A Unified Framework for Sequential Decisions under Uncertainty

Warren B Powell
Chief Analytics Officer, Optimal Dynamics
Professor Emeritus, Princeton University

Institute of Data Sciences
Texas A&M University
January 24, 2022
Laboratory sciences

• Materials science
  » Optimizing payloads: reactive species, biomolecules, fluorescent markers, ...
  » Controllers for robotic scientist for materials science experiments
  » Optimizing nanoparticles to maximize photoconductivity
Managing ride-hailing fleets

- Uber
  - Provides real-time, on-demand transportation.
  - Drivers are encouraged to enter or leave the system using pricing signals and informational guidance.

- Decisions:
  - How to price to get the right balance of drivers relative to customers.
  - Assigning and routing drivers to manage Uber-created congestion.
  - Real-time management of drivers.
  - Pricing (trips, new services, ...)
  - Policies (rules for managing drivers, customers, ...)
Managing ride-hailing fleets

• Now we have a logistic curve for each origin-destination pair \((i,j)\)

\[ P^r(p, a | \theta) = \frac{e^{\theta_0^r + \theta_j^r p + \theta_a^r a}}{1 + e^{\theta_0^r + \theta_j^r p + \theta_a^r a}} \]

• Number of offers for each \((i,j)\) pair is relatively small.

• Need to generalize the learning across hundreds to thousands of markets.
Fleet management

- Fleet management problem
  - Optimize the assignment of drivers to loads over time.
  - Tremendous uncertainty in loads being called in
An energy generation portfolio
Energy from wind

Wind power from all PJM wind farms

Jan     Feb     March    April    May    June    July    Aug    Sept    Oct    Nov     Dec

1 year

Megawatts
Mitigation

» Information from drones can be used to guide surface vehicles performing cleanup.

Sensing

» Drones fly over the ocean to detect the presence of oil.

Communication

» Drones communicate by sharing information
Multiagent supply chain management

Pratt & Whitney jet engines
- Over 1,000 parts
- Median lead time for a part is 120 days. Some lead times are over 300 days.
- Parts often require reworking.

Managing the supply chain
- Challenge is determining when to order parts given the long lead times, and production uncertainties.
- Suppliers work for multiple customers.
We are looking for opportunities for making better decisions where we have to deal with uncertainty.
GOALS & OBJECTIVES

» Reduce costs
» Increase profits
» Improve reliability
» Minimize waste
» Maximize strength

» Improve health
» Reduce risk
» Increase yield
» Reduce carbon production
» Minimize lives lost
GOALS & OBJECTIVES

Improve performance by making better decisions.

1ST STEP
What decisions are you making?

2ND STEP
How do we make effective decisions?

It seems we need to turn to "artificial intelligence"
Intelligent decisions

Artificial Intelligence

1. Making computers **behave like** humans
2. Making computers **smarter than** humans

1960s-1970s
Rule-based AI
- If eating meat, then drink red wine
- Given patient attributes, use this treatment.

1990s
Optimization
- Large scale linear & integer programming

2005
Machine Learning
- The new “AI”
- Neural networks

2020+
???

2015
Reinforcement Learning
- Making decisions
- Chess, Go, robots

1. Making computers **behave like** humans
2. Making computers **smarter than** humans

Intelligent decisions
WHAT’S NEXT IN AI?

*Sequential decision problems*, where we need to make decisions over time, as new information arrives.

I propose to unify 15 distinct fields that deal with dynamic decision making into a new field that I call *Sequential Decision Analytics*.

Sequential decision analytics includes all of reinforcement learning, but is broader, with a greater emphasis on uncertainty.
OUTLINE

→ The five layers of intelligence
→ Modeling sequential decision problems
→ Designing policies
  → Policy function approximations
  → Cost function approximations
  → Value function approximations
  → Direct lookahead approximations
→ A new educational field: sequential decision analytics
The five layers of intelligence
Modeling sequential decision problems
Designing policies
  - Policy function approximations
  - Cost function approximations
  - Value function approximations
  - Direct lookahead approximations
A new educational field: sequential decision analytics
THE 5 LAYERS OF INTELLIGENCE

1. Information acquisition and storage
2. Communication
3. Transactions and execution
4. Learning
5. Decisions

Data layers

Statistics/machine learning

Decision analytics

“reinforcement learning”

Data science

Optimal Dynamics

© WARREN POWELL 2021
THE 5 LAYERS OF INTELLIGENCE

- Information acquisition and storage
- Communication
- Transactions and execution
- Learning
- Decisions

- Statistics/machine learning
- Decision analytics

- Data science
- “reinforcement learning”

Optimal Dynamics
THE 5 LAYERS OF INTELLIGENCE

Data layers
- Information acquisition and storage
- Communication
- Transactions and execution
- Learning
- Decisions

Decision analytics
- "reinforcement learning"

Statistics/machine learning
- Data science
MACHINE LEARNING

Types of Learning

Pattern Matching
- What is the voice saying?
- What is in the picture?
- What is the email asking for?

Classification
- What product should I recommend for this customer?
- What treatment should I recommend for this patient?

Inference
- How will an increase in price affect market demand?
- What is the condition of a piece of equipment?

