Recent Trends in Control and Optimization of Stochastic Models

Bernd Heidergott
Vrije Universiteit, Econometrics

November 26, 2010

Outline of the Talk

A Few Thoughts About Control and Optimization

The World of Control

The general model of control is that an input process is fed to a system (box) which produces an output process.

Control/optimization is possible if the either the input process or the system behavior can be influenced by a controller.

The World of Control

The general model of control is that an input process is fed to a system (box) which produces an output process.

Control/optimization is possible if the either the input process or the system behavior can be influenced by a controller.

The World of Control

The general model of control is that an input process is fed to a system (box) which produces an output process.

Control/optimization is possible if the either the input process or the system behavior can be influenced by a controller.

White Box vs. Black Box

The world of comtrol/optimization has two main paradigms.

The Black Box Paradigm

No (or only very little) information is available on how the input leads to the output (think of the economy).

The White Box Paradigm

A detailed model is available on the relation between the input and the output (think of a queueing system).

This talk is about white box control/optimization.

White Box vs. Black Box

The world of comtrol/optimization has two main paradigms.

The Black Box Paradigm

No (or only very little) information is available on how the input leads to the output (think of the economy).

The White Box Paradigm

A detailed model is available on the relation between the input and the output (think of a queueing system).

This talk is about white box control/optimization.

White Box vs. Black Box

The world of comtrol/optimization has two main paradigms.

The Black Box Paradigm

No (or only very little) information is available on how the input leads to the output (think of the economy).

The White Box Paradigm

A detailed model is available on the relation between the input and the output (think of a queueing system).

This talk is about white box control/optimization.

White Box vs. Black Box

The world of comtrol/optimization has two main paradigms.

The Black Box Paradigm

No (or only very little) information is available on how the input leads to the output (think of the economy).

The White Box Paradigm

A detailed model is available on the relation between the input and the output (think of a queueing system).

This talk is about white box control/optimization.

A More Detailed Model

- In the following we assume that the system can be modeled by a Markov chain P (possibly on a general state space).
- We assume that the system dynamics can be influenced through the choice of some design parameter θ and we write P_θ to indicate this.
- The output process is denoted by $\{X_\theta(t)\}$ where $t \in \mathbb{R}_+$ or $t \in \mathbb{N}$.

We assume that P_θ is given to us in a closed-form analytical expression (white box), so we can do math with it.

A More Detailed Model

- In the following we assume that the system can be modeled by a Markov chain P (possibly on a general state space).
- We assume that the system dynamics can be influenced through the choice of some design parameter θ and we write P_θ to indicate this.
- The output process is denoted by $\{X_\theta(t)\}$ where $t \in \mathbb{R}_+$ or $t \in \mathbb{N}$.

We assume that P_θ is given to us in a closed-form analytical expression (white box), so we can do math with it.

A More Detailed Model

- In the following we assume that the system can be modeled by a Markov chain P (possibly on a general state space).
- We assume that the system dynamics can be influenced through the choice of some design parameter θ and we write P_θ to indicate this.
- The output process is denoted by $\{X_\theta(t)\}$ where $t \in \mathbb{R}_+$ or $t \in \mathbb{N}$.

We assume that P_θ is given to us in a closed-form analytical expression (white box), so we can do math with it.

A More Detailed Model

- In the following we assume that the system can be modeled by a Markov chain P (possibly on a general state space).
- We assume that the system dynamics can be influenced through the choice of some design parameter θ and we write P_θ to indicate this.
- The output process is denoted by $\{X_\theta(t)\}$ where $t \in \mathbb{R}_+$ or $t \in \mathbb{N}$.

We assume that P_θ is given to us in a closed-form analytical expression (white box), so we can do math with it.

A More Detailed Model

- In the following we assume that the system can be modeled by a Markov chain P (possibly on a general state space).
- We assume that the system dynamics can be influenced through the choice of some design parameter θ and we write P_θ to indicate this.
- The output process is denoted by $\{X_\theta(t)\}$ where $t \in \mathbb{R}_+$ or $t \in \mathbb{N}$.

We assume that P_θ is given to us in a closed-form analytical expression (white box), so we can do math with it.

A Real Control Problem

Consider a controller that regulates the amount of insulin given to a diabetics patient.

It is of key importance that the control algorithm

- quickly adapts to a changing environment, and
- quickly stabilizes after a regime change.

Non-Stationary Control

Control in a changing environment is very hard!

We don't even know what "optimal" control means in the non-stationary problem.

A Real Control Problem

Consider a controller that regulates the amount of insulin given to a diabetics patient.

It is of key importance that the control algorithm

- quickly adapts to a changing environment, and
- quickly stabilizes after a regime change.

Non-Stationary Control

Control in a changing environment is very hard!

We don't even know what "optimal" control means in the non-stationary problem.

A Real Control Problem

Consider a controller that regulates the amount of insulin given to a diabetics patient.

It is of key importance that the control algorithm

- quickly adapts to a changing environment, and
- quickly stabilizes after a regime change.

Non-Stationary Control

Control in a changing environment is very hard!

We don't even know what "optimal" control means in the non-stationary problem.

A Real Control Problem

Consider a controller that regulates the amount of insulin given to a diabetics patient.

It is of key importance that the control algorithm

- quickly adapts to a changing environment, and
- quickly stabilizes after a regime change.

Non-Stationary Control

Control in a changing environment is very hard!

We don't even know what "optimal" control means in the non-stationary problem.

A Real Control Problem

Consider a controller that regulates the amount of insulin given to a diabetics patient.

It is of key importance that the control algorithm

- quickly adapts to a changing environment, and
- quickly stabilizes after a regime change.

Non-Stationary Control

Control in a changing environment is very hard!

We don't even know what "optimal" control means in the non-stationary problem.

