# Multistage robust convex optimization problems: A sampling based approach

Fabrizio Dabbene/ Francesca Maggioni/ Georg Ch. Pflug

November 2019

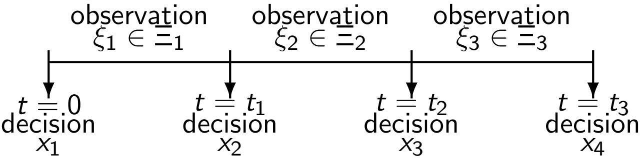
# Multistage robust (linear) programs

$$\begin{aligned}
\mathrm{RO}_{H+1} := \min_{x_1} c_1^\top x_1 + \\
+ \sup_{\xi_1 \in \Xi_1} \left[ \min_{x_2(\xi_1)} c_2^\top(\xi_1)\, x_2(\xi_1) + \sup_{\xi_2 \in \Xi_2} \left[ \cdots + \sup_{\xi_H \in \Xi_H} \left[ \min_{x_{H+1}(\underline{\xi}_H)} c_{H+1}^\top(\xi_H)\, x_{H+1}\left(\underline{\xi}_H\right) \right]\right]\right] \\
\text{s.t. } Ax_1 = h_1,\ x_1 \geq 0 \\
T_1(\xi_1)x_1 + W_2(\xi_1)x_2(\xi_1) = h_2(\xi_1),\ \forall \xi_1 \in \Xi_1 \\
\vdots \\
T_H(\xi_H)x_H(\underline{\xi}_{H-1}) + W_{H+1}(\xi_H)x_{H+1}(\underline{\xi}_H) = h_{H+1}(\xi_H),\ \forall \xi_H \in \Xi_H \\
x_t(\underline{\xi}_{t-1}) \geq 0 \quad \forall \xi_{t-1} \in \Xi_{t-1}; \qquad t = 2, \ldots, H+1\ ,
\end{aligned}$$

where $c_1 \in \mathbb{R}^{n_1}$ and $h_1 \in \mathbb{R}^{m_1}$ are known vectors and $A \in \mathbb{R}^{m_1 \times n_1}$ is a known matrix. The uncertain parameter vectors and matrices affected by the parameters $\xi_t \in \Xi_t$ are then given by $h_t \in \mathbb{R}^{m_t}$, $c_t \in \mathbb{R}^{n_t}$, $T_{t-1} \in \mathbb{R}^{m_t \times n_{t-1}}$, and $W_t \in \mathbb{R}^{m_t \times n_t}$, $t = 2, \ldots, H+1$.

$\Xi_t$ are compact sets in $\mathbb{R}^{k_t}$

# Non-anticipativity



# Replacing a huge infinite constraint set by a finite random extraction of the constraints

Consider the problem

$$\mathrm{RO} \quad : \quad \min_{x \in \mathbb{X}} \left\{ c^\top x : \sup_{\xi \in \Xi} f(x, \xi) \leq 0 \right\}, \tag{1}$$

where $x \in \mathbb{X} \subseteq \mathbb{R}^n$ is the optimization variable, $\mathbb{X}$ is convex and closed and $x \mapsto f(x, \xi) : \mathbb{X} \times \Xi \to \mathbb{R}$ is convex for all $\xi \in \Xi$. Suppose that $\Xi$ is compact and $\mathbb{P}$ is a probability measure on it with nonvanishing density. Let $\xi^{(1)}, \dots, \xi^{(N)}$ be independent samples from $\Xi$, sampled according to $\mathbb{P}^N = \mathbb{P} \times \cdots \times \mathbb{P}$. The "scenario" approximation of problem (1) is defined as follows

$$\mathrm{SO}^N \quad : \quad \min_{x \in \mathbb{X}} \left\{ c^\top x : \max_{1 \leq i \leq N} f(x, \xi^{(i)}) \leq 0 \right\}, \tag{2}$$

Problem $(SO^N)$ is a random problem and its solution is random. However it is solvable with standard solvers.

