Stochastic programming
Methods based of the stochastic program Lagrangian

Fabian Bastin
fabian.bastin@cirrelt.ca
Université de Montréal – CIRRELT

Motivation

Reference: Birge and Louveaux, Sections 3.4, 5.8 and 6.4.

We consider the general two-stage nonlinear program

$$\begin{aligned} \inf z\ &= f^1(x) + \mathcal{Q}(x) \\ \text{s.t. } & g_i^1(x) \leq 0,\ i = 1, \dots, m_1, \end{aligned}$$

where $\mathcal{Q}(x) = E_{\boldsymbol{\xi}}[Q(x, \xi)]$, and

$$\begin{aligned} Q(x, \xi) = \inf\ & f^2(y(\xi), \xi) \\ \text{t.q. } & b_i^2(x, \xi) + g_i^2(y(\xi), \xi) \leq 0,\ i = 1, \dots, m_2. \end{aligned}$$

We assume that all functions $f^2(\cdot, \xi)$, $b_i^2(\cdot, \xi)$ et $g_i^2(\cdot, \xi)$ are continuous in the first argument for any given ξ, and measurable in ξ for any fixed first argument.

Properties - assumptions

We can extend the definitions of K_1 and K_2.

$$\begin{aligned} K_1 &= \{x \mid g_i^1(x) \leq 0,\ i = 1, \dots, m_1\}, \\ K_2(\xi) &= \{x \mid \exists y(\xi) \text{ t.q. } b_i^2(x, \xi) + g_i^2(y(\xi), \xi) \leq 0,\ i = 1, \dots, m_2\}, \\ K_2 &= \{x \mid \mathcal{Q}(x) < \infty\}. \end{aligned}$$

Assumptions.

1. Convexity. The function f^1 is convex on $\mathbb{R}^{n_1}$, g_i^1 is convex on $\mathbb{R}^{n_1}$ $(i = 1, \dots, m_1)$, $f^2(\cdot, \xi)$ is convex on $\mathbb{R}^{n_2}$ for all $\xi \in \Xi$, $g_i^2(\cdot, \xi)$ is convex on $\mathbb{R}^{n_2}$ $(i = 1, \dots, m_2)$, for all $\xi \in \Xi$, and $b_i^2(\cdot, \xi)$ is convex on $\mathbb{R}^{n_1}$ $(i = 1, \dots, m_2)$, for all $\xi \in \Xi$.
2. Slater condition If $\mathcal{Q}(x) < \infty$, for almost every (a.e.) $\xi \in \Xi$, $\exists$ some $y(\xi)$ s.t. $b_i^2(x, \xi) + g_i^2(y(\xi), \xi) < 0$ $(i = 1, \dots, m_2)$.

Convexity

Slater condition is a regularity condition allowing to apply the KKT conditions.

Theorem

Under assumptions 1 and 2, the recourse function $Q(x, \xi)$ is a convex function in x for all $\xi \in \Xi$.

Theorem

If the recourse feasible set (i.e. in y, for given ξ) is bounded for any $x \in \mathbb{R}^{n_1}$, then the function $Q(x, \xi)$ is lower semi-continuous in x for all $\xi \in \Xi$.

Lower semi-continuity

We say that a function f is lower semi-continuous in x_0 if for every $\varepsilon > 0$, $\exists$ a neighborhood U of x_0 such that $f(x) > f(x_0) - \varepsilon$ for all $x \in U$. Equivalently

$$\liminf_{x \to x_0} f(x) \geq f(x_0),$$

where

$$\liminf_{x \to a} f(x) = \lim_{\varepsilon \to 0} \inf\{f(x) : x \in \mathrm{dom}(f) \cap B(a; \varepsilon) \backslash \{a\}\}.$$

The function f is lower semi-continuous if it is lower semi-continuous in any point in its domain.

A function is lower semi-continuous iff $\{x \in X : f(x) > \alpha\}$ is an open set for every $\alpha \in \mathbb{R}$, or, in a similar way, $\{x \in X : f(x) \leq \alpha\}$ is a closed set for every $\alpha \in \mathbb{R}$.

