Monte Carlo Sampling Approach to Solving Stochastic Multistage Programs

A. Shapiro
School of Industrial and Systems Engineering,
Georgia Institute of Technology,
Atlanta, Georgia 30332-0205, USA

Reinforcement Learning from Batch Data and Simulation
December 2020

Multistage stochastic programming.

Let ξ_t be a random (stochastic) process. Denote $\xi_{[t]} := (\xi_1, .., \xi_t)$ the history of the process ξ_t up to time t. The values of the decision vector x_t, chosen at stage t, may depend on the information $\xi_{[t]}$ available up to time t, but not on the future observations. The decision process has the form

$$\text{decision}(x_0) \rightsquigarrow \text{observation}(\xi_1) \rightsquigarrow \text{decision}(x_1) \rightsquigarrow \\ ... \rightsquigarrow \text{observation}(\xi_T) \rightsquigarrow \text{decision}(x_T).$$

Risk neutral T-stage stochastic programming problem:

$$\begin{array}{cl} \min\limits_{x_1, x_2(\cdot), \ldots, x_T(\cdot)} & f_1(x_1) + \mathbb{E}\left[\sum_{t=2}^{T} f_t\left(x_t, \xi_t\right)\right] \\ \text{s.t.} & x_1 \in \mathcal{X}_1,\ x_t \in \mathcal{X}_t(x_{t-1}, \xi_t),\ t = 2, \ldots, T. \end{array}$$

In linear case, $f_t(x_t, \xi_t) := c_t^\top x_t$ and

$$\mathcal{X}_t(x_{t-1}, \xi_t) := \{x_t : B_t x_{t-1} + A_t x_t = b_t,\ x_t \geq 0\},\ \ t = 2, ..., T.$$

Optimization is performed over feasible policies (also called decision rules). A policy is a sequence of (measurable) functions $x_t = x_t(\xi_{[t]})$, $t = 1, ..., T$. Each $x_t(\xi_{[t]})$ is a function of the data process up to time t, this ensures the *nonanticipative* property of a considered policy.

If the number of realizations (scenarios) of the process ξ_t is finite, then the above (linear) problem can be written as one large (linear) programming problem.

If we measure computational complexity, of the "true" problem, in terms of the number of scenarios required to approximate true distribution of the random data process with a reasonable accuracy, the conclusion is rather pessimistic.

Distributionally robust approach

Static case. The problem is formulated in the following minimax form

$$\min_{x \in \mathcal{X}} \sup_{P \in \mathfrak{M}} \mathbb{E}_P[F(x, \omega)],$$

where $\mathfrak{M}$ is a specified set of probability measures (distributions) on a sample space $(\Omega, \mathcal{F})$, and $F : \mathcal{X} \times \Omega \to \mathbb{R}$ is an objective function. It is assumed that for every $x \in \mathcal{X}$ the random variable $F_x(\omega) = F(x, \omega)$ is $\mathcal{F}$-measurable and the expectation

$$\mathbb{E}_P[F_x] = \int_\Omega F_x(\omega) dP(\omega),$$

with respect to every $P \in \mathfrak{M}$, is well defined and finite valued.

Popular approaches to define the set $\mathfrak{M}$ are either: (i) by distributions in some sense close to a specified reference distribution $\mathbb{P}$, or (ii) by moment constraints.

With the set $\mathfrak{M}$ is associated the (worst-distribution) functional

$$\mathcal{R}(Z) := \sup_{P \in \mathfrak{M}} \mathbb{E}_P[Z]$$

defined on a linear space $\mathcal{Z}$ of $\mathcal{F}$-measurable variables $Z : \Omega \to \mathbb{R}$.

There are two, somewhat natural, frameworks for duality analysis of this risk functional. In case (i), the set $\mathfrak{M}$ is assumed to consist of probability measures absolutely continuous with respect to $\mathbb{P}$, and

$$\mathfrak{M} = \{P : dP/d\mathbb{P} \in \mathfrak{A}\},$$

where $\mathfrak{A}$ is a set of densities.