A Real Control Problem

Consider a controller that regulates the amount of insulin given to a diabetics patient.

It is of key importance that the control algorithm

- quickly adapts to a changing environment, and
- quickly stabilizes after a regime change.

Non-Stationary Control

Control in a changing environment is very hard!

We don't even know what "optimal" control means in the non-stationary problem.

A Real Control Problem

Consider a controller that regulates the amount of insulin given to a diabetics patient.

It is of key importance that the control algorithm

- quickly adapts to a changing environment, and
- quickly stabilizes after a regime change.

Non-Stationary Control

Control in a changing environment is very hard!

We don't even know what "optimal" control means in the non-stationary problem.

Stationary Control

Let's turn to Stationary Control

- Assume that P_θ is homogeneous (time independent) and we want to control the steady-state behavior of the system.
- Suppose the system is operating at θ. A natural question to ask is the following: *Would it be better to change θ a little bit?*

Challenge

Compare $\pi_{\theta+\Delta}$ with π_θ in an efficient way.

Nota Bene: Never forget that we are doing this only because we cannot do better, or as Hinderer said: "In the long-run we are all dead."

Let's turn to Stationary Control

- Assume that P_θ is homogeneous (time independent) and we want to control the steady-state behavior of the system.
- Suppose the system is operating at θ. A natural question to ask is the following: *Would it be better to change θ a little bit?*

Challenge

Compare $\pi_{\theta+\Delta}$ with π_θ in an efficient way.

Nota Bene: Never forget that we are doing this only because we cannot do better, or as Hinderer said: "In the long-run we are all dead."

Let's turn to Stationary Control

- Assume that P_θ is homogeneous (time independent) and we want to control the steady-state behavior of the system.
- Suppose the system is operating at θ. A natural question to ask is the following: *Would it be better to change θ a little bit?*

Challenge

Compare $\pi_{\theta+\Delta}$ with π_θ in an efficient way.

Nota Bene: Never forget that we are doing this only because we cannot do better, or as Hinderer said: "In the long-run we are all dead."

Let's turn to Stationary Control

- Assume that P_θ is homogeneous (time independent) and we want to control the steady-state behavior of the system.
- Suppose the system is operating at θ. A natural question to ask is the following: *Would it be better to change θ a little bit?*

Challenge

Compare $\pi_{\theta+\Delta}$ with π_θ in an efficient way.

Nota Bene: Never forget that we are doing this only because we cannot do better, or as Hinderer said: "In the long-run we are all dead."

Let's turn to Stationary Control

- Assume that P_θ is homogeneous (time independent) and we want to control the steady-state behavior of the system.
- Suppose the system is operating at θ. A natural question to ask is the following: *Would it be better to change θ a little bit?*

Challenge

Compare $\pi_{\theta+\Delta}$ with π_θ in an efficient way.

Nota Bene: Never forget that we are doing this only because we cannot do better, or as Hinderer said: "In the long-run we are all dead."

The Stationary Control Model: The Basic Set Up

- Let $\Theta \subset \mathbb{R}$ denote the set of feasible parameters.
- Let P_θ denote a Markov kernel on $(S, \mathcal{S})$.
- Assume that for each $\theta \in \Theta$, P_θ admits a unique stationary distribution π_θ.
- Assume that the group inverse of P_θ exists for all $\theta \in \Theta$.

The Stationary Control Model: The Basic Set Up

- Let $\Theta \subset \mathbb{R}$ denote the set of feasible parameters.
- Let P_θ denote a Markov kernel on $(S, \mathcal{S})$.
- Assume that for each $\theta \in \Theta$, P_θ admits a unique stationary distribution π_θ.
- Assume that the group inverse of P_θ exists for all $\theta \in \Theta$.

The Stationary Control Model: The Basic Set Up

- Let $\Theta \subset \mathbb{R}$ denote the set of feasible parameters.
- Let P_θ denote a Markov kernel on $(S, \mathcal{S})$.
- Assume that for each $\theta \in \Theta$, P_θ admits a unique stationary distribution π_θ.
- Assume that the group inverse of P_θ exists for all $\theta \in \Theta$.

The Stationary Control Model: The Basic Set Up

- Let $\Theta \subset \mathbb{R}$ denote the set of feasible parameters.
- Let P_θ denote a Markov kernel on $(S, \mathcal{S})$.
- Assume that for each $\theta \in \Theta$, P_θ admits a unique stationary distribution π_θ.
- Assume that the group inverse of P_θ exists for all $\theta \in \Theta$.

The Stationary Control Model: The Basic Set Up

- Let $\Theta \subset \mathbb{R}$ denote the set of feasible parameters.
- Let P_θ denote a Markov kernel on $(S, \mathcal{S})$.
- Assume that for each $\theta \in \Theta$, P_θ admits a unique stationary distribution π_θ.
- Assume that the group inverse of P_θ exists for all $\theta \in \Theta$.

What is the Group Inverse?

- Let

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} P_\theta^n = \Pi_\theta,$$

provided the limit exists. Then for any probability distribution μ on $(S, \mathcal{S})$ it holds that $\mu \Pi_\theta = \pi_\theta$.
- In the uni-chain case, Π_θ is a matrix with rows equal to π_θ.
- In the multi-chain case, Π_θ has a simple block structure where each block corresponds to the ergodic projector of that particular ergodic class.

The *group inverse* of P_θ is defined as

$$(I - P_\theta + \Pi_\theta)^{-1} = \sum_{n\geq 0} (P_\theta - \Pi_\theta)^n,$$

provided it exists.

What is the Group Inverse?

- Let

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} P_\theta^n = \Pi_\theta,$$

provided the limit exists. Then for any probability distribution μ on $(S, \mathcal{S})$ it holds that $\mu \Pi_\theta = \pi_\theta$.