Convexity (cont'd)

Corollary

The expected recourse function $\mathcal{Q}(x)$ is a convex function in x.

Corollary

The feasible set $K_2 = \{x \mid \mathcal{Q}(x) < \infty\}$ is closed and convex.

Corollary

If the recourse feasible set if bounded for any $x \in \mathbb{R}^{n_1}$, $\mathcal{Q}$ is a lower semi-continuous function.

Solution: existence

Theorem

If the recourse feasible set is bounded for any $x \in \mathbb{R}^{n_1}$, K_1 is bounded, f^1 is continuous, g_i^1 and g_i^2 are continuous for every i, and $K_1 \cap K_2 \neq \emptyset$, then the nonlinear stochastic two-stage program has an optimal solution and the infimum is reached.

Theorem

If it exists a x such that $x \in ri(dom(f^1(x))$, $x \in ri(dom(\mathcal{Q}(x)))$ and $g_i^1(x) < 0$ ($i = 1, \ldots, m_1$), then x^* is optimal iff $x^* \in K_1$ and $\exists\, \lambda_i^* \geq 0$ ($i = 1, \ldots, m$), such that $\lambda_i^* g_i^1(x^*) = 0$ ($i = 1, \ldots, m$), and

$$0 \in \partial f^1(x^*) + \partial \mathcal{Q}(x^*) + \sum_{i=1}^{m_1} \lambda_i^* \partial g_i^1(x^*).$$

Relative interior

The relative interior of a set S, denoted by $ri(S)$), is defined as its interior within the affine envelop of S. In other words

$$ri(S) = \{x \in S \,|\, \exists\, \varepsilon > 0, (B_\varepsilon(x) \cap \mathrm{aff}(S)) \subseteq S\},$$

where $\mathrm{aff}(S)$ is the affine envelop of S, and $B_\varepsilon(x)$ is a ball of radius ε centered at x. Any metric can be used for the ball construction: all define the same relative interior.

The affine envelop $\mathrm{aff}(S)$ of S is the set of all affine combinations of elements of S, i.e.

$$\mathrm{aff}(S) = \left\{ \sum_{i=1}^{k} \alpha_i x_i \middle| x_i \in S,\ \alpha_i \in \mathbb{R},\ \sum_{i=1}^{k} \alpha_i = 1, k = 1, 2, \ldots \right\}.$$

Basic principle of Lagrangian approach

- Problem: classical nonlinear methods assume that the (sub-)gradients of $\mathcal{Q}$ are available and cheap to obtain. Not the case here.
- Link the first and second stages in order to avoid optimization subproblems when building search directions.

Consider the stochastic problem under the form

$$\begin{aligned} \inf\ z &= f^1(x) + E_{\xi}[f^2(y(\xi), \xi)] \\ \text{s.t. } g_i^1(x) &\leq 0,\ i = 1, \ldots, m_1, \\ & b_i^2(x, \xi) + g_i^2(y(\xi), \xi) \leq 0, i = 1, \ldots, m_2\ a.s, \end{aligned}$$

where *a.s.* means almost surely.

Basic principle of Lagrangian approach (cont'd)

Let (λ, π) be a vector of (dual) multipliers associated to the constraints. We can form a dual problem as follows:

$$\max_{\pi(\xi)\geq 0} w = \theta(\pi),$$

where

$$\begin{aligned}\theta(\pi) = \inf_{x,\mathbf{y}} f^1(x) + E_{\boldsymbol{\xi}}[f^2(y(\xi), \xi)] + \\ E_{\xi}\left[\sum_{i=1}^{m_2} \pi_i(\xi)\left(b_i^2(x,\xi) + g_i^2(y(\xi),\xi)\right)\right] \\ \text{s.t. } g_i(x) \leq 0, i = 1, \ldots, m_1.\end{aligned}$$

We can see that we only consider the dual based on the second-stage constraints (linking first and second stages). We will establish some duality properties in case of finite distribution.