Prediction
- What will the market demand be in three days?
- How many loads will the shipper need to move in a week?
Every single machine learning method falls in one of these three circles.
Machine learning as an optimization problem

\[
\min_{f \in F, \theta \in \Theta} \frac{1}{N} \sum_{n=1}^{N} (y^n - f(x^n | \theta))^2
\]

The first step is choosing a mathematical model that will do the best job of fitting the data (but be careful of overfitting with neural networks).
Machine learning as an optimization problem

\[
\min_{f \in \mathcal{F}, \theta \in \Theta^f} \frac{1}{N} \sum_{n=1}^{N} (y^n - f(x^n, \theta))^2
\]

"Big dataset"

Searching over statistical models

These consist of functions \( f \in \mathcal{F} \) and tunable parameters \( \theta \in \Theta^f \)
Neural networks struggle with:

- **Noise** – Their high flexibility tends to fit the noise.
- **Structure** – It is difficult to communicate structure:
  - Monotonicity – Higher price means lower demand
  - Concavity – As with the newsvendor problem to the right.
MACHINE LEARNING

➢ Neural network for demand response:

- 100 data points, 300000 training iterations

![Graph showing demand response and price]
THE 5 LAYERS OF INTELLIGENCE

- Information acquisition and storage
- Communication
- Transactions and execution
- Learning
- Decisions

Data science
- “reinforcement learning”
- Decision analytics
- Statistics/machine learning
- Data layers
INFORMATION & DECISION PROCESSES
Information and decision processes

- There are parallels between the process of making “decisions” and a manufacturing line making “products”
# DECISIONS

Types of decisions.

<table>
<thead>
<tr>
<th>Physical Decisions</th>
<th>Financial Decisions</th>
<th>Informational Decisions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managing inventories</td>
<td>Pricing decisions</td>
<td>Sending/receiving information</td>
</tr>
<tr>
<td>Assigning drivers and moving trucks</td>
<td>Insurance decisions</td>
<td>Marketing and advertising</td>
</tr>
<tr>
<td>Scheduling nurses and energy generators</td>
<td>Managing investments</td>
<td>Running experiments (lab or field)</td>
</tr>
<tr>
<td></td>
<td>Hedging contracts</td>
<td>Testing drugs</td>
</tr>
</tbody>
</table>
THE TIME FRAMES FOR DECISIONS

Strategic planning and design – We simulate operational decisions so we understand how a system would respond to decisions far in the future:
  » How many gas turbines should a power grid have?
  » How should we design a building to withstand earthquakes?
  » What should the capacity of a levee or reservoir be?

Tactical planning decisions – We simulate operational decisions to help make decisions that impact the system in the near future,
  » How much energy generation should the grid plan for tomorrow?
  » How many gallons of water should be ordered in anticipation of a hurricane?
  » How to allocate traffic management personnel to handle storm evacuations?

Real-time decisions – These are decisions that impact the system now:
  » Making real-time ramping decisions for energy generators.
  » Notifying houses within a zone to begin evacuations before a storm.
  » Operating pumps to mitigate flooding during a storm.
THE TIME FRAMES FOR DECISIONS

Strategic planning and design – We simulate operational decisions so we understand how a system would respond to decisions far in the future:

We need to simulate decisions in the future that do not depend on the state of the system now.

Tactical planning decisions – We simulate operational decisions to help make decisions that impact the system in the near future,

We need to simulate decisions in the future that do depend on the state of the system now.

Real-time decisions – These are decisions that impact the system now:

We need to simulate the effect of a decision now (which depends on the state of the system) on the future.
Airlines around the world use tools that depend on this mathematical model to perform strategic and operational planning.
DETERMINISTIC OPTIMIZATION

Planning a path to your destination

Low dimensional decisions

![Graph with nodes and edges]

\[ x_{ij} = \begin{cases} 1 & \text{If we move from node } i \text{ to node } j \\ 0 & \text{Otherwise} \end{cases} \]

Optimizing facility locations

High dimensional decisions

![Graph with decision variables]

\[ x_i = \begin{cases} 1 & \text{If we locate a facility at location } i \\ 0 & \text{Otherwise} \end{cases} \]
In most settings, decisions are made over time...

Information that arrives after a decision is not known when we made the decision.
Inventory management

Inventory Ordering Decisions

Customer Demands (information)
SEQUENTIAL DECISIONS

Driver dispatch for truckload trucking

Decisions Assigning Drivers to Loads

$x_0$  $x_1$  $x_2$  $x_3$

$\hat{D}_1$  $\hat{D}_2$  $\hat{D}_3$  $\hat{D}_4$

Shippers Calling in Loads (information)
SEQUENTIAL DECISIONS

Testing new vaccines

Vaccination Decisions (what dosage, which people)

Patient Outcomes (information)
SEQEUNICAL DECISIONS

Financial Trading

Buy-sell Decisions (what assets, how much)

$\mathbf{x}_0$, $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$

$\hat{\mathbf{p}}_1$, $\hat{\mathbf{p}}_2$, $\hat{\mathbf{p}}_3$, $\hat{\mathbf{p}}_4$

Changes in Stock Prices

Optimal Dynamics
What is the value of a financial option?

Which driver should move a load?

What is the best policy for high-frequency trading?

Which load to accept now to move next week?

What price to accept for a spot load?

Where should drivers be domiciled?

How many dedicated drivers should we have?

Which physician should handle a procedure?

How many syringes should be sent to each vaccination site, and when?

Which customer tanks should we fill when we are in the area?

How many nurses should we have to serve local hospitals and doctor’s offices?

Which nurse should visit this doctor’s office today?

Where should a patient be assigned for specific treatment?

How many nurses should we have to serve local hospitals and doctor’s offices?

How much battery storage is needed to handle the variability of wind?

When should gas turbines be scheduled to handle drops in wind?

Which customer tanks should we fill when we are in the area?

What is the best policy for high-frequency trading?

When should inventory be ordered?

Which vendor should supply each part?

What price should be charged?

Which supplier should manufacture turbine blades?

How many aircraft should I order for delivery in five years?

How many jet engines should be made each day?

When should inventory be refilled at a fulfillment center?

Which fulfillment center should handle an order?

What bid should we place on Google for a set of ad-words?