- In the uni-chain case, Π_θ is a matrix with rows equal to π_θ.
- In the multi-chain case, Π_θ has a simple block structure where each block corresponds to the ergodic projector of that particular ergodic class.

The *group inverse* of P_θ is defined as

$$(I - P_\theta + \Pi_\theta)^{-1} = \sum_{n \geq 0} (P_\theta - \Pi_\theta)^n,$$

provided it exists.

What is the Group Inverse?

- Let
$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} P_\theta^n = \Pi_\theta,$$
provided the limit exists. Then for any probability distribution μ on $(S, \mathcal{S})$ it holds that $\mu\Pi_\theta = \pi_\theta$.
- In the uni-chain case, Π_θ is a matrix with rows equal to π_θ.
- In the multi-chain case, Π_θ has a simple block structure where each block corresponds to the ergodic projector of that particular ergodic class.

The *group inverse* of P_θ is defined as

$$(I - P_\theta + \Pi_\theta)^{-1} = \sum_{n\geq 0} (P_\theta - \Pi_\theta)^n,$$

provided it exists.

What is the Group Inverse?

- Let

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} P_\theta^n = \Pi_\theta,$$

provided the limit exists. Then for any probability distribution μ on $(S, \mathcal{S})$ it holds that $\mu\Pi_\theta = \pi_\theta$.

- In the uni-chain case, Π_θ is a matrix with rows equal to π_θ.
- In the multi-chain case, Π_θ has a simple block structure where each block corresponds to the ergodic projector of that particular ergodic class.

The *group inverse* of P_θ is defined as

$$(I - P_\theta + \Pi_\theta)^{-1} = \sum_{n\geq 0} (P_\theta - \Pi_\theta)^n,$$

provided it exists.

What is the Group Inverse?

- Let
$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} P_\theta^n = \Pi_\theta,$$
provided the limit exists. Then for any probability distribution μ on $(S, \mathcal{S})$ it holds that $\mu \Pi_\theta = \pi_\theta$.
- In the uni-chain case, Π_θ is a matrix with rows equal to π_θ.
- In the multi-chain case, Π_θ has a simple block structure where each block corresponds to the ergodic projector of that particular ergodic class.

The *group inverse* of P_θ is defined as

$$(I - P_\theta + \Pi_\theta)^{-1} = \sum_{n \geq 0} (P_\theta - \Pi_\theta)^n,$$

provided it exists.

What does the Group Inverse Mean?

Note that

$$\sum_{n\geq 0}(P_\theta - \Pi_\theta)^n = \underbrace{\sum_{n\geq 0}(P_\theta^n - \Pi_\theta)}_{=:D_\theta} + \Pi_\theta,$$

where D_θ is called the *deviation matrix* of P_θ.

Deviation Matrix

The group inverse/deviation matrix measures the speed of convergence of P_θ to its stationary regime.

For cost vector f, $D_P f$ is called the value function in MDP.

Existence is guaranteed in the finite case, otherwise via geometric (!) ergodicity

What does the Group Inverse Mean?

Note that

$$\sum_{n \geq 0} (P_\theta - \Pi_\theta)^n = \underbrace{\sum_{n \geq 0} (P_\theta^n - \Pi_\theta)}_{=:D_\theta} + \Pi_\theta,$$

where D_θ is called the *deviation matrix* of P_θ.

Deviation Matrix

The group inverse/deviation matrix measures the speed of convergence of P_θ to its stationary regime.

For cost vector f, $D_P f$ is called the value function in MDP.

Existence is guaranteed in the finite case, otherwise via geometric (!) ergodicity

What does the Group Inverse Mean?

Note that

$$\sum_{n\geq 0}(P_\theta - \Pi_\theta)^n = \underbrace{\sum_{n\geq 0}(P_\theta^n - \Pi_\theta)}_{=:D_\theta} + \Pi_\theta,$$

where D_θ is called the *deviation matrix* of P_θ.

Deviation Matrix

The group inverse/deviation matrix measures the speed of convergence of P_θ to its stationary regime.

For cost vector f, $D_P f$ is called the value function in MDP.

Existence is guaranteed in the finite case, otherwise via geometric (!) ergodicity

What does the Group Inverse Mean?

Note that

$$\sum_{n\geq 0}(P_\theta - \Pi_\theta)^n = \underbrace{\sum_{n\geq 0}(P_\theta^n - \Pi_\theta)}_{=:D_\theta} + \Pi_\theta,$$

where D_θ is called the *deviation matrix* of P_θ.

Deviation Matrix

The group inverse/deviation matrix measures the speed of convergence of P_θ to its stationary regime.

For cost vector f, $D_P f$ is called the value function in MDP.

Existence is guaranteed in the finite case, otherwise via geometric (!) ergodicity

What does the Group Inverse Mean?

Note that

$$\sum_{n\geq 0}(P_\theta - \Pi_\theta)^n = \underbrace{\sum_{n\geq 0}(P_\theta^n - \Pi_\theta)}_{=:D_\theta} + \Pi_\theta,$$

where D_θ is called the *deviation matrix* of P_θ.

Deviation Matrix

The group inverse/deviation matrix measures the speed of convergence of P_θ to its stationary regime.

For cost vector f, $D_P f$ is called the value function in MDP.

Existence is guaranteed in the finite case, otherwise via geometric (!) ergodicity

What does the Group Inverse Mean?

Note that

$$\sum_{n\geq 0}(P_\theta - \Pi_\theta)^n = \underbrace{\sum_{n\geq 0}(P_\theta^n - \Pi_\theta)}_{=:D_\theta} + \Pi_\theta,$$

where D_θ is called the *deviation matrix* of P_θ.