Duality

Theorem

We assume that all the functions of the stochastic program are convex, and there exists a finite optimal value, as well as a point strictly satisfying all the constraints. Assume moreover that the support of ξ is finite.

Then $z \geq w$ for all $x, y_1, \dots, y_S$ feasible in the primal formulation and $\pi_1, \dots, \pi_S$ feasible in the dual formulation.

Moreover, their optimal values coincide: $z^* = w^*$.

We thus assume from now that Ξ is finite, and we will describe some procedures exploiting the Lagrangian function.

Basic Lagrangian dual ascent method

Assumption: the dual problem always has a unique solution.

Algorithme 1: Lagrangian dual ascent method

Step 0. Let $\pi^0 \geq 0$, $\nu = 0$.

Step 1. Given $\pi = \pi^\nu$ in the dual problem, consider the solution $(x^\nu, y_1^\nu, \dots, y_S^\nu)$. For $s = 1, \dots, S$, define $\hat{\pi}_i^s$ as

$$\hat{\pi}_i^s = \begin{cases} 0 \text{ if } \pi_i^{\nu,s} = 0 \text{ and } b_i^2(x^\nu, s) + g_i^2(y_s^\nu, s) \leq 0, \\ b_i^2(x^\nu, s) + g_i^2(y_s^\nu, s) \text{ otherwise.} \end{cases}$$

If $\hat{\pi}^s = 0$ for all s, stop.

Step 2. Let λ^ν maximizing $\theta(\pi^\nu + \lambda\hat{\pi})$ on $\pi^\nu + \lambda\hat{\pi} \geq 0,\ \lambda \geq 0$. Set $\pi^{\nu+1} = \pi^\nu + \lambda^\nu\hat{\pi}$, $\nu = \nu + 1$. Return to Step 1.

Properties - convergence

- Under the assumption of unicity of dual solution, we can show that this algorithm always produces an ascent direction in θ.
- Either the algorithm converges to an optimal solution, or, assuming a bounded set of optimal solutions, it produces an infinite sequence where all limit points are optimal.
- Good performance if the number of dual iterations is small compared to the number of function evaluations that would be required by solving the original problem directly.
- Solving the dual program may be time-consuming, but should nevertheless be easier than solving the original problem directly, as the constraints linking the two stages (i.e. involving x and $y(\xi)$) now appear in the objective.

Another approach: scenarios

As Ξ is finite, we can consider that S scenarios exist, and we can link a non-anticipative decision $\hat{x}$ and the decisions of the scenario s with the terms $(\hat{x} - x_k)$ in the objective function.

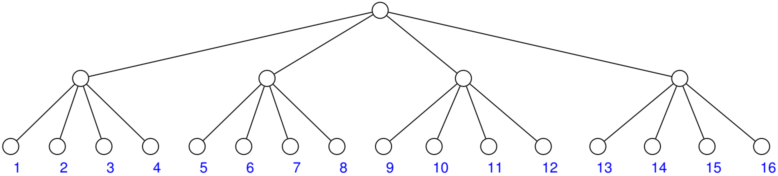
Consider the general problem

$$\begin{aligned} &\inf_{x \in \mathcal{N}} E\left[\sum_{t=0}^{T} f_{t+1}(\xi, x^t(\xi), x_{t+1}(\xi))\right], \\ &\text{s.t. } x_t(\xi) \in X_t(\xi) \text{ a.s.}, \end{aligned}$$

where x_t is the decision at stage t, and x^t is the decisions history at stage t, i.e. $x^t = \{x_1, \dots, x_t\}$. $\mathcal{N}$ designs the close linear subspace of nonanticipative processes. Denoting $\mathcal{A}$ the σ-field of all the events (i.e. the collection of events subsets, with the probability measure associated to ξ).

Scenarios tree

Conceptually, the sequence of random events can be organized in a tree. There are as many leaves as scenarios.