# The violation probability

Many authors have studied the approximation quality of (2) to the basic problem (1) coming up with convergence speed, cental limit type theorems and laws of large numbers.
It was the idea of Calafiore and Campi to look at the quality of the approximation in a different way, namely by studying the "violation probability distribution". The "violation probability" of the sample $\hat{\Xi}^N := \left\{\xi^{(1)}, \ldots, \xi^{(N)}\right\}$ is defined as

$$V(\hat{\Xi}^N) := \mathbb{P}\left\{\xi^{(N+1)} : \min_{x \in \mathbb{X}} \left\{c^\top x : \max_{1 \le i \le N+1} f(x, \xi^{(i)}) \le 0\right\} > v(\mathrm{SO}^N)\right\},$$

where also $\xi^{(N+1)}$ is sampled from $\mathbb{P}$. Here $v(\mathrm{SO}^N)$ is the optimal value of problem $\mathrm{SO}^N$). Notice that $V$ is a random variable taking its values in $[0, 1]$.

# Bounding the distribution of the violation probability

**Theorem.** [CCG Theorem, Calafiore (2010) and Campi/Garatti (2008)]. The distribution of $V$ under $\mathbb{P}$ is stochastically smaller (in the first order) than a random variable $Y_{N,n}$, which has the following compound distribution

$$Y_{N,n} = \begin{cases} 0, & \text{with probability } 1 - \binom{N}{n}^{-1} \\ Z_{N,n}, & \text{with probability } \binom{N}{n}^{-1}, \end{cases}$$

where $Z_{N,n}$ has a $Beta(n, N-n+1)$ distribution, that is for $\epsilon > 0$

$$\mathbb{P}\{V(\hat{\Xi}^N) > \epsilon\} \leq \mathbb{P}\{Y_{N,n} > \epsilon\} = n \int_\epsilon^1 (1-v)^{N-n} \, v^{n-1} \, dv =: B(N, \epsilon, n) \, .$$

These authors also show that

$$B(N, \epsilon, n) = \sum_{j=0}^{n} \binom{N}{j} \epsilon^j (1-\epsilon)^{N-j} \; .$$

For any probability level $\epsilon \in (0, 1)$ and confidence level $\beta \in (0, 1)$, let

$$N(\epsilon, \beta) := \min \left\{ N \in \mathbb{N} : \sum_{j=0}^{n} \binom{N}{j} \epsilon^j (1-\epsilon)^{N-j} \leq \beta \right\} \; .$$

Then $N(\epsilon, \beta)$ is a sample size which guarantees that the $\epsilon$-violation probability lies below $\beta$.

The CCG Theorem can also be applied to the problem

$$\min_x \sup_{\xi \in \Xi} \left\{ g(x,\xi) : x \in \mathbb{X}(\xi) \right\} , \tag{3}$$

where $x \mapsto g(x,\xi)$ is convex and $\mathbb{X}(\xi)$ are convex sets for all $\xi \in \Xi$. Set

$$f(x,\xi) = g(x,\xi) + \psi_{\mathbb{X}(\xi)}(x) ,$$

where $\psi$ is the indicator function

$$\psi_B(x) := \begin{cases} 0 & \text{if } x \in B \\ \infty & \text{otherwise.} \end{cases}$$

Then $f$ is convex in $x$ and (3) can be written as

$$\min_x \sup_{\xi \in \Xi} f(x,\xi) .$$

Finally, observe that this problem is equivalent to

$$\min_{x,\gamma} \left\{ \gamma : \sup_{\xi \in \Xi} f(x,\xi) - \gamma \leq 0 \right\} .$$

This problem is of the standard form. In this case, the dimension of the decision variable is $n+1$.