Deviation Matrix

The group inverse/deviation matrix measures the speed of convergence of P_θ to its stationary regime.

For cost vector f, $D_P f$ is called the value function in MDP.

Existence is guaranteed in the finite case, otherwise via geometric (!) ergodicity

All very nice but can we use the Deviation Matrix in Control?

Theorem 1

For $\Delta + \theta, \theta \in \Theta$, it holds that

$$\Pi_{\theta+\Delta} = \Pi_\theta + \Pi_{\theta+\Delta}(P_{\theta+\Delta} - P_\theta)D_\theta$$

Lets proof this basic fact. By simple algebra

$$\begin{aligned} P_\theta D_\theta &= P_\theta \sum_{n=0}^{\infty}(P_\theta^n - \Pi_\theta) \\ &= P_\theta \left(I - \Pi_\theta + \sum_{n=1}^{\infty}(P_\theta^n - \Pi_\theta) \right) = \Pi_\theta - I + D_\theta, \end{aligned}$$

which yields $I - \Pi_\theta = (I - P_\theta)D_\theta$, or $I = \Pi_\theta + (I - P_\theta)D_\theta$.
Multiplying by $\Pi_{\theta+\Delta}$ (and noting that $\Pi_{\theta+\Delta}\Pi_\theta = \Pi_\theta$, and $\Pi_\theta I = \Pi_\theta P_\theta$) proves the claim.

All very nice but can we use the Deviation Matrix in Control?

Theorem 1

For $\Delta + \theta, \theta \in \Theta$, it holds that

$$\Pi_{\theta+\Delta} = \Pi_\theta + \Pi_{\theta+\Delta}(P_{\theta+\Delta} - P_\theta)D_\theta$$

Lets proof this basic fact. By simple algebra

$$\begin{aligned} P_\theta D_\theta &= P_\theta \sum_{n=0}^{\infty}(P_\theta^n - \Pi_\theta) \\ &= P_\theta \left(I - \Pi_\theta + \sum_{n=1}^{\infty}(P_\theta^n - \Pi_\theta) \right) = \Pi_\theta - I + D_\theta, \end{aligned}$$

which yields $I - \Pi_\theta = (I - P_\theta)D_\theta$, or $I = \Pi_\theta + (I - P_\theta)D_\theta$.
Multiplying by $\Pi_{\theta+\Delta}$ (and noting that $\Pi_{\theta+\Delta}\Pi_\theta = \Pi_\theta$, and $\Pi_\theta I = \Pi_\theta P_\theta$) proves the claim.

All very nice but can we use the Deviation Matrix in Control?

Theorem 1

For $\Delta + \theta, \theta \in \Theta$, it holds that

$$\Pi_{\theta+\Delta} = \Pi_\theta + \Pi_{\theta+\Delta}(P_{\theta+\Delta} - P_\theta)D_\theta$$

Lets proof this basic fact. By simple algebra

$$\begin{aligned} P_\theta D_\theta &= P_\theta \sum_{n=0}^{\infty}(P_\theta^n - \Pi_\theta) \\ &= P_\theta \left(I - \Pi_\theta + \sum_{n=1}^{\infty}(P_\theta^n - \Pi_\theta) \right) = \Pi_\theta - I + D_\theta, \end{aligned}$$

which yields $I - \Pi_\theta = (I - P_\theta)D_\theta$, or $I = \Pi_\theta + (I - P_\theta)D_\theta$.
Multiplying by $\Pi_{\theta+\Delta}$ (and noting that $\Pi_{\theta+\Delta}\Pi_\theta = \Pi_\theta$, and $\Pi_\theta I = \Pi_\theta P_\theta$) proves the claim.

All very nice but can we use the Deviation Matrix in Control?

Theorem 1

For $\Delta + \theta, \theta \in \Theta$, it holds that

$$\Pi_{\theta+\Delta} = \Pi_\theta + \Pi_{\theta+\Delta}(P_{\theta+\Delta} - P_\theta)D_\theta$$

Lets proof this basic fact. By simple algebra

$$\begin{aligned} P_\theta D_\theta &= P_\theta \sum_{n=0}^{\infty}(P_\theta^n - \Pi_\theta) \\ &= P_\theta \left(I - \Pi_\theta + \sum_{n=1}^{\infty}(P_\theta^n - \Pi_\theta) \right) = \Pi_\theta - I + D_\theta, \end{aligned}$$

which yields $I - \Pi_\theta = (I - P_\theta)D_\theta$, or $I = \Pi_\theta + (I - P_\theta)D_\theta$.
Multiplying by $\Pi_{\theta+\Delta}$ (and noting that $\Pi_{\theta+\Delta}\Pi_\theta = \Pi_\theta$, and $\Pi_\theta I = \Pi_\theta P_\theta$) proves the claim.

All very nice but can we use the Deviation Matrix in Control?

Theorem 1

For $\Delta + \theta, \theta \in \Theta$, it holds that

$$\Pi_{\theta+\Delta} = \Pi_\theta + \Pi_{\theta+\Delta}(P_{\theta+\Delta} - P_\theta)D_\theta$$

Lets proof this basic fact. By simple algebra

$$\begin{aligned} P_\theta D_\theta &= P_\theta \sum_{n=0}^{\infty}(P_\theta^n - \Pi_\theta) \\ &= P_\theta \left(I - \Pi_\theta + \sum_{n=1}^{\infty}(P_\theta^n - \Pi_\theta) \right) = \Pi_\theta - I + D_\theta, \end{aligned}$$

which yields $I - \Pi_\theta = (I - P_\theta)D_\theta$, or $I = \Pi_\theta + (I - P_\theta)D_\theta$.

Multiplying by $\Pi_{\theta+\Delta}$ (and noting that $\Pi_{\theta+\Delta}\Pi_\theta = \Pi_\theta$, and $\Pi_\theta I = \Pi_\theta P_\theta$) proves the claim.