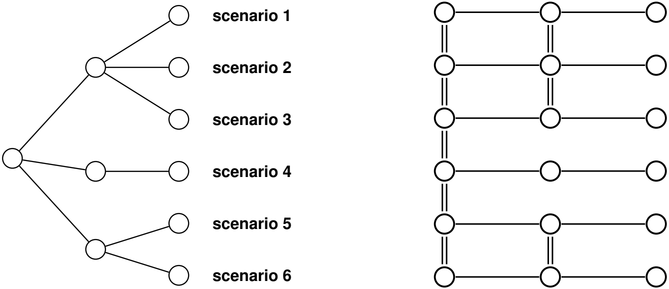


The scenarios are not necessarily equiprobable, the variables can be correlated,. . .

Nonanticipativity

We had to enforce nonanticipativity as we have created copies of the variables at each stage.

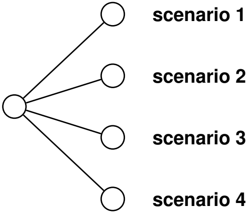
Graphically:

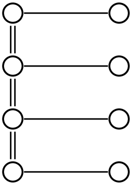


Nodal approach

- Create a vector of variables for each node in the tree.
- This vector corresponds to the decision to take given known realizations of the random variables at this stage (i.e. at this point in the graph).
- Index the nodes by $l = 1, 2, \ldots, \mathcal{L}$.
- We have to know the parent of each node. Let $A(l)$ be the ancestor of node $l \in \mathcal{L}$ in the scenarios tree.

Nonanticipativity: 2 stages





Mathematically,

- the nonanticipativity aims to require that $x^t(\xi)$ has to be $\mathcal{A}^t$-measurable, where $\mathcal{A}^t$ is the σ-field of events until time t;
- or, in other words, $x^t(\xi) = E[x^t(\xi) \,|\, \mathcal{A}^t]$ a.s., $t = 0, \ldots$;
- using the projection operator $\Pi^t : z \to \Pi^t z := E[z|\mathcal{A}^t]$, this is equivalent to

$$(I - \Pi^t)x^T = 0, \;\; t = 0, \ldots$$

Nonanticipativity (cont'd)

More simply, and as seen in the introduction, let S_s^t be the set of scenarios identical to the scenario s at time t. We have to ensure that

$$x_{its} = x_{its'}, \ \forall i \in N, \ \forall t \in T, \ \forall s \in S, \ \forall s' \in S_s^t.$$

In other terms, the scenarios share identical decisions as long as they share the same history, i.e. the observed realizations until a given stage are the same. As soon as the scenarios diverge, the associated decisions can be different.

Idea: explicitly include the nonanticipativity constraints, and place them in the objective function, rather than in the second stage constraints.

Dual ascent and scenarios

Consider the maximization of the dual built on the nonanticipativity constraints:

$$\max_{\pi(\xi)\geq 0} w = \theta(\pi),$$

where

$$\begin{aligned}\theta(\pi) = \inf_{x\in\mathcal{X}} z =& E\left[\sum_{t=0}^{T} f_{t+1}(\xi, x^t(\xi), x_{t+1}(\xi)\right] \\ &+ E\left[\sum_{t=0}^{T} \pi^t(\xi)(I - \Pi_t)x^t(\xi)\right],\end{aligned}$$

where π^t corresponds to the components of π associated to the t first periods.

Dual ascent and scenarios: two stages

We still have to describe the projection operator. For simplicity, consider two steps. The primal problem becomes (omitting all the constraints except the nonanticipativity ones)

$$\begin{aligned} \min z \; &= \sum_{s=1}^{S} p_s \left(f^1(x_s) + f^2(x_s, y_s) \right) \\ \text{s.t. } x_k - &\sum_{s=1}^{S} p_s x_s = 0, \;\; s = 1, \ldots, S. \end{aligned}$$

The dual problem can now be written as

$$\begin{aligned} \max_{\pi} \theta(\pi) = \min_{x,y} \sum_{s=1}^{S} p_s \Bigg(& f^1(x_s) + f^2(x_s, y_s)) \\ & + \pi_s \left(x_s - \sum_{s=1}^{S} p_s x_s \right) \Bigg). \end{aligned}$$