In that case

$$\mathcal{R}(Z) = \sup_{\zeta \in \mathfrak{A}} \int_{\Omega} Z(\omega)\zeta(\omega) d\mathbb{P}(\omega).$$

Cumulative distribution function of a random variable $Z(\omega)$ (with respect to $\mathbb{P}$) is $F_Z(t) = \mathbb{P}(Z \leq t)$. Two random variables Z, Z' are distributionally equivalent if $F_Z = F_{Z'}$, i.e., $\mathbb{P}(Z \leq t) = \mathbb{P}(Z' \leq t)$ for all $t \in \mathbb{R}$.

Definition 1 *It is said that a risk measure $\mathcal{R} : \mathcal{Z} \to \overline{\mathbb{R}}$ is law invariant, with respect to the reference distribution $\mathbb{P}$, if for any distributionally equivalent $Z, Z' \in \mathcal{Z}$, it follows that $\mathcal{R}(Z) = \mathcal{R}(Z')$.*

That is, law invariant risk measure $\mathcal{R}(Z)$ is a function of the cdf $F = F_Z$. We sometimes write $\mathcal{R}(F)$ directly as a function of cdf F. Value $\mathcal{R}(F)$ can be estimated by $\mathcal{R}(\widehat{F}_N)$, where $\widehat{F}_N$ is an empirical estimate of the cdf F based on a sample of size N. Consequently solving the distributionally robust problem can be approached by the Sample Average Approximation (SAA) method.

How law invariance of $\mathcal{R}$ can be formulated in terms of the uncertainty set $\mathfrak{M}$? Let $\mathcal{Z} := L_p(\Omega, \mathcal{F}, \mathbb{P})$ and

$$\mathcal{R}(Z) = \sup_{\zeta \in \mathfrak{A}} \int_{\Omega} \zeta(\omega) Z(\omega) d\mathbb{P}(\omega),$$

where $\mathfrak{A} \subset \mathcal{Z}^* = L_q(\Omega, \mathcal{F}, \mathbb{P})$ is a set of density functions.

It is said that the uncertainty set $\mathfrak{A}$ is law invariant if $\zeta \in \mathfrak{A}$ and ζ' is distributionally equivalent to ζ implies that $\zeta' \in \mathfrak{A}$.

Theorem 1 *(i) If the uncertainty set $\mathfrak{A}$ is law invariant, then the corresponding functional $\mathcal{R}$ is law invariant. (ii) Conversely, if the functional $\mathcal{R}$ is law invariant and the set $\mathfrak{A}$ is convex and weakly* closed, then $\mathfrak{A}$ is law invariant.*

In case the functional $\mathcal{R}$ is law invariant, it can be considered as a function of the cdf F_Z. Given a random sample $Z_1, ..., Z_N \sim \mathbb{P}$, we can approximate F_Z by the empirical cdf

$$\widehat{F}_N(z) := \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{(-\infty, z]}(Z_i).$$

Consequently we can approximate $\mathcal{R}(Z) = \mathcal{R}(F_Z)$ by $\mathcal{R}(\widehat{F}_N)$.

Suppose now that $\xi_1, ..., \xi_N$ is a sample of the random vector $\xi = \xi(\omega)$. Then we can estimate distributionally robust problem by the SAA problem:

$$\operatorname*{Min}_{x \in \mathcal{X}} \mathcal{R}(\widehat{F}_{x,N}). \tag{1}$$

Example

Consider $\mathcal{Z} := L_1(\Omega, \mathcal{F}, \mathbb{P})$ and

$$\mathfrak{A} := \left\{ \zeta : 1 - \beta_1 \leq \zeta(\omega) \leq 1 + \beta_2,\ \omega \in \Omega,\ \int_\Omega \zeta d\mathbb{P} = 1 \right\},$$

where $\beta_1 \in (0, 1]$ and $\beta_2 \geq 0$. The corresponding functional $\mathcal{R}$ is

$$\mathcal{R}(Z) = (1 - \beta_1)\mathbb{E}_{\mathbb{P}}[Z] + \beta_1 \mathsf{AV@R}_\alpha(Z),$$

where $\alpha = \beta_1/(\beta_1 + \beta_2)$ and

$$\mathsf{AV@R}_\alpha(Z) = \frac{1}{\alpha} \int_{1-\alpha}^{1} F_Z^{-1}(t) dt = \inf_{t \in \mathbb{R}} \left\{ t + \alpha^{-1} \mathbb{E}_P[Z - t]_+ \right\}.$$