All very nice but can we use the Deviation Matrix in Control?

Theorem 1

For $\Delta + \theta, \theta \in \Theta$, it holds that

$$\Pi_{\theta+\Delta} = \Pi_\theta + \Pi_{\theta+\Delta}(P_{\theta+\Delta} - P_\theta)D_\theta$$

Lets proof this basic fact. By simple algebra

$$\begin{aligned} P_\theta D_\theta &= P_\theta \sum_{n=0}^{\infty}(P_\theta^n - \Pi_\theta) \\ &= P_\theta \left(I - \Pi_\theta + \sum_{n=1}^{\infty}(P_\theta^n - \Pi_\theta) \right) = \Pi_\theta - I + D_\theta, \end{aligned}$$

which yields $I - \Pi_\theta = (I - P_\theta)D_\theta$, or $I = \Pi_\theta + (I - P_\theta)D_\theta$.
Multiplying by $\Pi_{\theta+\Delta}$ (and noting that $\Pi_{\theta+\Delta}\Pi_\theta = \Pi_\theta$, and $\Pi_\theta I = \Pi_\theta P_\theta$) proves the claim.

Should we switch from θ to $\theta + \Delta$?

Suppose you want to minimize $\pi_\theta f$.

By Theorem 1 (in vectorial form) it holds that

$$\pi_{\theta+\Delta} f - \pi_\theta f = \pi_{\theta+\Delta}(P_{\theta+\Delta} - P_\theta) D_\theta f.$$

Since $\pi_{\theta+\Delta}$ is positive, we have the following

- condition for switching from θ to $\theta + \Delta$:

$$\pi_{\theta+\Delta} f - \pi_\theta f \leq 0 \ \text{ if } \ (P_{\theta+\Delta} - P_\theta) D_\theta f \leq 0,$$

- condition for not switching from θ to $\theta + \Delta$:

$$\pi_{\theta+\Delta} f - \pi_\theta f \geq 0 \ \text{ if } \ (P_{\theta+\Delta} - P_\theta) D_\theta f \geq 0.$$

Should we switch from θ to $\theta + \Delta$?

Suppose you want to minimize $\pi_\theta f$.

By Theorem 1 (in vectorial form) it holds that

$$\pi_{\theta+\Delta} f - \pi_\theta f = \pi_{\theta+\Delta}(P_{\theta+\Delta} - P_\theta) D_\theta f.$$

Since $\pi_{\theta+\Delta}$ is positive, we have the following

- condition for switching from θ to $\theta + \Delta$:

$$\pi_{\theta+\Delta} f - \pi_\theta f \leq 0 \;\text{ if }\; (P_{\theta+\Delta} - P_\theta) D_\theta f \leq 0,$$

- condition for not switching from θ to $\theta + \Delta$:

$$\pi_{\theta+\Delta} f - \pi_\theta f \geq 0 \;\text{ if }\; (P_{\theta+\Delta} - P_\theta) D_\theta f \geq 0.$$

Should we switch from θ to $\theta + \Delta$?

Suppose you want to minimize $\pi_\theta f$.

By Theorem 1 (in vectorial form) it holds that

$$\pi_{\theta+\Delta} f - \pi_\theta f = \pi_{\theta+\Delta}(P_{\theta+\Delta} - P_\theta) D_\theta f.$$

Since $\pi_{\theta+\Delta}$ is positive, we have the following

- condition for switching from θ to $\theta + \Delta$:

$$\pi_{\theta+\Delta} f - \pi_\theta f \leq 0 \ \text{ if } \ (P_{\theta+\Delta} - P_\theta) D_\theta f \leq 0,$$

- condition for not switching from θ to $\theta + \Delta$:

$$\pi_{\theta+\Delta} f - \pi_\theta f \geq 0 \ \text{ if } \ (P_{\theta+\Delta} - P_\theta) D_\theta f \geq 0.$$

Should we switch from θ to $\theta + \Delta$?

Suppose you want to minimize $\pi_\theta f$.

By Theorem 1 (in vectorial form) it holds that

$$\pi_{\theta+\Delta} f - \pi_\theta f = \pi_{\theta+\Delta}(P_{\theta+\Delta} - P_\theta) D_\theta f.$$

Since $\pi_{\theta+\Delta}$ is positive, we have the following

- condition for switching from θ to $\theta + \Delta$:

$$\pi_{\theta+\Delta} f - \pi_\theta f \leq 0 \text{ if } (P_{\theta+\Delta} - P_\theta) D_\theta f \leq 0,$$

- condition for not switching from θ to $\theta + \Delta$:

$$\pi_{\theta+\Delta} f - \pi_\theta f \geq 0 \text{ if } (P_{\theta+\Delta} - P_\theta) D_\theta f \geq 0.$$

Should we switch from θ to $\theta + \Delta$?

Suppose you want to minimize $\pi_\theta f$.

By Theorem 1 (in vectorial form) it holds that

$$\pi_{\theta+\Delta} f - \pi_\theta f = \pi_{\theta+\Delta}(P_{\theta+\Delta} - P_\theta) D_\theta f.$$

Since $\pi_{\theta+\Delta}$ is positive, we have the following

- condition for switching from θ to $\theta + \Delta$:

$$\pi_{\theta+\Delta} f - \pi_\theta f \leq 0 \ \text{ if } \ (P_{\theta+\Delta} - P_\theta) D_\theta f \leq 0,$$

- condition for not switching from θ to $\theta + \Delta$:

$$\pi_{\theta+\Delta} f - \pi_\theta f \geq 0 \ \text{ if } \ (P_{\theta+\Delta} - P_\theta) D_\theta f \geq 0.$$

Should we switch from θ to $\theta + \Delta$?