# An Example illustrating the violation probability

The original problem:

$$\left\| \begin{array}{ll} \text{Maximize} & x \\ \text{subject to} & x^2 + y^2 \le 1 \end{array} \right.$$

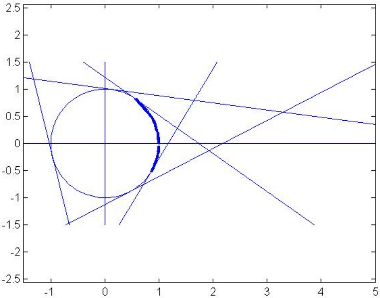
The reformulation as a problem with an infinite number of linear constraints:

$$\left\| \begin{array}{lll} \text{Maximize} & x & \\ \text{subject to} & x\cos(\xi) + y\sin(\xi) \le 1 & \text{for all } 0 \le \xi \le 2\pi \end{array} \right.$$

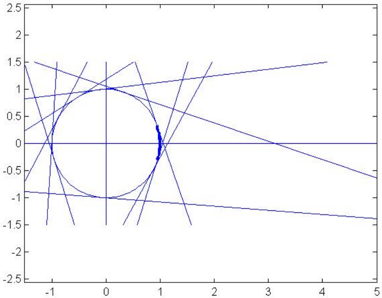
The randomly sampled problem, $\xi^{(i)} \sim Uniform[0, 2\pi]$ :

$$\left\| \begin{array}{lll} \text{Maximize} & x & \\ \text{subject to} & x\cos(\xi^{(i)}) + y\sin(\xi^{(i)}) \le 1 & \text{for } i = 1, \ldots, N \end{array} \right.$$

# Illustration




$N = 5$




$N = 10$

The random violation probability is represented by the blue arc length (relative to the total circumference $2\pi$).

# Extending the notion of violation probability to the multistage case

In the multistage situation, we have to respect the non-anticipativity conditions. Based on a finite random selection $\hat{\Xi}_1^{N_1}, \ldots, \hat{\Xi}_H^{N_H}$

$$
\begin{aligned}
\hat{\Xi}_1^{N_1} &= \{\xi_1^{(1)}, \ldots, \xi_1^{(N_1)}\} , \\
\hat{\Xi}_2^{N_2} &= \{\xi_2^{(1)}, \ldots, \xi_2^{(N_2)}\} , \\
&\vdots \\
\hat{\Xi}_H^{N_H} &= \{\xi_H^{(1)}, \ldots, \xi_H^{(N_H)}\} ,
\end{aligned}
$$

we generate a random tree $\hat{\mathcal{T}}^{N_1,\ldots,N_H}$, where $\{\xi_1^{(1)}, \ldots, \xi_1^{(N_1)}\}$ are the successors of the root, and recursively all nodes at stage $t$ get all values from $\Xi_t^{N_t}$ as successors. Notice that the number of nodes at stage $t+1$ of the tree is $\bar{N}_t := \prod_{s=1}^{t} N_s$. The total number of nodes of the tree is $N_{\mathrm{tot}} := 1 + \sum_{i=1}^{H} \bar{N}_i$.

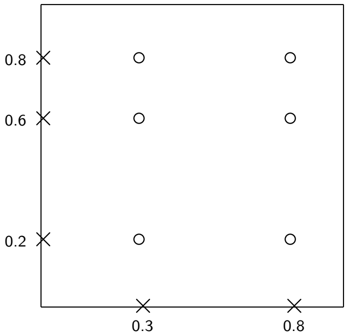
# The violation probability in the multistage situation

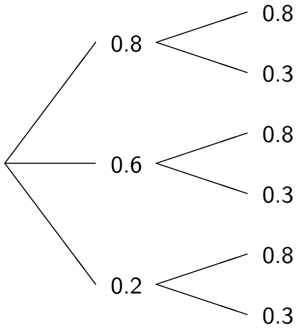
The violation probability $V_t$ at stage $t$ is defined in the following way. Given the random tree $\hat{\mathcal{T}}^{N_1,\ldots,N_H}$, suppose that we sample an additional element $\xi_t^{(N_{t+1})}$ in $\Xi_t$ and form the extended tree $\hat{\mathcal{T}}^{N_1,\ldots,N_t+1,\ldots,N_H}$. Then

$$V_t(\hat{\mathcal{T}}^{N_1,\ldots,N_H}) = \mathbb{P}\{\xi_t^{(N_t+1)} : v(\hat{\mathcal{T}}^{N_1,\ldots,N_t+1,\ldots,N_H}) > v(\hat{\mathcal{T}}^{N_1,\ldots,N_H})\}.$$