Multistage setting

For a family of probability distributions P of the data process $(\xi_1, ..., \xi_T)$ it is tempting to write the distributionally robust analogue of the risk neutral problem as

$$\min_{\pi \in \Pi} \sup_{P \in \mathfrak{M}} \mathbb{E}_P \left[\sum_{t=1}^{T} f_t(x_t, \xi_t) \right]$$

with Π being the set of policies satisfying the feasibility constraints.

However this formulation does not explicitly specify dynamics of the considered problem. Even worse, at the moment it is not well defined since it is not clear what "feasibility for a.e. realization of the data process" means.

Consider the nested functional (recall that ξ_1 is deterministic)

$$\mathfrak{R}(Z) := \sup_{P \in \mathfrak{M}} \mathbb{E}_{P|\xi_1} \left[\text{ess sup}_{P \in \mathfrak{M}} \mathbb{E}_{P|\xi_2} \left[\cdots \text{ess sup}_{P \in \mathfrak{M}} \mathbb{E}_{P|\xi_{[T-1]}}[Z] \right] \right],$$

where $\xi_{[t]} = (\xi_1, ..., \xi_t)$. This functional can be represented in the dual form

$$\mathfrak{R}(Z) = \sup_{P \in \widehat{\mathfrak{M}}} \mathbb{E}_P[Z]$$

for some set $\widehat{\mathfrak{M}}$ of probability measures.

Note that $\mathfrak{M} \neq \widehat{\mathfrak{M}}$.

This leads to the *nested* formulation of the distributionally robust problem

$$\min_{\pi \in \Pi} f_1(x_1) + \mathfrak{R}\Big[\sum_{t=2}^{T} f_t(x_t^{\pi}, \xi_t) \Big],$$

where Π is the set of feasible policies.

For the nested formulation it is possible to write dynamic programming equations with the respective cost-to-go (value) functions $V_t(x_{t-1}, \xi_{[t]})$ given by the optimal value of the problem

$$\begin{array}{ll} \min\limits_{x_t \in \mathcal{X}_t} & f_t(x_t, \xi_t) + \text{ess sup}\limits_{P \in \mathfrak{M}} \mathbb{E}_{P|\xi_{[t]}}[V_{t+1}(x_t, \xi_{[t+1]})] \\ \text{s.t.} & B_t x_{t-1} + A_t x_t = b_t \end{array}$$

at stages $t = T, T-1, ..., 1$, and $V_{T+1}(\cdot, \cdot)$ omitted.

The rectangular case:

$$\mathfrak{M} = \{P = P_1 \times \cdots \times P_T : P_t \in \mathfrak{M}_t,\ t = 1, ..., T\}$$

where $\mathfrak{M}_t$ is a set of marginal distributions of random vector ξ_t. In the risk neutral setting, when $\mathfrak{M}_t$ are singletons, this corresponds to the stagewise independence condition.

In the rectangular case the dynamic equations simplify to

$$\begin{array}{ll} \min\limits_{x_t \in \mathcal{X}_t} & f_t(x_t, \xi_t) + \sup\limits_{Q_{t+1} \in \mathfrak{M}_{t+1}} \mathbb{E}_{Q_{t+1}}[V_{t+1}(x_t, \xi_{t+1})] \\ \text{s.t.} & B_t x_{t-1} + A_t x_t = b_t. \end{array}$$

Approximate dynamic programming

Basic idea is to approximate the value functions

$$\mathcal{V}_{t+1}(x_t) = \sup_{Q_{t+1} \in \mathfrak{M}_{t+1}} \mathbb{E}_{Q_{t+1}}[V_{t+1}, \xi_{t+1})]$$

by a class of computationally manageable functions. When functions $\mathcal{V}_t(\cdot)$ are convex it is natural to approximate these functions by piecewise linear functions given by maximum of cutting hyperplanes.