Suppose you want to minimize $\pi_\theta f$.

By Theorem 1 (in vectorial form) it holds that

$$\pi_{\theta+\Delta} f - \pi_\theta f = \pi_{\theta+\Delta}(P_{\theta+\Delta} - P_\theta) D_\theta f.$$

Since $\pi_{\theta+\Delta}$ is positive, we have the following

- condition for switching from θ to $\theta + \Delta$:

$$\pi_{\theta+\Delta} f - \pi_\theta f \leq 0 \text{ if } (P_{\theta+\Delta} - P_\theta) D_\theta f \leq 0,$$

- condition for not switching from θ to $\theta + \Delta$:

$$\pi_{\theta+\Delta} f - \pi_\theta f \geq 0 \text{ if } (P_{\theta+\Delta} - P_\theta) D_\theta f \geq 0.$$

Can we do better?

Since a Markov transition probability is a collection of conditional probabilities we can construct a new kernel $\hat{P}_\theta$ as follows.

We check for each state s what the best choice for θ is, i.e., we solve

$$\min_{\theta' \in \Theta} (P_{\theta'}(s) - P_\theta(s)) D_\theta f$$

In Markov decision process theory this is called *policy iteration*.

Can we do better?

Since a Markov transition probability is a collection of conditional probabilities we can construct a new kernel $\hat{P}_\theta$ as follows.

We check for each state s what the best choice for θ is, i.e., we solve

$$\min_{\theta' \in \Theta} (P_{\theta'}(s) - P_\theta(s)) D_\theta f$$

In Markov decision process theory this is called *policy iteration*.

Can we do better?

Since a Markov transition probability is a collection of conditional probabilities we can construct a new kernel $\hat{P}_\theta$ as follows.

We check for each state s what the best choice for θ is, i.e., we solve

$$\min_{\theta' \in \Theta} (P_{\theta'}(s) - P_\theta(s)) D_\theta f$$

In Markov decision process theory this is called *policy iteration*.

Can we do better?

Since a Markov transition probability is a collection of conditional probabilities we can construct a new kernel $\hat{P}_\theta$ as follows.

We check for each state s what the best choice for θ is, i.e., we solve

$$\min_{\theta' \in \Theta} (P_{\theta'}(s) - P_\theta(s)) D_\theta f$$

In Markov decision process theory this is called *policy iteration*.

Is it Applicable?

The problem with any policy-iteration-like approach is that D_θ is usually not available in closed analytical form.

An obvious analytical way around this is to work with

$$\sum_{n=0}^{k}(P_\theta^n - \Pi_\theta)$$

as *approximation* of D_θ, and

$$\sum_{n=0}^{k}(P_\theta^n - \Pi_\theta)f$$

as *approximate value function* for cost function f.

Is it Applicable?

The problem with any policy-iteration-like approach is that D_θ is usually not available in closed analytical form.

An obvious analytical way around this is to work with

$$\sum_{n=0}^{k} (P_\theta^n - \Pi_\theta)$$

as *approximation* of D_θ, and

$$\sum_{n=0}^{k} (P_\theta^n - \Pi_\theta) f$$

as *approximate value function* for cost function f.

Is it Applicable?

The problem with any policy-iteration-like approach is that D_θ is usually not available in closed analytical form.

An obvious analytical way around this is to work with

$$\sum_{n=0}^{k}(P_\theta^n - \Pi_\theta)$$

as *approximation* of D_θ, and

$$\sum_{n=0}^{k}(P_\theta^n - \Pi_\theta)f$$

as *approximate value function* for cost function f.

Online Control

Let's Take a Step back: What is really needed?

Note that in our approximative formula for $\pi_{\theta+\Delta} f - \pi_\theta f$ we actually need the term

$$(P_{\theta+\Delta} - P_\theta) D_\theta f,$$

which can be rewritten as follows

$$\begin{aligned}
(P_{\theta+\Delta} - P_\theta) D_\theta f &= (P_{\theta+\Delta} - P_\theta) \sum_{n \geq 0} (P_\theta^n - \Pi_\theta) f \\
&= \sum_{n \geq 0} (P_{\theta+\Delta} - P_\theta)(P_\theta^n - \Pi_\theta) f \\
&= \sum_{n=0}^{\infty} (P_{\theta+\Delta} - P_\theta) P_\theta^n f.
\end{aligned}$$

Let's Take a Step back: What is really needed?

Note that in our approximative formula for $\pi_{\theta+\Delta} f - \pi_\theta f$ we actually need the term

$$(P_{\theta+\Delta} - P_\theta) D_\theta f,$$

which can be rewritten as follows

$$\begin{aligned}
(P_{\theta+\Delta} - P_\theta) D_\theta f &= (P_{\theta+\Delta} - P_\theta) \sum_{n \geq 0} (P_\theta^n - \Pi_\theta) f \\
&= \sum_{n \geq 0} (P_{\theta+\Delta} - P_\theta)(P_\theta^n - \Pi_\theta) f \\
&= \sum_{n=0}^{\infty} (P_{\theta+\Delta} - P_\theta) P_\theta^n f.
\end{aligned}$$

Let's Take a Step back: What is really needed?