How many aircraft should I order for delivery in five years?
Even small sequential decision problems explode dramatically as we plan into the future.
OUTLINE

→ The five layers of intelligence
→ Modeling sequential decision problems
→ Designing policies
  → Policy function approximations
  → Cost function approximations
  → Value function approximations
  → Direct lookahead approximations
→ A new educational field: sequential decision analytics
The biggest challenge when making decisions under uncertainty is **modeling**.

Everyone writing out a deterministic optimization model, or machine learning model, knows how to write out their problem mathematically...

\[
\text{Min } E \left\{ \sum cx \right\} \\
Ax = b \\
x \geq 0
\]

...we lack a standard modeling framework for sequential decisions.
Any sequential decision problems can be written:

$$(S_0, x_0, W_1, S_1, x_1, W_2, \ldots, S_t, x_t, W_{t+1}, S_{t+1}, \ldots, S_T)$$

Each time we make a decision, we receive a contribution $C(S_t, x_t)$.

Decisions are made with a method or policy $X^\pi(S_t)$.

The goal is to find the policy that maximizes expected contributions:

$$\max_\pi \mathbb{E}\{\sum_{t=0}^T C(S_t, X^\pi(S_t)) | S_0\}$$
MODELING SEQUENTIAL DECISION PROBLEMS

Every sequential decision problem can be modeled using 5 core components

» State variables $S_t = (R_t, I_t, B_t)$
  • Physical state $R_t$, other information $I_t$, belief state $B_t$.

» Decision variables $(x_t, a_t, u_t)$
  • Made with policy $X^\pi(S_t | \theta)$ (or $A^\pi(S_t)$ or $U^\pi(S_t)$)

» Exogenous information $W_{t+1}$
  • What do we learn for the first time between $t$ and $t + 1$?

» Transition function $S_{t+1} = S^M(S_t, x_t, W_{t+1})$
  • How do the state variables evolve over time?

» Objective function
  • $\max_{\pi} \mathbb{E}_{S_0} \mathbb{E}_{W_1, \ldots, W_T | S_0} \sum_{t=0}^{T} C(S_t, X^\pi(S_t))$

These five elements describe any sequential decision problem.
We start by identifying:

» What are the performance metrics you are focusing on?

» What decisions are involved?

» What are the sources of uncertainty and new information?

» What information is needed to compute metrics, make decisions, and model their evolution over time (“state variables”)
OUTLINE

→ The five layers of intelligence
→ Modeling sequential decision problems
→ Designing policies
  → Policy function approximations
  → Cost function approximations
  → Value function approximations
  → Direct lookahead approximations
→ A new educational field: sequential decision analytics
Policy Definition
A policy is a method that maps a state variable into a decision ... any method.
# DESIGNING POLICIES

## Policies and the English language

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Format</th>
<th>Prejudice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavior</td>
<td>Formula</td>
<td>Principle</td>
</tr>
<tr>
<td>Belief</td>
<td>Grammar</td>
<td>Procedure</td>
</tr>
<tr>
<td>Bias</td>
<td>Habit</td>
<td>Process</td>
</tr>
<tr>
<td>Canon</td>
<td>Laws/bylaws</td>
<td>Protocols</td>
</tr>
<tr>
<td>Code</td>
<td>Manner</td>
<td>Recipe</td>
</tr>
<tr>
<td>Commandment</td>
<td>Method</td>
<td>Ritual</td>
</tr>
<tr>
<td>Conduct</td>
<td>Mode</td>
<td>Rule</td>
</tr>
<tr>
<td>Control law</td>
<td>Mores</td>
<td>Style</td>
</tr>
<tr>
<td>Convention</td>
<td>Orthodox</td>
<td>Syntax</td>
</tr>
<tr>
<td>Culture</td>
<td>Patterns</td>
<td>Technique</td>
</tr>
<tr>
<td>Customs</td>
<td>Plans</td>
<td>Template</td>
</tr>
<tr>
<td>Duty</td>
<td>Policies</td>
<td>Tenet</td>
</tr>
<tr>
<td>Etiquette</td>
<td>Practice</td>
<td>Tradition</td>
</tr>
<tr>
<td>Fashion</td>
<td>Precedent</td>
<td>Way of life</td>
</tr>
</tbody>
</table>

[http://tinyurl.com/policiesanddecisions](http://tinyurl.com/policiesanddecisions)
Designing policies

Every sequential decision problem can be modeled using 5 core components

> State variables $S_t = (R_t, I_t, B_t)$
  > Physical state $R_t$, other information $I_t$, belief state $B_t$.

> Decision variables $(x_t, a_t, u_t)$
  > Made with policy $X^\pi(S_t|\theta)$ (or $A^\pi(S_t)$ or $U^\pi(S_t)$)

> Exogenous information $W_{t+1}$
  > What do we learn for the first time between $t$ and $t+1$?

> Transition function $S_{t+1} = S_t^M(S_t, x_t, W_{t+1})$
  > How do the state variables evolve over time?

> Objective function
  > $\max_{\pi} \mathbb{E}_{S_0}\mathbb{E}_{W_1,...,W_T|S_0}\sum_{t=0}^{T} C(S_t, X^\pi(S_t))$

These five elements describe any sequential decision problem.
Evaluating policies

1) Theoretically
   - Optimality proofs
   - Regret bounds
   - Asymptotic convergence

2) Through numerical simulations

3) In the field
BRIDGING MACHINE LEARNING & SEQUENTIAL DECISIONS

Machine learning

\[
\min_{f \in F, \theta \in \Theta} \frac{1}{N} \sum_{n=1}^{N} \left( y^n - f(x^n, \theta) \right)^2
\]

Searching over functions

“Big dataset”

Sequential decisions

\[
\max_{\pi = (f \in F, \theta \in \Theta)} \frac{1}{N} \sum_{n=1}^{N} \sum_{t=0}^{T} C(S^n_t, X^\pi(S^n_t | \theta))
\]

Searching over policies

System model

System model:

\[
S_{t+1} = S^M(S_t, x_t, W_{t+1})
\]
DESIGNING POLICIES

Two fundamental strategies for designing policies

Policy search – Search over a class of methods for making decisions to optimize some metric over time.
  » Finding the best class of policy.
  » Finding the best policy within the class.