Dual ascent and scenarios: algorithm

```
Algorithme 2: Lagrandian dual ascent method
    Step 0. Let $\pi^0 \geq 0$, $\nu = 0$.
    Step 1. Given $\pi = \pi^\nu$ in the dual problem, compute the
            solution $(x_1^\nu, \dots, x_S^\nu, y_1^\nu, \dots, y_S^\nu)$.
    Step 2. If $x_s - \sum_{k=1}^S p_k x_k = 0$, $s = 1, \dots, S$, stop: the
            solution is optimal.
            Otherwise, define $\hat{\pi}_s = x_s - \sum_{k=1}^S p_k x_k$, and go
            to Step 3.
    Step 3. Let $\lambda^\nu$ maximize $\theta(\pi^\nu + \lambda\hat{\pi})$ on
            $\pi^\nu + \lambda\hat{\pi} \geq 0, \ \lambda \geq 0$. Set $\pi^{\nu+1} = \pi^\nu + \lambda^\nu\hat{\pi}$,
            $\nu = \nu + 1$. Return to Step 1.
```

Augmented Lagrangian

Unfortunately, this type of procedure is usually slow as there is a linearization at a single point of θ only. It is nevertheless easy to implement and can give good results, especially for small size problems.

In order to improve the performances, we will turn to another technique inspired by nonlinear programming: augmented Lagrangian approaches.

The basic idea in an augmented Lagrangian procedure is to add a penalty of $\theta(\pi)$ and to build iterations by including this term.

Moreover, with augmented Lagrangian techniques, we can try to exploit the problem structure in order to decompose the problem.

Augmented Lagrangien

A way to limit the difficulties associated to penalty methods is to add the penalization term to the Lagrangian rather than the objective, leading to the Augmented Lagrangian function:

$$\mathcal{L}_A(x, \lambda; \mu) \overset{def}{=} f(x) + \sum_{i \in \mathcal{E}} \lambda_i c_i(x) + \frac{1}{2\mu} \sum_{i \in \mathcal{E}} c_i^2(x).$$

We will then minimize $\mathcal{L}_A(x, \lambda; \mu)$ with respect to x, and use some rules to update the Lagrange multipliers, at each iteration.

Augmented Lagrangian and Stochastic Programming

We again develop the ideas in the two-stages contact, while the approach can be easily generalized in the multi-stage context.

Recall that we aim to solve

$$\max_{\pi(\xi) \geq 0} w = \theta(\pi),$$

where

$$\begin{aligned}\theta(\pi) = \inf_{x \in \mathcal{X}} z = & E\left[\sum_{t=0}^{T} f_{t+1}(\xi, x^t(\xi), x_{t+1}(\xi))\right] \\ & + E\left[\sum_{t=0}^{T} \pi^t(\xi)(I - \Pi_t)x^t(\xi)\right],\end{aligned}$$

Penalizing the nonanticipativity constraints, we obtain the following program, where $\hat{x}$ is nonanticipative.

Augmented Lagrangian and SP (cont'd)

$$
\begin{aligned}
\theta(\rho) = \inf z = f^1(\hat{x}) + \sum_{s=1}^{S} \left(p_s f^2(y_s, \xi_s) + \rho_s^T (x_s - \hat{x}) + \frac{r}{2} \|x_s - \hat{x}\|^2 \right) \\
\text{s.t. } g_i^1(\hat{x}) \leq 0,\ i = 1, \ldots, m_1, \\
t_i^2(x_s, \xi_s) + g_i^2(y_s, \xi_S) \leq 0,\ i = 1, \ldots, m_2,\ s = 1, \ldots, S.
\end{aligned}
$$

r if the augmented Lagrangian penalty parameter, and ρ_s is the vector of dual variables associated to the nonanticipativity constraints of the scenario s.

The method that we will develop aims to contract the pair $(\hat{x}^{\nu+1}, \rho^{\nu+1})$ around a saddle point.