Note that in our approximative formula for $\pi_{\theta+\Delta} f - \pi_\theta f$ we actually need the term

$$(P_{\theta+\Delta} - P_\theta) D_\theta f,$$

which can be rewritten as follows

$$\begin{aligned}
(P_{\theta+\Delta} - P_\theta) D_\theta f &= (P_{\theta+\Delta} - P_\theta) \sum_{n \geq 0} (P_\theta^n - \Pi_\theta) f \\
&= \sum_{n \geq 0} (P_{\theta+\Delta} - P_\theta)(P_\theta^n - \Pi_\theta) f \\
&= \sum_{n=0}^{\infty} (P_{\theta+\Delta} - P_\theta) P_\theta^n f.
\end{aligned}$$

Introduce the Potential Matrix

- For given cost function f, let $g_\theta(s)$ be defined as follows

$$g_\theta(s) = \sum_{n=0}^{\infty}(P_\theta^n(s)f - \pi_\theta f).$$

  The vector g_θ is called *bias vector* in MDP.
- The matrix

$$H_\theta(s,u) = g_\theta(s) - g_\theta(u), \quad s,u \in S$$

  is called the *potential matrix*.

Introduce the Potential Matrix

- For given cost function f, let $g_\theta(s)$ be defined as follows

$$g_\theta(s) = \sum_{n=0}^{\infty}(P_\theta^n(s)f - \pi_\theta f).$$

The vector g_θ is called *bias vector* in MDP.

- The matrix

$$H_\theta(s,u) = g_\theta(s) - g_\theta(u), \quad s,u \in S$$

is called the *potential matrix*.

Introduce the Potential Matrix

- For given cost function f, let $g_\theta(s)$ be defined as follows

$$g_\theta(s) = \sum_{n=0}^{\infty} (P_\theta^n(s)f - \pi_\theta f).$$

  The vector g_θ is called *bias vector* in MDP.

- The matrix

$$H_\theta(s,u) = g_\theta(s) - g_\theta(u), \quad s,u \in S$$

  is called the *potential matrix*.

Estimating the Potential Matrix

- The potential matrix can be written as

$$H_\theta(s,u) = \sum_{n=0}^{\infty} (P_\theta^n(s)f - P_\theta^n(u)f).$$

- Let $\tau_\theta(s,u)$ be the time until the version started in s and the version started in u couple, then

$$H_\theta(s,u) = \mathbb{E}\left[\sum_{n=0}^{\tau_\theta(s,u)} f(X_\theta(n,s)) - f(X_\theta(n,u))\right],$$

where $X_\theta(n,r)$ the n-th state of a P_θ-Markov chain started in state $r \in S$.

Note that $H_\theta(s,u)$ can be estimated from the observing a single-sample path using cut-and-past methods.

Estimating the Potential Matrix

- The potential matrix can be written as

$$H_\theta(s,u) = \sum_{n=0}^{\infty} (P_\theta^n(s)f - P_\theta^n(u)f).$$

- Let $\tau_\theta(s,u)$ be the time until the version started in s and the version started in u couple, then

$$H_\theta(s,u) = \mathbb{E}\left[\sum_{n=0}^{\tau_\theta(s,u)} f(X_\theta(n,s)) - f(X_\theta(n,u))\right],$$

where $X_\theta(n,r)$ the n-th state of a P_θ-Markov chain started in state $r \in S$.

Note that $H_\theta(s,u)$ can be estimated from the observing a single-sample path using cut-and-past methods.

Estimating the Potential Matrix

- The potential matrix can be written as

$$H_\theta(s,u) = \sum_{n=0}^{\infty} (P_\theta^n(s)f - P_\theta^n(u)f).$$

- Let $\tau_\theta(s,u)$ be the time until the version started in s and the version started in u couple, then

$$H_\theta(s,u) = \mathbb{E}\left[\sum_{n=0}^{\tau_\theta(s,u)} f(X_\theta(n,s)) - f(X_\theta(n,u))\right],$$

where $X_\theta(n,r)$ the n-th state of a P_θ-Markov chain started in state $r \in S$.

Note that $H_\theta(s,u)$ can be estimated from the observing a single-sample path using cut-and-past methods.

Estimating the Potential Matrix

- The potential matrix can be written as

$$H_\theta(s,u) = \sum_{n=0}^{\infty} (P_\theta^n(s)f - P_\theta^n(u)f).$$

- Let $\tau_\theta(s,u)$ be the time until the version started in s and the version started in u couple, then

$$H_\theta(s,u) = \mathbb{E}\left[\sum_{n=0}^{\tau_\theta(s,u)} f(X_\theta(n,s)) - f(X_\theta(n,u))\right],$$

where $X_\theta(n,r)$ the n-th state of a P_θ-Markov chain started in state $r \in S$.

Note that $H_\theta(s,u)$ can be estimated from the observing a single-sample path using cut-and-past methods.

Estimating the Potential Matrix

- The potential matrix can be written as

$$H_\theta(s,u) = \sum_{n=0}^{\infty} (P_\theta^n(s)f - P_\theta^n(u)f).$$

- Let $\tau_\theta(s,u)$ be the time until the version started in s and the version started in u couple, then

$$H_\theta(s,u) = \mathbb{E}\left[\sum_{n=0}^{\tau_\theta(s,u)} f(X_\theta(n,s)) - f(X_\theta(n,u))\right],$$

where $X_\theta(n,r)$ the n-th state of a P_θ-Markov chain started in state $r \in S$.

Note that $H_\theta(s,u)$ can be estimated from the observing a single-sample path using cut-and-past methods.

How to use the Potential Matrix in Control?

- Recall that we are want to evaluate

$$(P_{\theta+\Delta} - P_\theta)D_\theta f.$$

- Using the potential matrix, this can be written as

$$(P_{\theta+\Delta}(s) - P_\theta(s))D_\theta f = \int_{S\times S} (P_{\theta+\Delta}(s, ds') - P_\theta(s, du'))H_\theta(s', u')$$

- The above formula can be made useful for on-line control.

How to use the Potential Matrix in Control?

- Recall that we are want to evaluate

$$(P_{\theta+\Delta} - P_{\theta})D_{\theta}f.$$

- Using the potential matrix, this can be written as

$$(P_{\theta+\Delta}(s) - P_{\theta}(s))D_{\theta}f = \int_{S\times S}(P_{\theta+\Delta}(s,ds') - P_{\theta}(s,du'))H_{\theta}(s',u')$$

- The above formula can be made useful for on-line control.