Lookahead approximations – Approximate the impact of a decision now on the future.
  » The contribution from the first period, plus
  » An approximation of the sum of contributions in future time periods resulting from the first decision.
1) Policy function approximation (PFA)

These are analytical functions that specify what to do given what we know.

Examples:

a) Order-up-to inventory policy $\theta = (\theta_{min}, \theta_{max})$

b) Buy when the price goes below $\theta_{min}$ and sell when it goes above $\theta_{max}$

c) Lookup tables, linear/nonlinear models, neural networks, nonparametric models, …
Policy search

2) Cost function approximations (CFAs)

These are parameterized optimization problems:

a) Find the shortest path to a destination, but add a buffer $\theta$ (e.g. 15 minutes) to make sure you arrive on time.

b) Optimize energy generation for tomorrow to meet forecasted demand, but add reserves $\theta$ in case of a generator failure.

c) Advertise the product $x$ which solves:

$$X^{UCB}(S^n|\theta) = \max_x (\text{Estimated revenue}_x^n + \theta \cdot \text{Standard deviation of estimated revenue}_x^n)$$

Now solve:

$$\max_\theta \mathbb{E}\{\sum_{n=1}^N C(S^n, X^\pi(S^n|\theta)) | S_0\}$$

Parametric CFAs are widely used in industry, yet dismissed by the academic research community. This is actually quite a powerful strategy.
Cost function approximations

- Lookup table
  - We can organize potential catalysts into groups
  - Scientists using domain knowledge can estimate correlations in experiments between similar catalysts.
Cost function approximations

- Correlated beliefs: Testing one material teaches us about other materials
Cost function approximations

- Cost function approximations (CFA)
  - Upper confidence bounding
    \[ X^{UCB} (S^n | \theta^{UCB}) = \arg\max_x (\bar{\mu}_x^n + \theta^{UCB} \sqrt{\log n / N_x^n}) \]
  - Interval estimation
    \[ X^{IE} (S^n | \theta^{IE}) = \arg\max_x (\bar{\mu}_x^n + \theta^{IE} \bar{\sigma}_x^n) \]
  - Thompson sampling
    \[ x^n = \arg\max_x \hat{\mu}_x^n \quad \hat{\mu}_x^n \sim N(\bar{\mu}_x^n, \theta^{TS} \bar{\sigma}_x^{2,n}) \]
Cost function approximations

- Picking $\theta^{IE} = 0$ means we are evaluating each choice at the mean.
Cost function approximations

- Picking $\theta^{IE} = 2$ means we are evaluating each choice at the 95th percentile.
Cost function approximations

- Optimizing the policy
  - We optimize $\theta^{IE}$ to maximize:
    \[
    \max_{\theta^{IE}} F(\theta^{IE}) = \mathbb{E} F\left(x^{\pi,N}, W\right)
    \]
    where
    \[
    x^n = X^{IE}(S^n | \theta^{IE}) = \arg\max_x \left( \bar{\mu}_x^n + \theta^{IE} \bar{\sigma}_x^n \right)
    \]
    \[
    S^n = (\bar{\mu}_x^n, \bar{\sigma}_x^n)
    \]

- Notes:
  - This can handle any belief model, including correlated beliefs, nonlinear belief models.
  - All we require is that we be able to simulate a policy.
Hybrid direct lookahead/CFA

• An energy storage problem:
Hybrid direct lookahead/CFA

- Forecasts evolve over time as new information arrives:

  Rolling forecasts, updated each hour.

  Forecast made at midnight:

  Actual
Hybrid direct lookahead/CFA

- Benchmark policy – Deterministic lookahead

\[ X_t^{D-LA}(S_t) = \arg \min_{x_t, (\bar{x}_t, t'=t+1, \ldots, t+H)} \left( C(S_t, x_t) + \sum_{t'=t+1}^{t+H} \bar{c}_{t', \bar{x}_{t'}} \right) \]

\[
\begin{align*}
\tilde{x}_{tt'}^{wd} + \beta \tilde{x}_{tt'}^{rd} + \tilde{x}_{tt'}^{gd} &\leq f_{tt'}^D \\
\tilde{x}_{tt'}^{gd} + \tilde{x}_{tt'}^{gr} &\leq f_{tt'}^G \\
\tilde{x}_{tt'}^{rd} + \tilde{x}_{tt'}^{rg} &\leq \tilde{R}_{tt'} \\
\tilde{x}_{tt'}^{wr} + \tilde{x}_{tt'}^{gr} &\leq R_{t'}^{\text{max}, ttt} \\
\tilde{x}_{tt'}^{wr} + \tilde{x}_{tt'}^{wd} &\leq f_{tt'}^E \\
\tilde{x}_{tt'}^{wr} + \tilde{x}_{tt'}^{gr} &\leq \gamma_{\text{ch} \arg e} \\
\tilde{x}_{tt'}^{rd} + \tilde{x}_{tt'}^{rg} &\leq \gamma_{\text{disch} \arg e}
\end{align*}
\]
Hybrid direct lookahead/CFA