Progressive hedging

- Main reference: R. T. Rockafellar and R. J.-B. Wets, *Scenarios and policy aggregation in optimization under uncertainty*, Mathematics of Operations Research 16(1):119–147 (1991).
  The paper focussed on the multistage version.
- The method performs a complete separation of the problems, scenario by scenario, at each iteration, reducing the cost per iteration.
- The number of iterations can however increase…
- But for structures problems, it is possible to exploit parallelism, and therefore solve large-scale problems.
- Observation: the Lagrangian function is not separable with respect to the scenarios due to the term $(\hat{x} - x_s)$.

Progressing hedging: basic principle

Alternatively,

1. fix $\hat{x}$ and obtain the solutions x_s, $s = 1, \ldots, S$,
2. fix x_s, $s = 1, \ldots, S$, and compute $\hat{x}$.

In other words, we work scenario by scenario, and we enforce progressively the nonanticipativity constraints.

At iteration ν, we solve the subproblems

$$
\begin{aligned}
\inf z = \sum_{ks=1}^{S} p_s \left(f^1(x_s) + f^2\left(y_s, \xi_s\right) + \left(\rho_s^\nu\right)^T \left(x_s - \hat{x}^\nu\right) + \frac{r}{2} \|x_s - \hat{x}^\nu\|^2 \right) \\
\text{s.t. } g_i^1(x_s) \leq 0,\ i = 1, \ldots, m_1,\ s = 1, \ldots, S, \\
t_i^2\left(x_s, k\right) + g_i^2\left(y_s, \xi_s\right) \leq 0,\ i = 1, \ldots, m_2,\ s = 1, \ldots, S.
\end{aligned}
$$

$\hat{x}^\nu$ is not necessarily a feasible solution!

Decomposition

However, it is easy to decompose the problem!
Algorithm:

Step 0. Assume that we have a nonanticipative solution x^0, an initial vector ofmultipliers ρ^0, and $r > 0$. $\nu \leftarrow 0$. Go to Step 1.

Step 1. Let $\left(x_s^{\nu+1}, y_s^{\nu+1}\right)$ be a solution of the previous Lagrangian program, for $s = 1, \ldots, S$. Set

$$\hat{x}^{\nu+1} = \left(\hat{x}_1^{\nu+1}, \ldots, \hat{x}_S^{\nu+1}\right)^T,$$

$$\text{where } \hat{x}_s^{\nu+1} = \sum_{s=1}^{S} p_s x_s^{\nu+1}, \; s = 1, \ldots, S.$$

Decomposition (cont'd)

Step 2. Let $\rho_s^{\nu+1} = \rho_s^{\nu} + r\left(x_s^{\nu+1} - \hat{x}_s^{\nu+1}\right)$, $s = 1, \dots, S$.
If $\hat{x}^{\nu+1} = \hat{x}^{\nu}$ and $\rho^{\nu+1} = \rho^{\nu}$, stop: $(\hat{x}^{\nu}, \rho^{\nu})$ is optimal.
Otherwise, set $\nu = \nu + 1$ and return to Step 1.

- Step 2 simply consists to take the expected value of $x^{\nu+1}$ as $\hat{x}^{\nu+1}$.
- The basis of this approach is therefore the contraction of the pair $(\hat{x}^{\nu}, \rho^{\nu})$ around a saddle point rather than a dual ascent strategy.
- In practice, the optimality test is replaced by a convergence test for instance if the following quantity is small enough (De Silva et Abramson):

$$\sqrt{\|\hat{x}^{\nu} - \hat{x}^{\nu+1}\|^2 + \sum_{s=1}^{S} p_s \|x_s^{\nu} - \hat{x}^{\nu}\|^2}.$$

Generalization

$$\begin{aligned} \min_{x} \ & \sum_{s\in\mathcal{S}} p_s f\left(x^{(s)}, \xi^{(s)}\right) \\ \text{s.t. } & x^{(s)} \in \mathcal{X}^{(s)}, \\ & x_t^{(s)} \text{ is nonanticipative}, \ t = 1, \ldots, T. \end{aligned}$$