How to use the Potential Matrix in Control?

- Recall that we are want to evaluate

$$(P_{\theta+\Delta} - P_\theta)D_\theta f.$$

- Using the potential matrix, this can be written as

$$(P_{\theta+\Delta}(s) - P_\theta(s))D_\theta f = \int_{S\times S} (P_{\theta+\Delta}(s, ds') - P_\theta(s, du'))H_\theta(s', u')$$

- The above formula can be made useful for on-line control.

How to use the Potential Matrix in Control?

- Recall that we are want to evaluate

$$(P_{\theta+\Delta} - P_\theta)D_\theta f.$$

- Using the potential matrix, this can be written as

$$(P_{\theta+\Delta}(s) - P_\theta(s))D_\theta f = \int_{S\times S} (P_{\theta+\Delta}(s, ds') - P_\theta(s, du'))H_\theta(s', u')$$

- The above formula can be made useful for on-line control.

On-line Control and Sample-Path based learning

- Suppose the chain is in state $s = X_\theta(n)$.
- Then simulate then next state according to $P_{\theta+\Delta}$, denoted by $X_{\theta+\Delta}(s)$, and the next state under P_θ, denoted by $X_\theta(s)$.
- If
$$H_\theta(X_{\theta+\Delta}(s), X_\theta(s)) \leq 0,$$
then in state s, the design parameter should be switched to $\theta + \Delta$ (in order to minimize costs) .

This way the best choice for each state s can be found. This leads to *on-line control*, resp. *on-line learning*.

On-line Control and Sample-Path based learning

- Suppose the chain is in state $s = X_\theta(n)$.
- Then simulate then next state according to $P_{\theta+\Delta}$, denoted by $X_{\theta+\Delta}(s)$, and the next state under P_θ, denoted by $X_\theta(s)$.
- If

$$H_\theta(X_{\theta+\Delta}(s), X_\theta(s)) \leq 0,$$

  then in state s, the design parameter should be switched to $\theta + \Delta$ (in order to minimize costs) .

This way the best choice for each state s can be found. This leads to *on-line control*, resp. *on-line learning*.

On-line Control and Sample-Path based learning

- Suppose the chain is in state $s = X_\theta(n)$.
- Then simulate then next state according to $P_{\theta+\Delta}$, denoted by $X_{\theta+\Delta}(s)$, and the next state under P_θ, denoted by $X_\theta(s)$.
- If
$$H_\theta(X_{\theta+\Delta}(s), X_\theta(s)) \leq 0,$$
then in state s, the design parameter should be switched to $\theta + \Delta$ (in order to minimize costs) .

This way the best choice for each state s can be found. This leads to *on-line control*, resp. *on-line learning*.

On-line Control and Sample-Path based learning

- Suppose the chain is in state $s = X_\theta(n)$.
- Then simulate then next state according to $P_{\theta+\Delta}$, denoted by $X_{\theta+\Delta}(s)$, and the next state under P_θ, denoted by $X_\theta(s)$.
- If

$$H_\theta(X_{\theta+\Delta}(s), X_\theta(s)) \leq 0,$$

then in state s, the design parameter should be switched to $\theta + \Delta$ (in order to minimize costs) .

This way the best choice for each state s can be found. This leads to *on-line control*, resp. *on-line learning*.

On-line Control and Sample-Path based learning

- Suppose the chain is in state $s = X_\theta(n)$.
- Then simulate then next state according to $P_{\theta+\Delta}$, denoted by $X_{\theta+\Delta}(s)$, and the next state under P_θ, denoted by $X_\theta(s)$.
- If
$$H_\theta(X_{\theta+\Delta}(s), X_\theta(s)) \leq 0,$$
then in state s, the design parameter should be switched to $\theta + \Delta$ (in order to minimize costs) .

This way the best choice for each state s can be found. This leads to *on-line control*, resp. *on-line learning*.

On-line Control and Sample-Path based learning

- Suppose the chain is in state $s = X_\theta(n)$.
- Then simulate then next state according to $P_{\theta+\Delta}$, denoted by $X_{\theta+\Delta}(s)$, and the next state under P_θ, denoted by $X_\theta(s)$.
- If
$$H_\theta(X_{\theta+\Delta}(s), X_\theta(s)) \leq 0,$$
then in state s, the design parameter should be switched to $\theta + \Delta$ (in order to minimize costs) .

This way the best choice for each state s can be found. This leads to *on-line control*, resp. *on-line learning*.

Summary

- The basic techniques are rather straightforward.
- Markov decision processes techniques can be made fruitful for on-line control.
- Simulation offers an interesting alternative for computing the input data for on-line control (read the potential matrix).

Summary

- The basic techniques are rather straightforward.
- Markov decision processes techniques can be made fruitful for on-line control.
- Simulation offers an interesting alternative for computing the input data for on-line control (read the potential matrix).

Summary

- The basic techniques are rather straightforward.
- Markov decision processes techniques can be made fruitful for on-line control.
- Simulation offers an interesting alternative for computing the input data for on-line control (read the potential matrix).

Summary

- The basic techniques are rather straightforward.
- Markov decision processes techniques can be made fruitful for on-line control.
- Simulation offers an interesting alternative for computing the input data for on-line control (read the potential matrix).

Impressive!?

Well, I hope so, but don't forget that everything said applies only to the simple (=stationary) control problem.

Impressive!?

Well, I hope so, but don't forget that everything said applies only to the simple (=stationary) control problem.

Impressive!?

Well, I hope so, but don't forget that everything said applies only to the simple (=stationary) control problem.