- Benchmark policy – Deterministic lookahead

\[ X_t^{D-LA} (S_t | \theta) = \arg\min_{x_t, (\tilde{x}_{tt'}, t'=t+1,...,t+H)} \left( C(S_t, x_t) + \sum_{t'=t+1}^{t+H} \tilde{c}_{tt'} \tilde{x}_{tt'} \right) \]

\[ \tilde{x}_{tt'}^{wd} + \beta \tilde{x}_{tt'}^{rd} + \tilde{x}_{tt'}^{gd} \leq f_{tt'}^D \]
\[ \tilde{x}_{tt'}^{gd} + \tilde{x}_{tt'}^{gr} \leq f_{tt'}^G \]
\[ \tilde{x}_{tt'}^{rd} + \tilde{x}_{tt'}^{rg} \leq \tilde{R}_{tt'} \]
\[ \tilde{x}_{tt'}^{wr} + \tilde{x}_{tt'}^{gr} \leq R_{\text{max}} \tilde{x}_{tt'} \]
\[ \tilde{x}_{tt'}^{wr} + \tilde{x}_{tt'}^{wd} \leq \theta_{tt'} - tf_{tt'} \]
\[ \tilde{x}_{tt'}^{wr} + \tilde{x}_{tt'}^{gr} \leq \gamma^{ch} \text{arg e} \]
\[ \tilde{x}_{tt'}^{rd} + \tilde{x}_{tt'}^{rg} \leq \gamma^{\text{disch arg e}} \]
Hybrid direct lookahead/CFA

- Benchmark policy – Deterministic lookahead

\[
X_t^{D-LA}(S_t) = \arg \min_{x_t, (\tilde{x}_{t'\cdot t+1=1}, \ldots, t+H)} \left( C(S_t, x_t) + \sum_{t' = t+1}^{t+H} \tilde{c}_{tt'} \tilde{x}_{tt'} \right)
\]

\[
\begin{align*}
\tilde{x}_{tt'}^{wd} + \beta \tilde{x}_{tt'}^{rd} + \tilde{x}_{tt'}^{gd} & \leq f_{tt'}^D \\
\tilde{x}_{tt'}^{gd} + \tilde{x}_{tt'}^{gr} & \leq f_{tt'}^G \\
\tilde{x}_{tt'}^{rd} + \tilde{x}_{tt'}^{rg} & \leq \tilde{R}_{tt'} \\
\tilde{x}_{tt'}^{wr} + \tilde{x}_{tt'}^{gr} & \leq R_{\text{max}}^{\max} \tilde{x}_{tt'} \\
\tilde{x}_{tt'}^{wr} + \tilde{x}_{tt'}^{wd} & \leq \theta_{tt'-t} f_{tt'}^E \\
\tilde{x}_{tt'}^{wr} + \tilde{x}_{tt'}^{gr} & \leq \gamma_{\text{arg e}} \\
\tilde{x}_{tt'}^{rd} + \tilde{x}_{tt'}^{rg} & \leq \gamma_{\text{disch arg e}}
\end{align*}
\]
Hybrid direct lookahead/CFA

- Lookahead policies peek into the future
  - Optimize over deterministic lookahead model
Hybrid direct lookahead/CFA

- Lookahead policies peek into the future
  - Optimize over deterministic lookahead model

\[ x^L_{t+2}(S_t) = \arg\min C(S_{t+2}, X_{t+2}) + \sum_{i=0}^{p-1} C(S_{t+2}, X_{t+2+i}) \]
Hybrid direct lookahead/CFA

- Lookahead policies peek into the future
  » Optimize over deterministic lookahead model
Hybrid direct lookahead/CFA

- Lookahead policies peek into the future
  - Optimize over deterministic lookahead model
Hybrid direct lookahead/CFA

- One-dimensional contour plots – perfect forecast

\[ \theta_i^* = 1 \] for perfect forecasts.
Hybrid direct lookahead/CFA

- One-dimensional contour plots-uncertain forecast
Energy storage optimization

• Tuning the parameters

\[ \theta_i = 1 \quad \theta_i \in [0,1] \quad \theta_i \in [0.5,1.5] \quad \theta_i \in [1,2] \]

Starting points for stochastic search for \( \theta \)

Percent improvement over \( \theta = 1 \)

Performance of the lookup policy obtained by the SGF-CFA method with batch size of 12, \( \text{eta}=1 \) and \( \theta=1.5 \)

\( \theta = 1 \)
Cost function approximations

• Other applications
  » Airlines optimizing schedules with schedule slack to handle weather uncertainty.
  » Manufacturers using buffer stocks to hedge against production delays and quality problems.
  » Grid operators scheduling extra generation capacity in case of outages.
  » Adding time to a trip planned by Google maps to account for uncertain congestion.

See [http://tinyurl.com/cfapolicy](http://tinyurl.com/cfapolicy) for more on parametric cost function approximations.
Policy search

• Tuning the policy (PFAs or CFAs):
  » We need to maximize
    \[
    \max_{\theta} F(\theta) = \mathbb{E}\left\{\sum_{t=0}^{T} C(S_t, X_t^\pi(S_t|\theta)) | S_0\right\}
    \]
  » We cannot compute the expectation, so we run simulations:
Policy function approximations

- How do we search for the best $\theta$?
  - Derivative-based
    - Stochastic gradient methods:
      \[
      \theta^{n+1} = \theta^n + \alpha_n \nabla_{\theta} F(\theta^n, W^{n+1})
      \]
  - Derivative-free
    - Build a belief model $\tilde{F}(\theta) \approx \mathbb{E} F(\theta, W)$ that approximates our function.

- Both of these approaches are sequential decision problems!
DESIGNING POLICIES

Two fundamental strategies for designing policies

**Policy search** – Search over a class of methods for making decisions to optimize some metric over time.

» Finding the best class of policy.

» Finding the best policy within the class.