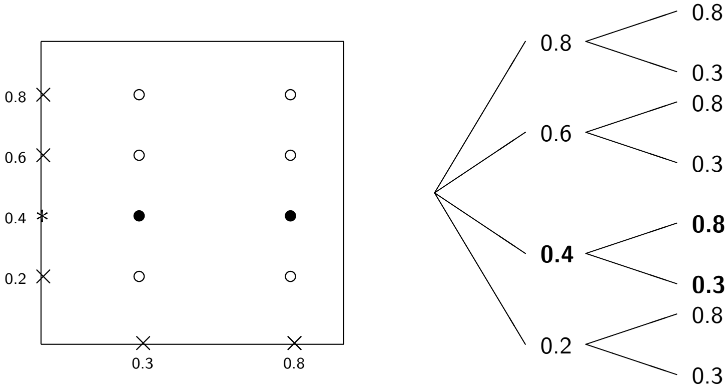
Here $v(\mathcal{T})$ is the value of the multistage optimization problem on the tree $\mathcal{T}$.

# Illustration

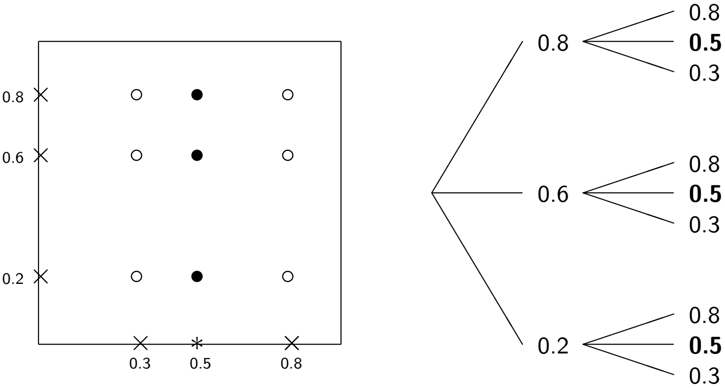



The original sampled tree $\hat{\mathcal{T}}^{3,2}$.




The randomly extended tree $\hat{\mathcal{T}}^{4,2}$. A new observation in stage 1 is added. The new nodes are in bold.




The randomly extended tree $\hat{\mathcal{T}}^{3,3}$. A new observation in stage 2 is added. The new nodes are in bold.

# The structure of the multistage robust decision problem

The decision problem can be written as

$$\widehat{\mathrm{RO}}_{H+1}^{N_1\cdots N_H}: \min_{(x_{1,\cdot})\in I_1} \max_i \min_{(x_{2,\cdot})\in I_2} \max_i \min_{(x_{3,\cdot})\in I_3} \ldots \max_i f_i(x_{1,i},\ldots,x_{H,i},x_{H+1,i})$$

where $I_j$ are the constraint sets induced by non-anticipativity and the functions $f_i$ are defined as

$$f_i(x_{1,i},\ldots,x_{H+1,i}) = c_1^\top x_{1,i} + \psi_{\mathbb{X}_1}(x_{1,i}) + \sum_{t=2}^{H+1}(c_{t,i}^\top x_{t,i} + \psi_{\mathbb{X}_t(x_{t-1,i},\xi_{p_t(i)})}(x_{t,i})) ,$$

where $\psi.$ are the convex indicator functions and

$$\mathbb{X}_t(x_{t-1},\xi_{t-1}) := \{x_t \geq 0 : T_{t-1}(\xi_{t-1})x_{t-1} + W_t(\xi_{t-1})x_t = h_t(\xi_{t-1})\}.$$