Cutting planes type approach. In the risk neutral setting this was introduced in Pereira and Pinto (1991), based on the nested method of Birge (1985). This became known as the Stochastic Dual Dynamic Programming (SDDP) method.

Consider the linear multistage program and the rectangular setting. First, the marginal distributions of ξ_t, $t = 2, ..., T$, are discretized by generating Monte Carlo samples from the reference distribution (the SAA approach).

For the constructed discretized problem the value functions are approximated. For trial decisions $\bar{x}_t$, $t = 1, ..., T-1$, at the backward step of the SDDP algorithm, piecewise linear approximations $\mathfrak{V}_t(\cdot)$ of the value functions $\mathcal{V}_t(\cdot)$ are constructed by solving problems

$$\operatorname*{Min}_{x_t \in \mathbb{R}^{n_t}} (c_t^j)^{\mathsf{T}} x_t + \mathfrak{V}_{t+1}(x_t) \text{ s.t. } B_t^j \bar{x}_{t-1} + A_t^j x_t = b_t^j,\ x_t \geq 0,$$

$j = 1, ..., N_t$, and their duals, going backward in time $t = T, ..., 1$.

By construction

$$\mathcal{V}_t(\cdot) \geq \mathfrak{V}_t(\cdot),\ t = 2, ..., T.$$

Therefore the optimal value of

$$\operatorname*{Min}_{x_1 \in \mathbb{R}^{n_1}} c_1^\top x_1 + \mathfrak{V}_2(x_1) \text{ s.t. } A_1 x_1 = b_1,\ x_1 \geq 0$$

gives a lower bound for the optimal value $\widehat{v}_N$ of the SAA problem.

In the *risk neutral* setting,

$$v^0 \geq \mathbb{E}[\widehat{v}_N],$$

and hence *on average* $\widehat{v}_N$ is also a lower bound for the optimal value v^0 of the *true* problem.

The approximate value functions $\mathfrak{V}_2, ..., \mathfrak{V}_T$ and a feasible first stage solution $\bar{x}_1$ define a feasible policy. That is for a realization (sample path) $\xi_1, ..., \xi_T$ of the data process, $\bar{x}_t = \bar{x}_t(\xi_{[t]})$ are computed recursively in $t = 2, ..., T$ as a solution of

$$\underset{x_t \geq 0}{\text{Min}} \, c_t^{\mathsf{T}} x_t + \mathfrak{V}_{t+1}(x_t) \text{ s.t. } B_t \bar{x}_{t-1} + A_t x_t = b_t.$$

In the *forward step* of the SDDP algorithm M sample paths (scenarios) are generated and the corresponding $\bar{x}_t$, $t = 2, ..., T$, are used as trial points in the next iteration of the backward step.

Note that the functions $\mathfrak{V}_2, ..., \mathfrak{V}_T$ and $\bar{x}_1$ define a feasible policy also for the *true* problem.

Periodical infinite horizon multistage programs

Consider infinite horizon problem with discount factor $\gamma \in (0,1)$

$$\min_{\pi \in \Pi} \; f_1(x_1) + \mathfrak{R}\left[\sum_{t=2}^{\infty} \gamma^{t-1} f_t\left(x_t, \xi_t\right)\right],$$

where Π is a set of policies satisfying the feasibility constraints

$$x_t \in \mathcal{X}_t, \; B_t x_{t-1} + A_t x_t = b_t.$$

Suppose the **rectangular** setting, and that the problem has periodic structure with period $m \in \mathbb{N}$.

That is

- The functional $\mathcal{R} : \mathcal{Z} \to \mathbb{R}$ is a law invariant.

- Vectors ξ_t and $\xi_{t+\mathsf{m}}$ have the same (reference) distribution, with support $\Xi \subset \mathbb{R}^d$, for $t \geq 2$ (recall that ξ_1 is deterministic).