**Lookahead approximations** – Approximate the impact of a decision now on the future.

» The contribution from the first period, plus

» An approximation of the sum of contributions in future time periods resulting from the first decision.
Lookahead approximations

- Lookahead approximations combine:
  - The immediate contribution (or cost) of a decision made now...
  - ... and an approximation of future contributions (or costs)
This looks like scary mathematics, but it is what all of us are doing when we make decisions now that consider what might happen in the future.

The challenge is … how to compute it!!!

Lookahead policies are based on solving

\[
X_t^*(S_t) = \arg\max_x \left( C(S_t, x_t) + \mathbb{E} \left\{ \max_\pi \left( \mathbb{E} \sum_{t' = t+1}^{T} C(S_t, X_{t'}^\pi(S_{t'})) \middle| S_{t+1} \right) \right\} \right)
\]
Lookahead policies are based on solving

\[ X^*_t(S_t) = \arg\max_x \left( C(S_t, x_t) + \mathbb{E}\left[ \max_{\pi} \mathbb{E}\left[ \sum_{t=1}^{T} C(S_t, X^*_t(S_t'), |S_{t+1}|, |S_t, x_t|) \right] \right] \right) \]
**Lookahead approximations**

Approximate the impact of a decision now on the future

\[
X^*_t(S_t) = \arg \max_x \left( C(S_t, x_t) + \mathbb{E} \left\{ \max_\pi \mathbb{E} \sum_{t'=t+1}^T C(S_t, X^\pi_{t'}(S_{t'})) \mid S_{t+1} \right\} \right)
\]

**3) Value function approximations (VFAs)**

\[
X^*_t(S_t) = \arg \max_x \left( C(S_t, x_t) + \mathbb{E} \left\{ V_{t+1}(S_{t+1}) \mid S_t, x_t \right\} \right)
\]

\[
X^{VFA}_t(S_t) = \arg \max_x \left( C(S_t, x_t) + \mathbb{E} \left\{ \overline{V}^x_{t+1}(S_{t+1}) \mid S_t, x_t \right\} \right)
\]

\[
= \arg \max_x \left( C(S_t, x_t) + \overline{V}^x_t(S_t) \right)
\]

\[
= \arg \max_x \overline{Q}_t(S_t, x_t) \quad ("Q-learning")
\]
DESIGNING POLICIES

Lookahead approximations
Approximate the impact of a decision now on the future

\[ X_t^*(S_t) = \arg \max_x \left( C(S_t, x_t) + \mathbb{E} \left\{ \max_{\pi} \left( \mathbb{E} \sum_{t'=t+1}^T C(S_{t'}, X_{t'}^{\pi}(S_{t'})) | S_{t+1} \right) | S_t, x_t \right\} \right) \]

3) Value function approximations (VFAs)

\[ X_t^*(S_t) = \arg \max_x \left( C(S_t, x_t) + \mathbb{E} \left\{ V_{t+1}(S_{t+1}) | S_t, x_t \right\} \right) \]
\[ X_t^{VFA}(S_t) = \arg \max_x \left( C(S_t, x_t) + \mathbb{E} \left\{ \bar{V}_{t+1}(S_{t+1}) | S_t, x_t \right\} \right) \]
\[ = \arg \max_x \left( C(S_t, x_t) + \bar{V}_x(S_t) \right) \]
\[ = \arg \max_x \bar{Q}_t(S_t, x_t) \quad ("Q-learning") \]
Lookahead approximations
Approximate the impact of a decision now on the future

$$X_t^*(S_t) = \arg \max_x \left( C(S_t, x_t) + \mathbb{E} \left\{ \max_\pi \left( \mathbb{E} \sum_{t'=t+1}^{T} C(S_t, X_{t'}^\pi(S_{t'})) | S_{t+1} \right) | S_t, x_t \right\} \right)$$

3) Value function function approximations (VFAs)

$$X_t^*(S_t) = \arg \max_x \left( C(S_t, x_t) + \mathbb{E} \left\{ V_{t+1}(S_{t+1}) | S_t, x_t \right\} \right)$$

$$X_t^{VFA}(S_t) = \arg \max_x \left( C(S_t, x_t) + \mathbb{E} \left\{ \overline{V}_{t+1}(S_{t+1}) | S_t, x_t \right\} \right)$$

$$= \arg \max_x \left( C(S_t, x_t) + \overline{V}_t^x(S_t^x) \right)$$

$$= \arg \max_x \overline{Q}_t(S_t, x_t) \quad \text{("Q-learning")}$$
Lookahead approximations
Approximate the impact of a decision now on the future

\[ X_t^*(S_t) = \arg\max_x \left( C(S_t, x_t) + \mathbb{E} \left\{ \max_{\pi} \left\{ \mathbb{E} \sum_{t'=t+1}^T C(S_t, X_{t'}^\pi(S_{t'})) | S_{t+1} \right\} | S_t, x_t \right\} \right) \]

3) Value function approximations (VFAs)

\[ X_t^* (S_t) = \arg\max_{x_t} \left( C(S_t, x_t) + \mathbb{E} \left\{ V_{t+1}(S_{t+1}) | S_t, x_t \right\} \right) \]

\[ X_t^{VFA} (S_t) = \arg\max_{x_t} \left( C(S_t, x_t) + \mathbb{E} \left\{ \bar{V}_{t+1}(S_{t+1}) | S_t, x_t \right\} \right) \]

\[ = \arg\max_{x_t} \left( C(S_t, x_t) + \bar{V}_t^x(S_t^x) \right) \]

\[ = \arg\max_{x_t} \tilde{Q}_t (S_t, x_t) \quad ("Q-learning") \]

© WARREN POWELL 2021
Locational marginal prices on the grid

LMPs – Locational marginal prices

$977/MW !!!

