Stochastic Optimization
IDA PhD course 2011ht

Stefanie Kosuch
PostDok at TCSLab
www.kosuch.eu/stefanie/

7. Lecture: Inner and Outer Approximation Methods
24. November 2011

Linköping University

Linköping University

Outline

Basics of Multi-Stage Programming

- Information obtained in several "stages"
- Recourse action can be taken in each stage
- Recourse action depends on:
  - → Information available
  - → Previous decisions
- Recourse action does not depend on:
  - → Information to be released later ($\neq$ det. multi-period optimization)
- Objective: Minimize total expected cost
- Crucial: Discretization of time and random distribution(s)

Linköping University

Main Differences to Online Optimization

- Statistic Information used
- Parameters modeled as random variables
- Decision(s) made "offline"
- Actual decision computed (for all scenarios)
- Objective: Good solution on average

Which approach to choose?

Questions to be answered:

- Do I have a finite, infinite or undefined number of periods/stages?
- Do I have statistical information about the uncertain parameters?
- Do I have the infrastructure and/or time to make decisions online?
- Will I use my solution once or several times?

Outline

General Linear Multi-Stage Model

$$\min_{x_0 \geq 0} \quad c_0^\intercal x_0 + \mathbb{E}_{\zeta_1}\left[\mathcal{Q}^1(x_{[0]}, \zeta_{[1]})\right]$$

$$\text{s.t.} \quad A_{00} x_0 \geq b_0,$$

$$\left.\begin{aligned} \mathcal{Q}^t(x_{[t-1]}, \zeta_{[t]}) = \min_{x_t \geq 0} \quad & c_t^\intercal x_t + \mathbb{E}_{\zeta_{t+1}|\zeta_{[t]}}\left[\mathcal{Q}^{t+1}(x_{[t]}, \zeta_{[t+1]})\right] \\ \text{s.t.} \quad & \sum_{s=0}^{t} A_{ts}(\zeta_{[t]}) x_s \geq b_t(\zeta_{[t]}). \end{aligned}\right\} \quad \forall t \in \{1, \dots, t-1\}$$

$$\mathcal{Q}(x_{[T-1]}, \zeta_{[T]}) = \min_{x_T \geq 0} \quad c_T^\intercal x_T$$

$$\text{s.t.} \quad \sum_{s=0}^{T} A_{Ts}(\zeta_{[T]}) x_s \geq b_T(\zeta_{[T]}).$$

$\zeta_{[t]} = (\zeta_1, \dots, \zeta_t)$
$x_{[t-1]} = (x_0, x_1, \dots, x_{t-1})$
$x_t = x_t(\zeta_{[t]}) \ \forall t \geq 1$

Linköping University

Outline

Linköping University

Definition

Scenario

- Sequence of outcomes $\zeta_1^k, \ldots, \zeta_T^k$ of $\zeta_1, \ldots, \zeta_T$
- Outcome $\zeta_{[T]}^k$ of $\zeta_{[T]} = (\zeta_1, \ldots, \zeta_T)$

Idea

Introduce sequence of decision vectors $(x_0^k, x_1^k, \ldots, x_T^k)$ ($\forall$ scenarios $\zeta_{[T]}^k$)

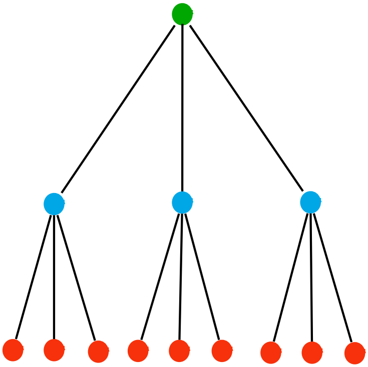
Problem

Some decision vectors should be equal (e.g. $x_0^k = x_0^{k'}$ $\forall k, k'$)

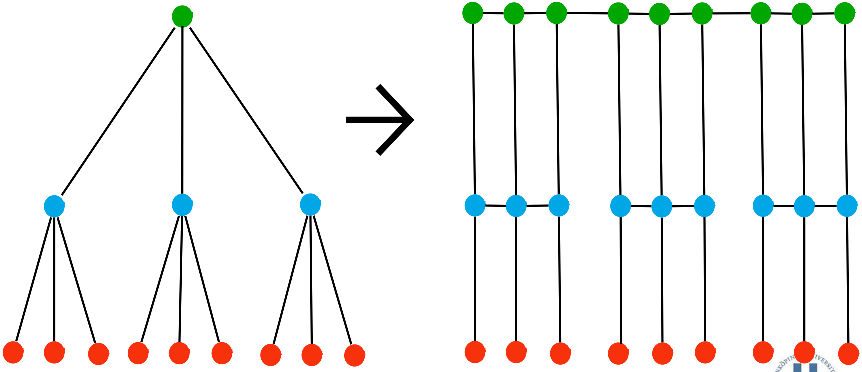
Solution

Non-anticipativity constraints: $x_t^k = x_t^{k'}$ if $\zeta_{[t]}^k = \zeta_{[t]}^{k'}$

Scenario Tree



Scenario Tree



Outline

Linköping University

Aim

Solve Stochastic Programming problems approximately...