# Upper and lower bounds

From the representation

$$\widehat{\mathrm{RO}}_{H+1}^{N_1 \cdots N_H} : \min_{(x_{1,\cdot}) \in I_1} \max_i \min_{(x_{2,\cdot}) \in I_2} \max_i \min_{(x_{3,\cdot}) \in I_3} \ldots \max_i f_i(x_{1,i}, \ldots, x_{H,i}, x_{H+1,i}) \tag{4}$$

one sees that

(i) A lower bound is obtained by relaxing the non-anticipativity constraints

(ii) An upper bound is obtained by shifting some or all max operators to the right in formula (4)

# The violation probability in stage 1

For the violation probability at stage 1, keep the samples $\hat{\Xi}_2, \ldots, \hat{\Xi}_H$ fixed and consider only at the dependency on $\xi_1$, summarized in the objective function $\bar{f}(x_1, \xi_1)$. The decision problem at stage 1 is of the form

$$\min_{x_1} \max_{\xi_1} \bar{f}(x_1, \xi_1) \ .$$

Therefore we get the estimate from the CCG Theorem

$$\mathbb{P}\left\{ V_1(\hat{\Xi}_1^{N_1}, \ldots, \hat{\Xi}_H^{N_H}) > \epsilon \right\} \leq B(N_1, \epsilon, n_1 + 1) \ ,$$

where $n_1 = dim(x_1)$.

# The violation probability at stage $t$

Similarly, at stage $t$, there are $\bar{N}_{t-1} = \prod_{s=1}^{t-1} N_s$ nodes of the tree. The violation probability $V_{t,j}$ at stage $t$ and a fixed node $j$ is stochastically dominated by $Y_{N_t,n_t+1}$, given before, i.e.

$$\mathbb{P}\left\{ V_{t,j}(\hat{\Xi}_1^{N_1}, \ldots, \hat{\Xi}_H^{N_H}) > \epsilon \right\} \leq \mathbb{P}\left\{ Y_{N_t,n_t+1} > \epsilon \right\}.$$

Notice that this bound does not depend on $j$. Now

$$\begin{aligned} V_t(\hat{\Xi}_1^{N_1}, \ldots, \hat{\Xi}_H^{N_H}) &= \mathbb{P}\left\{ \text{Violation at any node at stage } t | \hat{\Xi}_1^{N_1}, \ldots, \hat{\Xi}_H^{N_H} \right\} \\ &\leq \sum_{j=1}^{\bar{N}_{t-1}} \mathbb{P}\left\{ \text{Violation at node } j \text{ at stage } t \, | \hat{\Xi}_1^{N_1}, \ldots, \hat{\Xi}_H^{N_H} \right. \\ &= \sum_{j=1}^{\bar{N}_{t-1}} V_{t,j}(\hat{\Xi}_1^{N_1}, \ldots, \hat{\Xi}_H^{N_H}) \, . \end{aligned}$$

Now, we use a modification of a result by Frank, Nelson and Schweizer (1987) solving a problem by Kolmogorov/Makarov.

**Lemma.** Let $X_1, \ldots, X_K$ be a sequence of possibly dependent random variables, where each of them has left-continuous distribution function $F$ and survival function $\bar{F} = 1 - F$. Then

$$\mathbb{P}\{\sum_{j=1}^{K} X_i \geq \epsilon\} \leq K\bar{F}(\epsilon/K),$$

leading to the main Theorem

**Theorem.** [Violation probability at stage $t$ of sampled scenario tree]

Given an accuracy level $\epsilon \in (0,1)$, let $\bar{N}_{t-1} = \prod_{s=1}^{t-1} N_s$ and $\epsilon_t := \epsilon/\bar{N}_{t-1}$. Then, the probability of violation at stage $t$, $V_t(\hat{\Xi}^{N_1}, \ldots, \hat{\Xi}^{N_H})$ is bounded by

$$\mathbb{P}\left\{V_t(\hat{\Xi}^{N_1}, \ldots, \hat{\Xi}^{N_H}) > \epsilon\right\} \leq \bar{N}_{t-1} B(N_t, \epsilon_t, n_t + 1),$$

where $n_t = dim(x_t)$.