NEW JERSEY HUB: 977.87
EASTERN HUB: 208.20
WESTERN HUB: 1,000.40
DOMINION HUB: 362.57
OHIO HUB: 77.27

ARKANSAS HUB: 247.21
LOUISIANA HUB: 33.51
TEXAS HUB: 24.74

MINN HUB: 83.40
ILLINOIS HUB: 83.35
MICHIGAN HUB: 69.30
INDIANA HUB: 62.44

N I L L I N O I S H U B : 70.45
CHICAGO HUB: 74.23
AEP-DAYTON HUB: 196.31

© WARREN POWELL 2021
Imagine 25 large storage devices spread around the PJM grid:
Imagine 25 large storage devices spread around the PJM grid:
The value of grid level storage
The value of grid level storage
The value of grid level storage
The value of grid level storage

Monday

Time: 05  :10  :15  :20

...
Derivatives are used to estimate a piecewise linear approximation.
Approximate dynamic programming for energy storage

- With luck, your objective function improves
Approximate dynamic programming for energy storage

- Congested grid:
  - Green and blue circles indicate energy storage
Approximate dynamic programming for energy storage

- Congested grid:
  - Green and blue circles indicate energy storage
NOTABLE APPLICATIONS

Reinforcement Learning

» Major achievement — Playing Go

High-Dimensional ADP

» Major achievement — Optimizing a 5000 truck fleet
4) Direct lookahead policies (DLAs) – Here we create an approximation called the approximate lookahead model:

\[(\tilde{S}_{tt}, \tilde{x}_{tt}, \tilde{W}_{t,t+1}, \tilde{S}_{t,t+1}, \tilde{x}_{t,t+1}, \tilde{W}_{t,t+2}, ..., \tilde{S}_{tt'}, \tilde{x}_{tt'}, \tilde{W}_{t,t'+1}, ... )\]

There are six classes of approximations we can introduce. Our direct lookahead policy now requires solving:

\[X_t^{DLA}(S_t|\theta) = \arg \max_x \left( C(S_t, x_t) + \mathbb{E} \left\{ \max_{\pi} \sum_{t'=t+1}^{t+H} \mathbb{E} \left\{ C(\tilde{S}_{t'}, \tilde{x}_{t',\pi}(\tilde{S}_{t'})) \right\} \right| S_t, x_t \right) \]

- Sampled information process
- Restricted horizon
- Limited decisions
- “Policy-within-a-policy”
- Reduced state space
DESIGNING POLICIES

Direct Lookahead Policies (DLAs)

» Tilde variables are used to model approximate lookahead

The lookahead model

The base model

(t, \tilde{x}_t, \tilde{W}_t, \ldots, S_{t+1}, \tilde{x}_{t+1}, \tilde{W}_{t+1}, \ldots, S_{t+3}, \tilde{x}_{t+3}, \tilde{W}_{t+3}, \ldots)

Optimal Dynamics
DESIGNING POLICIES

Direct Lookahead Policies (DLAs)

» Tilde variables are used to model approximate lookahead
Direct Lookahead Policies (DLAs)

» Tilde variables are used to model approximate lookahead
Examples of Lookahead Models

- **The deterministic lookahead model**
  - This is what is most widely used in practice.
  - Standard approach is to use a “best estimate” (which means deterministic) of travel times in the future.
  - This is often referred to as “model predictive control”

- **Robust optimization** - We could use the 90th percentile of travel times.

- **Stochastic programming** – We represent the future using, say, 20 samples.

- **Approximate dynamic programming applied to approximate lookahead model**

- **Chance constrained programming** – Impose constraint on the probability of being late.

See Chapter 19 at [http://tinyurl.com/RLandSO](http://tinyurl.com/RLandSO)
The unit commitment problem (for PJM)
Planning tomorrow’s schedule
The unit commitment problem (for PJM)
Planning tomorrow’s schedule
The timing of decisions

- The day-ahead unit commitment problem
The timing of decisions

- Intermediate-term unit commitment problem
The timing of decisions

- Intermediate-term unit commitment problem
The unit commitment problem

- Day-ahead unit commitment
- Load curtailment notification
- Natural gas generation
- Tapping spinning reserve
The unit commitment problem
The unit commitment problem
DESIGNING POLICIES

Policy search policies

Policy function approximations (PFAs)
- Simple rules, functions
- Examples:
  - Order up to
  - Buy low, sell high

Cost function approximations (CFAs)
- Parameterized cost models
- Examples
  - Schedule slack for trips
  - Buffer stocks for inventory

Lookahead policies

Value function approximations (VFAs)
- Making a decision now using the value of being in a future state
- Examples:
  - The value of a truck driver
  - The value of holding an asset

Direct lookaheads (DLAs)
- Models that optimize over a planning horizon (deterministically/stochastically)
- Examples:
  - Google maps
  - Energy planning models
The four classes of policies are *universal* – they cover every method for making decisions described in the research literature or used in practice.

This means you are already using one of the four classes of policies (or a hybrid) in your own decisions. But are you doing the best you can?
THE UNIVERSAL FRAMEWORK FOR SEQUENTIAL DECISIONS
Warren B. Powell, Princeton University
http://tinyurl.com/powelljungle

\[
\max_{\pi} \mathbb{E} \left\{ \sum_{t=0}^{T} C(S_t, X^\pi(S_t)) | S_0 \right\}
\]
where \( S_{t+1} = S^M(S_t, X^\pi(S_t), W_{t+1}) \)
and given \((S_0, W_1, W_2, ..., W_t, ... )\)

The four classes of policies (PFAs, CFAs, VFAs and DLAs) are universal. Any sequential decision problem will use one of these four classes (or a hybrid), including whatever you might be doing now.