1) ...whose objective function is computationally hard to evaluate.
2) ...that involve continuously distributed random parameters.
3) ...that involve too many scenarios.
4) ...whose underlying distribution is unknown ("Black-Box model").
5) ...are structural difficult (lack of convexity, linearity, etc.).

Linköping University

Outer Approximation

- Problem Approximated
- Underlying distribution replaced by finite random sample
- Deterministic reformulation
- Famous example: Sample Average Approach

Inner Approximation

- Randomized Solution Algorithm
- Sampling during solution process
- Either: Find good solution over iterations
- Or: Problem approximated over iterations
- Famous examples:

  → Stochastic gradient algorithm (Stochastic approximation)

  → Stochastic Decomposition

Outline

Linköping University

Basic Idea

- Sample uniformly N random vectors from distribution
- Assign probability $p_k = \frac{1}{N}$ $(k = 1, \ldots, N)$
- Replace underlying distribution by obtained discrete finite distribution
- Solve approximated problem $\mathcal{P}(\chi^1, \ldots, \chi^N)$
- Hopefully:

1) $x(\mathcal{P}(\chi^1, \ldots, \chi^N)) \to x^*$ as $N \to \infty$

2) $\mathbb{P}\{x(\mathcal{P}(\chi^1, \ldots, \chi^N)) = x^*\}$ increases as $N \to \infty$

Linköping University

Problem type (in general)

$$\begin{aligned} \min \quad & \mathbb{E}[f(x,\chi)] \\ \text{s.t.} \quad & x \in X \end{aligned}$$

X independent of distribution of random vector

Includes:

- Minimization of expected value
- Simple Recourse problems
- Two-Stage Programming problems

Linköping University

Problem type (in general)

$$\begin{aligned} \min \quad & \mathbb{E}[f(x,\chi)] \\ \text{s.t.} \quad & x \in X \end{aligned}$$

Simple-Recourse problem

$$\min_{x \in X} \quad f(x) + \sum_{i=1}^{m} d_i \cdot \mathbb{E}\left[[g_i(x,\chi)]^+\right]$$

- $d_i > 0 \; \forall i \in \{1, \dots, m\}$
- $[x]^+ = max(0, x)$

Linköping University

Problem type (in general)

$$\begin{aligned} \min \quad & \mathbb{E}[f(x,\chi)] \\ \text{s.t.} \quad & x \in X \end{aligned}$$

Linear Two-Stage Problem

$$\begin{aligned} \min_{x\geq 0} \quad & c^T x + \mathbb{E}[\mathcal{Q}(x,\chi)] \\ \text{s.t.} \quad & Ax \geq b, \\ & \mathcal{Q}(x,\chi) = \min_{y\geq 0} \quad d^T y \\ & \qquad\qquad\qquad \text{s.t.} \quad T(\chi)x + W(\chi)y \geq h(\chi). \end{aligned}$$

Problem type (in general)

$$\begin{aligned} \min \quad & \mathbb{E}[f(x,\chi)] \\ \text{s.t.} \quad & x \in X \end{aligned}$$

Does not (in general) include:

- Chance-Constrained programming problems
  (Feasible set described by CDF)
- Multi-Stage Programming (conditional sampling needed)

Chance-Constrained Stochastic Optimization Problem

$$\min_{x \in X} \quad f(x)$$
$$\text{s.t.} \quad \mathbb{P}\{G(x,\chi) \leq 0\} \geq p$$

Linköping University

Deterministic Opt. Model $\rightarrow$ Stochastic Programming Model

$$\begin{aligned} \max_{x \in X} \quad & f(x) \\ \text{s.t.} \quad & G(x) \leq 0 \end{aligned} \qquad \rightarrow \qquad \begin{aligned} \min_{x \in X} \quad & f(x, \chi) \\ \text{s.t.} \quad & G(x, \chi) \leq 0 \end{aligned}$$

$\chi \in \Omega \subseteq \mathbb{R}^s$: random vector

Stochastic Programming Model → Deterministic Equivalent Model

$$\min_{x \in X} \quad \mathbb{E}[f(x,\chi)]$$

$\chi \in \Omega \subseteq \mathbb{R}^s$: random vector

$$\xrightarrow{SAA}$$

$$\min_{x \in X} \quad \frac{1}{N}\sum_{k=1}^{N} f(x,\chi^k)$$

$\chi^1, \ldots, \chi^N$: random sample

Linköping University

Convergence results

- $\hat{x}_N$: optimal solution of SAA with N samples ($\hat{x}_N$ is random variable)
- A: set of optimal solutions of the exact problem