# Almost sure convergence

**Theorem.** If all sample sizes $N_1, \ldots, N_H$ tend to infinity, then the optimal value of the sampled multistage program converges almost surely to the optimal value of the robust multistage program.

# A numerical example: An inventory problem

- A retailer replenishes his inventory at the beginning of each time period $t \in \{1, \ldots, H\}$ by orders $x_t^o$ - but without knowing the demand $\xi_t$ - at a cost of $d_t$ per unit of the product.
- The demand must be satisfied from the inventory with filling level $s_t^{inv}$. Unsatisfied demand may be backlogged at cost $p_t$ and inventory may be held in the warehouse with a unitary holding cost $h_t$.
- Lower and upper bounds on the orders $x_t^o$ at each period as well as on the cumulative orders are given. We assume that there is no demand at time $t = 1$ and that the demand at time $t$ lies within an interval centered around a nominal value $\bar{\xi}_t$ and uncertainty level $\rho \in [0, 1]$ resulting in a box uncertainty set as follows: $\Xi = \times_{t \in \mathbb{T}} \left\{ \xi_t \in \mathbb{R} : \left| \xi_t - \bar{\xi}_t \right| \leq \rho \bar{\xi}_t \right\}$.

The full model is

$$\mathrm{RO}_{H+1}(\mathcal{COC}) := \tag{5a}$$

$$\min_{x_t^o, x_t^c, s_t^{co}, s_t^{inv}} \left[ x_1^c + \max_{\underline{\xi} \in \Xi} \sum_{t \in \mathbb{T}} x_{t+1}^c(\underline{\xi}_t) \right] \tag{5b}$$

$$\text{s.t. } x_1^c \geq d_1 x_1^o + \max\left\{ h_1 s_1^{inv}, -p_1 s_1^{inv} \right\} \tag{5c}$$

$$x_{t+1}^c(\underline{\xi}_t) \geq d_{t+1} x_{t+1}^o(\underline{\xi}_t) + + \max\left\{ h_{t+1} s_{t+1}^{inv}(\underline{\xi}_t), -p_{t+1} s_{t+1}^{inv}(\underline{\xi}_t) \right\}, \quad t = 1, \ldots, H-1 \tag{5d}$$

$$x_{H+1}^c(\underline{\xi}_H) \geq \max\left\{ h_{H+1} s_{H+1}^{inv}(\underline{\xi}_H), -p_{H+1} s_{H+1}^{inv}(\underline{\xi}_H) \right\} \tag{5e}$$

$$s_2^{inv}(\underline{\xi}_1) = s_1^{inv} + x_1^o - \xi_1 \tag{5f}$$

$$s_{t+1}^{inv}(\underline{\xi}_t) = s_t^{inv}(\underline{\xi}_{t-1}) + x_t^o(\underline{\xi}_{t-1}) - \xi_t, \quad t = 2, \ldots, H \tag{5g}$$

$$s_2^{co}(\underline{\xi}_1) = s_1^{co} + x_1^o \tag{5h}$$

$$s_{t+1}^{co}(\underline{\xi}_t) = s_t^{co}(\underline{\xi}_{t-1}) + x_t^o(\underline{\xi}_{t-1}), \quad t = 2, \ldots, H \tag{5i}$$

$$\underline{x}_1^o \leq x_1^o \leq \bar{x}_1^o, \quad \underline{s}_1^{co} \leq s_1^{co} \leq \bar{s}_1^{co} \tag{5j}$$

