tr>
<tr>
<td>Call Options</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With No Penalty</td>
<td>8.842</td>
<td>9.281</td>
<td>9.904</td>
</tr>
<tr>
<td>With Penalty</td>
<td>8.842</td>
<td>8.869</td>
<td>8.896</td>
</tr>
<tr>
<td>Lower Bound</td>
<td>8.839</td>
<td>8.872</td>
<td>8.892</td>
</tr>
<tr>
<td>Put Options</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With No Penalty</td>
<td>11.128</td>
<td>12.006</td>
<td>13.304</td>
</tr>
<tr>
<td>With Penalty</td>
<td>11.128</td>
<td>11.158</td>
<td>11.225</td>
</tr>
<tr>
<td>Lower Bound</td>
<td>11.128</td>
<td>11.154</td>
<td>11.174</td>
</tr>
</tbody>
</table>

$\Rightarrow$ Stochastic volatility affects option values, but you don’t need to solve a two-dimensional DP to pin down the value.

Example: Valuing an Option with Stochastic Volatility – Dual Policies

- Call option with $\rho = 0.5$; Actions chosen just before dividend in dual problem, by scenario:
Example: Sequential Exploration

- Given $n$ (dependent) wells, what is the optimal drilling strategy?

- Wells must succeed on each of $k$ underlying factors to be “wet”

- From Bickel and Smith (2006); Bickel, Smith and Meyer (2007)

Example: Sequential Exploration

- The primal problem gets complex quickly.

<table>
<thead>
<tr>
<th>Wells (n)</th>
<th>Actions 1</th>
<th>States 1</th>
<th>Actions 2</th>
<th>States 2</th>
<th>Actions 3</th>
<th>States 3</th>
<th>Actions 4</th>
<th>States 4</th>
<th>(Any k)</th>
<th>Actions</th>
<th>States</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>10</td>
<td>9</td>
<td>18</td>
<td>17</td>
<td></td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>9</td>
<td>35</td>
<td>25</td>
<td>99</td>
<td>81</td>
<td>323</td>
<td>289</td>
<td>32</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>54</td>
<td>27</td>
<td>200</td>
<td>125</td>
<td>972</td>
<td>729</td>
<td>5,780</td>
<td>4,913</td>
<td>20</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>189</td>
<td>81</td>
<td>1,125</td>
<td>625</td>
<td>9,477</td>
<td>6,561</td>
<td>103,173</td>
<td>83,521</td>
<td>48</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>648</td>
<td>243</td>
<td>6,250</td>
<td>3,125</td>
<td>91,854</td>
<td>59,049</td>
<td>1,837,462</td>
<td>1.42E+06</td>
<td>112</td>
<td>32</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>2,187</td>
<td>729</td>
<td>34,375</td>
<td>15,625</td>
<td>8.96E+05</td>
<td>531,441</td>
<td>3.27E+07</td>
<td>2.41E+07</td>
<td>256</td>
<td>64</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>7,290</td>
<td>2,187</td>
<td>187,500</td>
<td>78,125</td>
<td>8.50E+06</td>
<td>4.79E+06</td>
<td>5.79E+08</td>
<td>4.10E+08</td>
<td>576</td>
<td>128</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>24,057</td>
<td>6,561</td>
<td>1,02E+09</td>
<td>390,625</td>
<td>8.13E+07</td>
<td>4.30E+07</td>
<td>1.03E+10</td>
<td>6.98E+09</td>
<td>1,280</td>
<td>256</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>78,732</td>
<td>19,683</td>
<td>5.47E+06</td>
<td>1.95E+06</td>
<td>7.75E+08</td>
<td>3.87E+08</td>
<td>1.81E+11</td>
<td>1.19E+11</td>
<td>2,816</td>
<td>512</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>255,879</td>
<td>59,049</td>
<td>2.93E+07</td>
<td>9.77E+06</td>
<td>7.36E+09</td>
<td>3.49E+09</td>
<td>3.20E+12</td>
<td>2.02E+12</td>
<td>6,144</td>
<td>1,024</td>
<td>4</td>
</tr>
</tbody>
</table>

- Perfect information dual:
  - Randomly generate well results, then choose drilling order.
  - Inner problem grows exponentially with the number of wells $n$ (albeit slowly), but is independent of the number of factors $k$.  

Example: Sequential Exploration – Relaxations and Penalties

☐ We consider a numerical example with 5 wells and 3 factors
  – 59,049 states; 91,854 actions
  – Numeric assumptions from Bickel, Smith and Meyer (2007)

☐ We consider information relaxations that provide perfect information in
  periods 0, 1, . . . 5.

☐ Penalties:
  
  No Penalty: \( z_t(a) = 0 \)
  
  Myopic Penalty: \( z_t(a) = r_t(a) - \mathbb{E}[r_t(a)|\mathcal{F}_t] \)
  
  Binary VF Penalty: \( z_t(a) = \hat{V}_t(a) - \mathbb{E}[\hat{V}_t(a)|\mathcal{F}_t] \)

  where \( \hat{V}_t \) is the value function for a binary version of the problem.

☐ Lower bound from using optimal policy from binary version of model.

Example: Sequential Exploration – Results

☐ The bounds get tighter as we delay the provision of perfect information
  and treat more periods “exactly.”

☐ Some penalty is required to get reasonably tight bounds.

<table>
<thead>
<tr>
<th></th>
<th>Filtration ( \mathcal{G}_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Upper Bounds:</td>
<td></td>
</tr>
<tr>
<td>No Penalty</td>
<td>71.14</td>
</tr>
<tr>
<td>With Myopic Penalty</td>
<td>36.78</td>
</tr>
<tr>
<td>With Binary VF Penalty</td>
<td>30.23</td>
</tr>
<tr>
<td>Lower Bound with Binary Policy</td>
<td>18.32</td>
</tr>
</tbody>
</table>
Summary and Conclusions

- Our dual approach is part of an iterative strategy for studying DPs:
  - Lower bounds from MCS with a given policy
  - Upper bounds from information relaxations and penalties

If bounds are too broad, refine.

- Can control effort and avoid:
  - Exact solution of large-scale DPs
  - Detailed modeling of uncertainty resolution (e.g., when is the state of the world known? is volatility observed?)

- Future research:
  - Infinite horizon and/or continuous time
  - “Automatic” generation of upper and lower bounds
  - Careful analysis of the complexity vs. accuracy tradeoff