- The functions $b_t(\cdot)$, $B_t(\cdot)$, $A_t(\cdot)$ and $f_t(\cdot, \cdot)$ have period m, i.e., are the same for $t = \tau$ and $t = \tau + \mathsf{m}$, $t = 2, ...,$ and the sets $\mathcal{X}_t$ are nonempty and $\mathcal{X}_t = \mathcal{X}_{t+\mathsf{m}}$ for all t.

This leads to the following periodical variant of Bellman equations for the value functions $\mathcal{V}_2(\cdot), ..., \mathcal{V}_{\mathsf{m}+1}(\cdot)$ of the dynamic equations:

$$\mathcal{V}_\tau(x_{\tau-1}) = \mathcal{R}[V_\tau(x_{\tau-1}, \xi_\tau)].$$

$$V_\tau(x_{\tau-1}, \xi_\tau) = \inf_{\substack{x_\tau \in \mathcal{X}_\tau \\ B_\tau x_{\tau-1} + A_\tau x_\tau = b_\tau}} f_\tau(x_\tau, \xi_\tau) + \gamma \mathcal{V}_{\tau+1}(x_\tau),$$

for $\tau = 2, ..., \mathsf{m}+1$, and $\mathcal{V}_{\mathsf{m}+2}$ replaced by $\mathcal{V}_2$ for $\tau = \mathsf{m}+1$. Consequently for $t \geq \mathsf{m}+2$ the corresponding value functions are defined recursively as $V_t(\cdot, \xi_t) = V_{t-\mathsf{m}}(\cdot, \xi_t)$, and hence $\mathcal{V}_t(\cdot) = \mathcal{V}_{t-\mathsf{m}}(\cdot)$.

Duals of periodical linear programs

Dual approach to construction of upper bounds was initiated in Leclére, Carpentier, Chancelier, Lenoir, Pacaud (2019),
we follow here the approach of Guigues, Shapiro, Cheng (2019)

Consider linear (risk neutral) multistage stochastic program

$$\begin{array}{ll} \min\limits_{x_t \geq 0} & \mathbb{E}\left[\sum_{t=1}^T c_t^\top x_t\right] \\ \text{s.t.} & A_1 x_1 = b_1, \\ & B_t x_{t-1} + A_t x_t = b_t,\ t = 2, ..., T. \end{array}$$

Dualization of the the feasibility constraints

The Lagrangian

$$L(x,\pi) = \mathbb{E}\left[\sum_{t=1}^{T} c_t^\top x_t + \pi_t^\top (b_t - B_t x_{t-1} - A_t x_t)\right]$$

in variables $x = (x_1(\xi_{[1]}), \ldots, x_T(\xi_{[T]}))$ and $\pi = (\pi_1(\xi_{[1]}), \ldots, \pi_T(\xi_{[T]}))$ with the convention that $x_0 = 0$. Dualization of the feasibility constraints leads to the following dual

$$\begin{array}{ll} \max\limits_{\pi} & \mathbb{E}\Big[\sum\limits_{t=1}^{T} b_t^\top \pi_t\Big] \\ \text{s.t.} & A_T^\top \pi_T \le c_T, \\ & A_{t-1}^\top \pi_{t-1} + \mathbb{E}_{|\xi_{[t-1]}}\left[B_t^\top \pi_t\right] \le c_{t-1},\ t = 2, ..., T. \end{array}$$

The optimization is over policies $\pi_t = \pi_t(\xi_{[t]})$, $t = 1, ..., T$.

Dynamic programming equations for the dual problem

Assume the stagewise independence condition and finite number of scenarios N_t per stage and respective probabilities p_{tj}. At the last stage $t = T$ we have the following problem

$$\begin{aligned} \max_{\pi_{T1},...,\pi_{TN_T}} \quad & \mathbb{E}[b_T^\top \pi_T] = \sum_{j=1}^{N_T} p_{Tj} b_{Tj}^\top \pi_{Tj} \\ \text{s.t.} \quad & A_{Tj}^\top \pi_{Tj} \leq c_{Tj},\ j = 1, ..., N_T, \\ & A_{T-1}^\top \pi_{T-1} + \sum_{j=1}^{N_T} p_{Tj} B_{Tj}^\top \pi_{Tj} \leq c_{T-1}. \end{aligned}$$

The optimal value $V_T(\pi_{T-1}, \xi_{T-1})$ and an optimal solution $(\bar{\pi}_{T1}, \ldots, \bar{\pi}_{TN_T})$ of that problem are functions of vectors π_{T-1} and c_{T-1} and matrix A_{T-1}.