Nonanticipativity:

$$x_t^{(s)} = E\left[x_t^{(s')} \,\middle|\, s' \in \mathcal{S}_t^{(s)}\right],$$

where $\mathcal{S}_t^{(s)} = \left\{s' \,\middle|\, \bar{\xi}_t^{(s)} = \bar{\xi}_t^{(s')}\right\}$. It can be reformulated as

$$x_t^{(s)} = \hat{x}_t^{(s)},$$

with

$$\hat{x}_t^{(s)} = \frac{\sum_{s'\in\mathcal{S}_t^{(s)}} p_{s'} x_t^{(s')}}{\sum_{s'\in\mathcal{S}_t^{(s)}} p_{s'}}.$$

Generalization

Augmented Lagrangian:

$$L(x,\lambda,\rho) = E\left[f\left(x^{(s)},\xi^{(s)}\right) + \sum_{t=1}^{T}\left(\lambda_t^{(s)'}\left(x_t^{(s)} - \hat{x}_t^{(s)}\right) \right.\right.$$
$$\left.\left. + \frac{\rho}{2}\left\| x_t^{(s)} - \hat{x}_t^{(s)} \right\|^2 \right) \right],$$

where λ is the Lagrange multipliers vector associated to the NA constraints, and $\rho > 0$ is a penalty parameter.

Given λ, the augmented Lagrangian program is

$$\begin{aligned} &\min_x L(x,\lambda,\rho) \\ &\text{s.t. } x^{(s)} \in \mathcal{X}^{(s)},\ s \in \mathcal{S}. \end{aligned} \qquad (1)$$

In order to achieve full separability, fix $\hat{x}_t^{(s)}$ and repeatedly solve the program by updating the Lagrange multipliers vector and the value of $\hat{x}_t^{(s)}$ between consecutive resolutions.

Progressive hedging algorithm (PHA)

Step 0. Set $\hat{x}^{(s),0} = \left(\hat{x}_1^{(s),0}, \ldots, \hat{x}_T^{(s),0}\right)$ and $k = 0$. Choose $\lambda^{(s),0} = \mathbf{0}$, $\rho^0 > 0$.

Step 1. Compute $x^{(s),k+1} = (x_1^{s,k+1}, \ldots, x_T^{s,k+1})$, $s = 1, \ldots, S$, by solving each scenario subproblem

$$\begin{aligned} \min_{x^{(s)}} \; & f\left(x^{(s)}, \xi^{(s)}\right) + \sum_{t=1}^{T} \left(\lambda_t^{(s)'}\left(x_t^{(s)} - \hat{x}_t^{(s),k}\right)\right. \\ & \left. + \frac{\rho^k}{2}\left\|x_t^{(s)} - \hat{x}_t^{(s),k}\right\|^2\right) \\ \text{s.t. } & x^{(s)} \in \mathcal{X}^{(s)}. \end{aligned}$$

Step 2. For $s = 1, \ldots, S$, $t = 1, \ldots, T$, set

$$\hat{x}_t^{(s),k+1} = \frac{\sum_{s' \in \mathcal{S}_t^{(s)}} p_{s'} x_t^{(s'),k+1}}{\sum_{s' \in \mathcal{S}_t^{(s)}} p_{s'}}.$$

Progressive hedging algorithm (PHA)

Step 3. Set ρ^{k+1} and

$$\lambda_t^{(s),k+1} = \lambda_t^{(s),k} + \rho^k \left(x_t^{(s),k+1} - \hat{x}_t^{(s),k+1} \right),$$

for $t = 1, \ldots, T$, $s \in \mathcal{S}$.

Step 4. Stop if convergence is achieved. Otherwise, set $k \leftarrow k + 1$ and return to Step 1.

Convergence

Theorem

Assume that the initial nonlinear stochastic program has only convex functions and a finite optimal value, as well as a strictly feasible point. Assume moreover that the support of ξ is finite. The progressive hedging algorithm converges to an optimal solution x^*, ρ^*.

- De Silva and Abramson have tested the progressive hedging algorithm for linear problems in portfolio management, proposed by Mulvey and Vladimirou.
- The subproblems where solved by means of interior points methods.
- Good speed-up convergence, but the algorithm remains sensitive to the choice of the penality parameter.