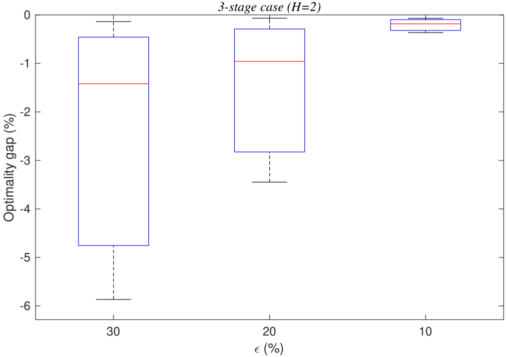
$$\underline{x}_t^o \leq x_t^o(\underline{\xi}_{t-1}) \leq \bar{x}_t^o, \quad \underline{s}_t^{co} \leq s_t^{co}(\underline{\xi}_{t-1}) \leq \bar{s}_t^{co}, \quad t = 2, \ldots, H \tag{5k}$$

# The required sample sizes

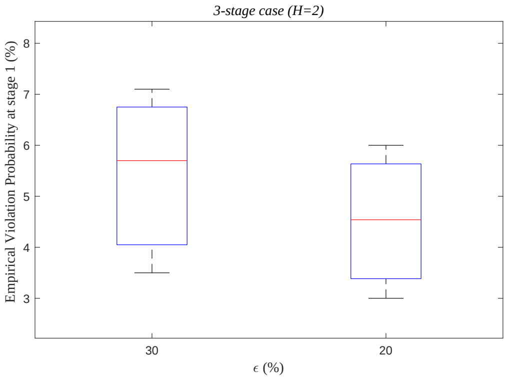
| $\epsilon$ (%) | $\beta$ | $N_1^*$ | $N_2^*$ |
|---|---|---|---|
| 30 | 0.05 | 43 | 1849 |
| 20 | 0.05 | 64 | 4096 |
| 10 | 0.05 | 127 | 16129 |

Sample sizes of the problem $\widehat{RO}_3^{N_1 N_2}$ in the tree-stage case ($H = 2$) with $n_1 = n_2 = 4$.

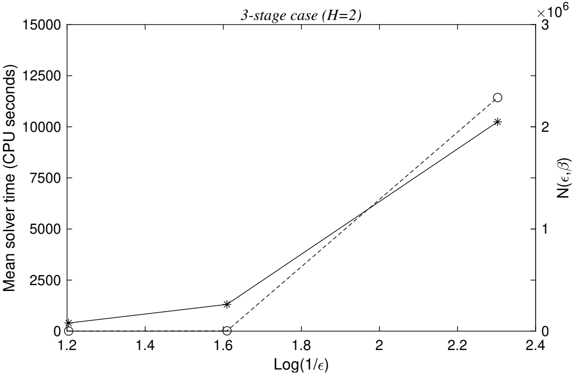
# The optimality gaps



# The empirical violation probabilites



# Mean solver times




Mean solver times (solid lines) and number of samples (dashed line) as a function of $\log(1/\epsilon)$.

# Conclusions

- Multistage robust optimization problems can be approximated by sampled versions. Almost sure convergence holds.
- We found bounds for the multistage violation probabilities in the sense of Calafiore and Campi/Garatti.
- The empirical violation probabilities are typically much smaller than the universal (worst case) upper bounds. There is much room for tightening these bounds.

References:


- F. Dabbene, F. Maggioni, G. Pflug. Multistage robust convex optimization problems: A sampling based approach. submitted.
- G. Calafiore. Random convex programs (2010). SIAM J. Optimization 20(6) 3427-3464
- M.C. Campi, S. Garatti. The exact feasibility of randomized solutions of uncertain convex programs (2008). SIAM J. Optimization 19 (3) 1211-1230.
- P. Vayanos, D. Kuhn, B. Rustem (2012). A constraint sampling approach for multi-stage robust optimization. Automatica 48 (3) 459-471.