And so on going backward in time we can write the respective dynamic programming equations for $t = T - 1, ..., 2$, as

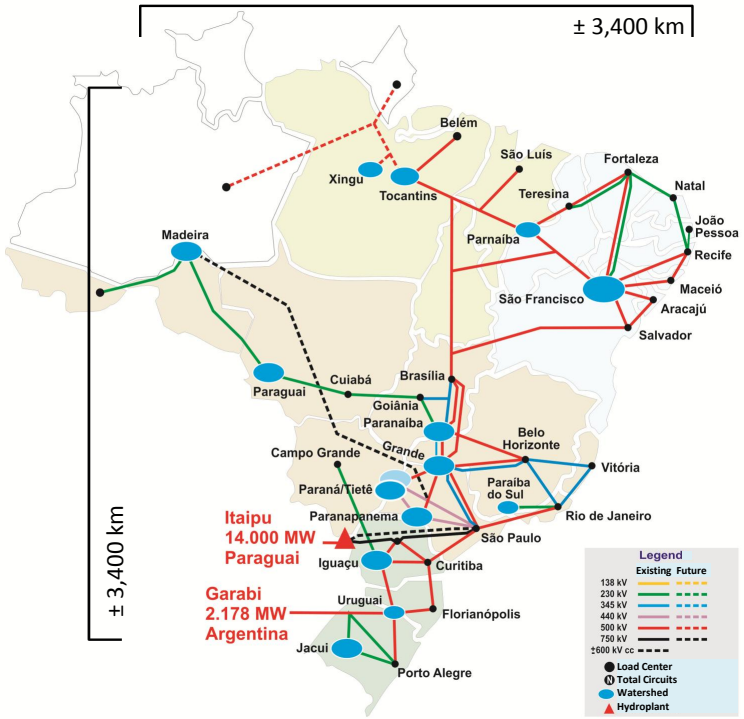
$$\begin{aligned} \max_{\pi_{t1},\ldots,\pi_{tN_t}} \quad & \sum_{j=1}^{N_t} p_{tj} \left[b_{tj}^\top \pi_{tj} + V_{t+1}(\pi_{tj}, \xi_{tj}) \right] \\ \text{s.t.} \quad & A_{t-1}^\top \pi_{t-1} + \sum_{j=1}^{N_t} p_{tj} B_{tj}^\top \pi_{tj} \leq c_{t-1}, \end{aligned}$$

with $V_t(\pi_{t-1}, \xi_{t-1})$ being the optimal value of the above problem. Finally at the first stage the following problem should be solved

$$\max_{\pi_1} b_1^\top \pi_1 + V_2(\pi_1, \xi_1).$$

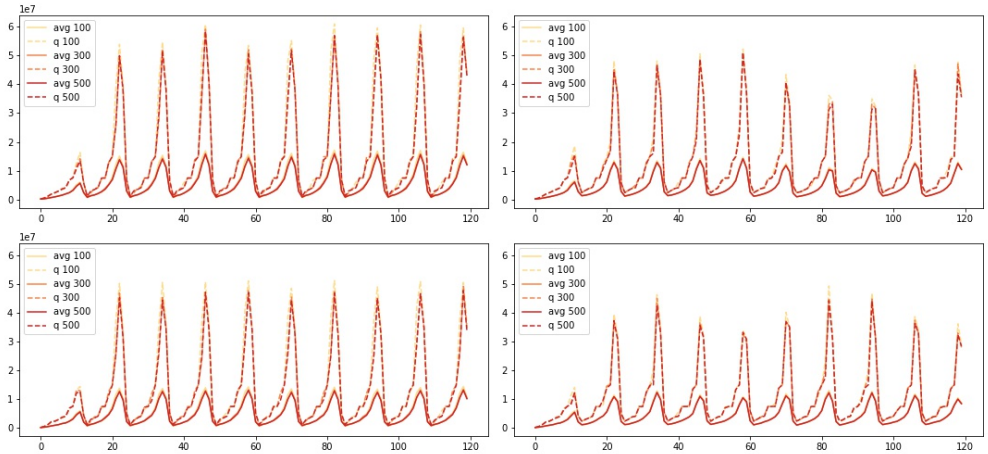
The Brazilian hydro power operation planning problem