The optimal policy (if we could solve it) is given by

\[
X^*(S_t) = \max_{x_t} \left\{ C(S_t, x_t) + \mathbb{E} \left\{ \max_{\pi} \mathbb{E} \left\{ \sum_{t'=t+1}^{t+H} C(S_{t'}, X^\pi(S_{t'})) | S_{t'}, x_t \right\} | S_t, x_t \right\} \right\}
\]
Policies in the “policy search” class are simplest
  - As a result, this is what you are most likely going to see used in practice.

  ... but ...

  - “The price of simplicity is tunable parameters”

  ... and ...

  - “Tuning is hard”!
Consider a basic energy storage problem

We are going to show that with minor variations in the characteristics of this problem, we can make each class of policy work best.
## AN ENERGY STORAGE PROBLEM

Each policy is best on certain problems

<table>
<thead>
<tr>
<th>Problem:</th>
<th>Problem description</th>
<th>PFA</th>
<th>CFA Error correction</th>
<th>VFA</th>
<th>Deterministic Lookahead</th>
<th>CFA Lookahead</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A stationary problem with heavy-tailed prices, relatively low noise, moderately accurate forecasts.</td>
<td>0.959</td>
<td>0.839</td>
<td>0.936</td>
<td>0.887</td>
<td>0.887</td>
</tr>
<tr>
<td>B</td>
<td>A time-dependent problem with daily load patterns, noseasonalities in energy and price, relatively low noise, less accurate forecasts.</td>
<td>0.714</td>
<td>0.752</td>
<td>0.712</td>
<td>0.746</td>
<td>0.746</td>
</tr>
<tr>
<td>C</td>
<td>A time-dependent problem with daily load, energy and price patterns, relatively high noise, forecast errors increase over horizon.</td>
<td>0.865</td>
<td>0.590</td>
<td>0.914</td>
<td>0.886</td>
<td>0.886</td>
</tr>
<tr>
<td>D</td>
<td>A time-dependent problem, relatively low noise, very accurate forecasts.</td>
<td>0.962</td>
<td>0.749</td>
<td>0.971</td>
<td>0.997</td>
<td>0.997</td>
</tr>
<tr>
<td>E</td>
<td>Same as (C), but the forecast errors are stationary over the planning horizon.</td>
<td>0.865</td>
<td>0.590</td>
<td>0.914</td>
<td>0.922</td>
<td>0.934</td>
</tr>
</tbody>
</table>

Joint research with Prof. Stephan Meisel, University of Muenster, Germany.

» ... any policy might be best depending on the data.
BRIDGING MACHINE LEARNING & SEQUENTIAL DECISIONS

Machine learning

\[
\min_{f \in F, \theta \in \Theta} \frac{1}{N} \sum_{n=1}^{N} (y^n - f(x^n | \theta))^2
\]

Searching over functions

“Big dataset”

Sequential decisions

\[
\max_{\pi = (f \in F, \theta \in \Theta)} \frac{1}{N} \sum_{n=1}^{N} \sum_{t=0}^{T} C(S^n_t, X^n \pi (S^n_t | \theta))
\]

Searching over policies

System model

- Policy function approximations
- Cost function approximations
- Value function approximations
- Direct lookahead approximations

- Analytical functions
- Optimization problem
- Optimization problem
- Optimization problem

© WARREN POWELL 2021
Choosing a policy class

A new book:

» First book to introduce a universal modeling framework, covering all four classes of policies.
» Describes the tools for modeling and solving any sequential decision problem, from simple learning problems to truckload fleets to complex supply chains.
» Aimed at a technical audience interested in writing software to develop models such as those described in this presentation.
» Provides the foundation for a new field we are calling sequential decision analytics.

http://tinyurl.com/RLandSO/
Choosing a policy class

An introductory book:

» Uses a teach-by-example style

» Illustrates how to model sequential decision problems using a rich set of examples

» Illustrates all four classes of policies

» Highlights uncertainty modeling

http://tinyurl.com/sdaexamplesprint
OUTLINE

→ The five layers of intelligence
→ Modeling sequential decision problems
→ Designing policies
  → Policy function approximations
  → Cost function approximations
  → Value function approximations
  → Direct lookahead approximations
→ A new educational field: sequential decision analytics
The core disciplines of decision analytics

Optimization

Each of these fields have well-defined communities, using common notation and established tools.

There are widely used textbooks that cover common material, with standard notational frameworks.

Simulation

The concepts are taught in hundreds of academic programs, producing thousands of graduates each year which can be hired by industry.

Machine learning
The fields that deal with decisions and uncertainty are completely fragmented.

- Sequential decision analytics is not a recognized field.
- There are 15 distinct communities that deal with decisions under uncertainty.
- Each community offers tools that work only for narrowly defined problem classes.
- Real applications require skills that span a wide range of problem settings.
The fields that deal with decisions and uncertainty are completely fragmented.

» Sequential decision analytics is not a recognized field.

» There are 15 distinct communities that deal with decisions under uncertainty

» Each community offers tools that work only for specific problems

» Real applications require skills that span a wide range of problem settings.

These will be the first books to present sequential decision problems and solution methods in a unified way.
An academic proposal

» We need to establish academic programs in engineering focusing broadly on sequential decision analytics, comparable to existing programs in machine learning.

» As with machine learning, this program could be centered on methodology, or different problem domains that span engineering (all fields), the sciences (all fields), business, finance, logistics, energy and health.

FOR MORE INFORMATION, VISIT
http://tinyurl.com/sdafield
Thank you!

See

http://tinyurl.com/sdafield

for an introduction to a field I am calling

“Sequential Decision Analytics”

My new book is available at

http://tinyurl.com/RLandSO/