General distributions

Under some mild (technical) assumptions:

- $\text{dist}_A(\hat{x}_N) \rightarrow 0$ as $N \rightarrow \infty$ (w.p.1)
- Convergence rate: order $N^{-\frac{1}{2}}$
- $\forall \epsilon > 0$: $\mathbb{P}\{\text{dist}_A(\hat{x}_N) \leq \epsilon\}$ increases exponentially

Linköping University

Convergence results

- $\hat{x}_N$: optimal solution of SAA with N samples ($\hat{x}_N$ is random variable)
- A: set of optimal solutions of the exact problem

Discrete distributions

- $\forall \chi \in \Omega \; f(\cdot, \chi)$ is convex
- $\mathbb{E}[f(\cdot, \chi)]$ is well defined and is finite valued
- X is closed and convex
- the exact problem has a unique optimal solution x^* s.t.

$$\exists c > 0 : \forall x \in X : \quad \mathbb{E}[f(x, \chi)] - \mathbb{E}[f(x^*, \chi)] \geq c \cdot \text{dist}(x, x^*)$$

N large enough $\overset{wp1}{\Longrightarrow}$ SAA has a **unique optimal solution** $\mathbf{\hat{x}_N = x^*}$

Further results

- Sample size (e.g. for ϵ-exact solutions) (with given probability bound)
- Error bounds

Linköping University

Outline

Linköping University

Basic Idea

- Basically: Gradient method
- At each iteration: Sample random parameters
- Compute new solution based on this sample
- Use gradient of function **inside** expectation
- Hopefully:

1) $x^k \to x^*$ as $k \to \infty$ w.h.p.

Problem type (mostly)

$$\begin{aligned} \text{min} \quad & \mathbb{E}[f(x,\chi)] \\ \text{s.t.} \quad & x \in X \end{aligned}$$

Where:

- X independent of distribution of random vector
- X convex set
- $f(\cdot,\chi)$ convex
- $f(\cdot,\chi)$ differentiable (nearly everywhere)

- $r^k = \nabla_x f(x, \chi^k)$
- $(\epsilon^k)_{k \in \mathbb{N}}$ is a σ-sequence

Stochastic Gradient Algorithm

```
k ← 0
Choose x^0 in X
while k < K_max do
  k ← k + 1
  Draw χ^k = (χ_1^k, ..., χ_n^k)
  Update x^k as follows:

        x^{k+1} ← x^k + ε^k r^k

  Project x^{k+1} on X
end while
return  Best found or last solution
```

Linköping University

Assumptions

1) $\forall x \in X$: $f(x, \chi)$ is a random variable with finite expectation
2) $\forall \chi \in \Omega$: $f(\cdot, \chi)$ is convex, proper, differentiable
3) $\exists m > 0$ s.t.
$$\forall x \in X, \forall \chi \in \Omega : \|\nabla_x f(x, \chi)\| \leq m$$
4) $\exists$ set of optimal solutions X^* and $c > 0$ s.t.:
$$\forall x \in X, x^* \in X^* : \quad \mathbb{E}[f(x, \chi)] - \mathbb{E}[f(x^*, \chi)] \geq c \cdot (\text{dist}_{X^*}(x))^2$$

Theorem

Under assumptions 1) - 4) we have:

$$\lim_{k \to \infty} \mathbb{E}\left[(dist_{X^*}(x^k))^2\right] = 0$$

Linköping University

Assumptions

1) $\forall x \in X$: $f(x,\chi)$ is a random variable with finite expectation
2) $\forall \chi \in \Omega$: $f(\cdot,\chi)$ is convex, proper, differentiable
3) $\exists m > 0$ s.t.
$$\forall x \in X, \forall \chi \in \Omega : \|\nabla_x f(x,\chi)\| \leq m$$
4) $\exists$ set of optimal solutions X^* and $c > 0$ s.t.:
$$\forall x \in X, x^* \in X^* : \quad \mathbb{E}[f(x,\chi)] - \mathbb{E}[f(x^*,\chi)] \geq c \cdot (\text{dist}_{X^*}(x))^2$$

Theorem

Let $d^0 = (dist_{X^*}(x^0))^2$ *and* $\epsilon^k = \frac{1}{ck+\frac{m^2}{cd^0}}$.
Under assumptions 1) - 4) we have:

$$\mathbb{E}\left[(dist_{X^*}(x^k))^2\right] \leq \frac{1}{\frac{c^2}{m^2}k + \frac{1}{d^0}}$$

QUESTIONS?

What about next week?