The Brazilian hydro power operation planning problem is a multistage, large scale (more than 200 power plants, of which 141 are hydro plants), stochastic optimization problem. On a high level, planning is for 5 years on monthly basis together with 5 additional years to smooth out the end of horizon effect. This results in 120-stage stochastic programming problem. Four energy equivalent reservoirs are considered, one in each one of the four interconnected main regions, SE, S, N and NE. The resulting policy obtained with the aggregate representation can be further refined, so as to provide decisions for each of the hydro and thermal power plants.

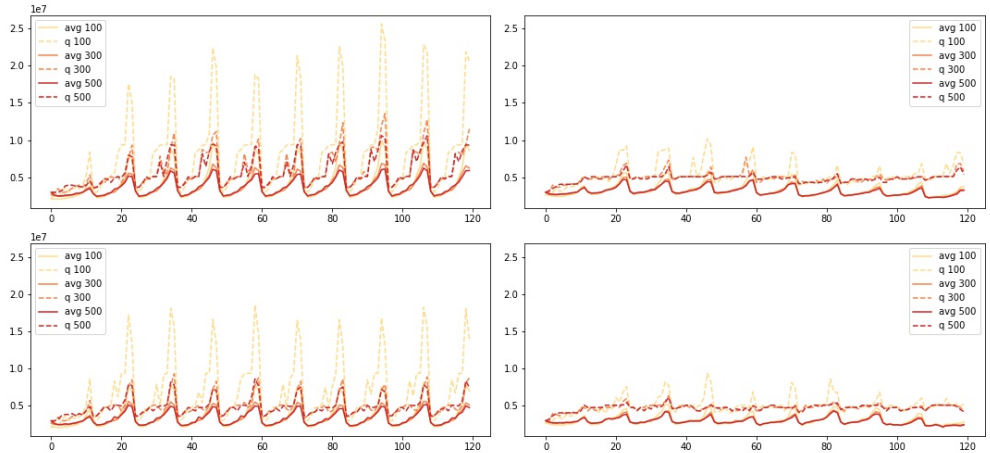

± 3,400 km
± 3,400 km
Belém
São Luis
Fortaleza
Xingu
Tocantins
Teresina
Natal
João Pessoa
Parnaíba
Recife
Madeira
São Francisco
Maceió
Aracajú
Salvador
Cuiabá
Brasília
Paraguai
Goiânia
Paranaíba
Belo Horizonte
Grande
Campo Grande
Vitória
Paraíba do Sul
Paraná/Tietê
Paranapanema
Rio de Janeiro
Itaipu
14.000 MW
Paraguai
São Paulo
Iguaçu
Curitiba
Garabi
2.178 MW
Argentina
Uruguai
Florianópolis
Jacui
Porto Alegre
Legend
Existing Future
138 kV
230 kV
345 kV
440 kV
500 kV
750 kV
±600 kV cc
Load Center
Total Circuits
Watershed
Hydroplant

Comparison of the classical and periodical SDDP (with 8 state variables, period $m = 12$)

Stored energy (in average value and 0.9 quantile) by periodical SDDP (on the left) and classical SDDP (on the right) for the SAA discretization problem (100 samples per stage) and the true problem (on the bottom) for the risk neutral case with discount factor $\gamma = 0.8$

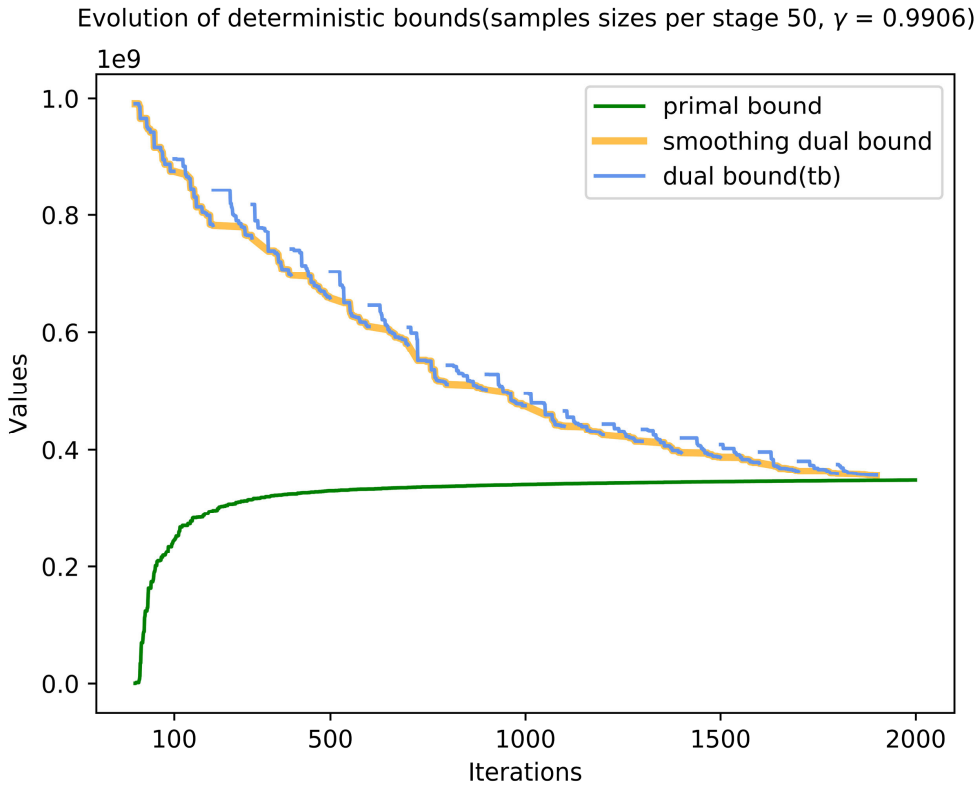


Individual stage costs (in average value and 0.9 quantile) by periodical SDDP (on the left) and classical SDDP (on the right) for the SAA discretization problem (on the above) and the true problem (on the bottom) for the risk neutral case with discount factor $\gamma = 0.9906$ (this γ corresponds to the annual discount rate of 12%, that is $1/\gamma^{12} = 1.12$)



Dual bounds for periodical problem

Hydro-thermal problem with 4 state variables, 50 samples per stage, discount factor $\gamma = 0.9906$ and period $\mathsf{m} = 12$. Evolution of deterministic bounds of primal and dual periodical programs.


Evolution of deterministic bounds(samples sizes per stage 50, γ = 0.9906)
1e9
1.0
0.8
0.6
0.4
0.2
0.0
Values
100
500
1000
1500
2000
Iterations
primal bound
smoothing dual bound
dual bound(tb)

Some references

Shapiro, A. , Dentcheva, D. & Ruszczynski, A. (2014). Lectures on stochastic programming: modeling and theory (2nd). Philadelphia: SIAM .

Shapiro, A. and Nemirovski, A., On complexity of stochastic programming problems, Continuous Optimization: Current Trends and Applications, pp. 111-144, Springer, 2005.

Shapiro, A. and Ding, L., Periodical Multistage Stochastic Programs, *SIAM Journal on Optimization*, vol. 30, pp. 2083 - 2102, 2020.

Ding, L., Ahmed, S. and Shapiro, A., A Python package for multi-stage stochastic programming,
http://www.optimization-online.org/DB_HTML/2019/05/7199.ht