A bundle-like approach to induce monotonicity in the Progressive Hedging algorithm

Felipe Atenas   Claudia Sagastizábal
PhD student supervised by Claudia Sagastizábal and Paulo Silva

Robustness and Resilience in Stochastic Optimization and Statistical Learning: Mathematical Foundations
Erice 2022

IMECC, State University of Campinas UNICAMP (Brazil)
Research supported by FAPESP grant 2019/20023-1

Outline

1. Motivation
2. Progressive Hedging Algorithm
3. Bundle methods
4. Bundle Progressive Hedging
5. Convergence analysis

Motivation

Why decomposition? Why descent?

- Large scale problems provide a suitable setting for applying decomposition methods.
- Decomposition in stochastic optimization problems amounts to considering separately the impact of different realizations of uncertainty.
- The Progressive Hedging algorithm uses this structure to solve separate scenario subproblems.
- Stopping test: can we use function values to evaluate the quality of the candidate point for a solution? Example: bundle methods.

Progressive Hedging Algorithm

Nonanticipativity constraint

1st stage decision $\rightarrow$ revealed information $\rightarrow$ 2nd stage decision $\rightarrow$ revealed information $\rightarrow$ 3rd stage decision $\dots$

One-stage stochastic primal problem

$$\begin{cases} \min & \displaystyle\sum_{s=1}^{S} p_s f_s(x_s) \\ \text{s.t.} & x_s = z, \text{ for } s = 1, \ldots, S \end{cases}$$

One-stage stochastic primal problem

$$\begin{cases} \min & \displaystyle\sum_{s=1}^{S} p_s f_s(x_s) \\ \text{s.t.} & x_s = \mathbb{E}[x], \text{ for } s = 1, \ldots, S \end{cases}$$

One-stage stochastic primal problem

$$\begin{cases} \min & \displaystyle\sum_{s=1}^{S} p_s f_s(x_s) \\ \text{s.t.} & x_s = \mathbb{E}[x], \text{ for } s = 1, \ldots, S \quad (\leftarrow w_s) \end{cases}$$

Relax the coupling constraints to obtain a separable Lagrangian, for $w \in \mathcal{N}^{\perp}$:

$$\begin{aligned} L(x, w) &= \mathbb{E}[f(x)] + \langle w, x - \mathbb{E}[x] \rangle_S \\ &= \sum_{s=1}^{S} p_s \left(f_s(x_s) + \langle w_s, x_s \rangle \right) \end{aligned}$$

Stochastic formulation: dual function

- Separable Lagrangian:

$$L(x, w) = \sum_{s=1}^{S} p_s L_s(x_s, w_s)$$

where $L_s(x_s, w_s) = f_s(x_s) + \langle w_s, x_s \rangle$.

Stochastic formulation: dual function

- Separable Lagrangian:

$$L(x, w) = \sum_{s=1}^{S} p_s L_s(x_s, w_s)$$

where $L_s(x_s, w_s) = f_s(x_s) + \langle w_s, x_s \rangle$.

- Coupling constraints with separable objective function:

$$\max_{w \in \mathcal{N}^\perp} \min_{x} L(x, w)$$

Stochastic formulation: dual problem

- Separable dual function:

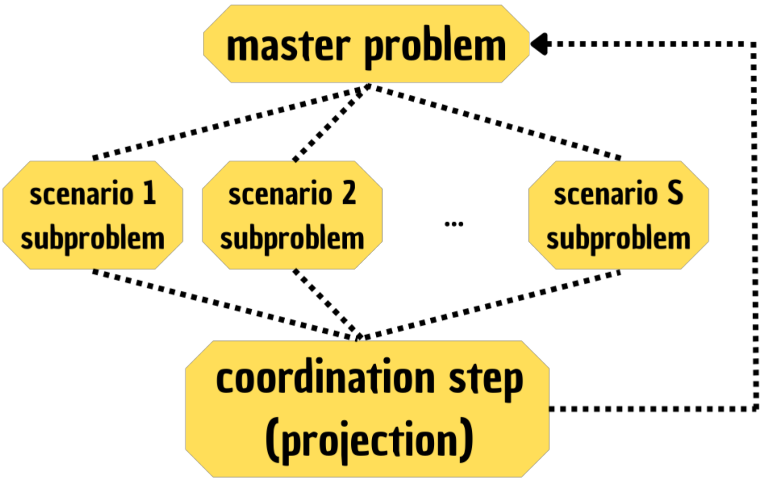
$$h(w) = \sum_{s=1}^{S} p_s h_s(w_s)$$

for $h_s(w_s) = \max\{-L_s(x_s, w_s)\}$.

- Dual problem:

$$\begin{cases} \min\limits_{w} & h(w) \\ \text{s.t.} & w \in \mathcal{N}^{\perp} \qquad \left(\sum_{s=1}^{S} p_s w_s = 0\right) \end{cases}$$

Decomposition for different scenarios



Progressive Hedging Algorithm: primal subproblems

The PHA employs an Augmented Lagrangian that replaces $\mathbb{E}[x]$ with $\mathbb{E}[x^k]$.

$$\mathbb{E}[f(x)] + \langle w, x - \mathbb{E}[x] \rangle_S + \frac{r}{2} \| x - \mathbb{E}[x] \|_S^2$$

$$\downarrow$$

$$\mathbb{E}[f(x)] + \langle w, x - \mathbb{E}[x^k] \rangle_S + \frac{r}{2} \| x - \mathbb{E}[x^k] \|_S^2$$

- Advantage: preserve separability.
- Drawback: the quality of this approximation is not checked.

Progressive Hedging Algorithm [RW91, R18]

- **Subproblem solutions**: for each $s = 1, \ldots, S$,

$$x_s^{k+1/2} = \arg\min \left\{ f_s(x_s) + \langle w_s^k, x_s \rangle + \frac{r}{2} |x_s - x_s^k|^2 \right\}$$

- **Primal projection** (onto $\mathcal{N}$): for each $s = 1, \ldots, S$,

$$x_s^{k+1} = \mathbb{E}\left[x^{k+1/2} \right].$$

- **Dual update**: for each $s = 1, \ldots, S$,

$$w_s^{k+1} = w_s^k + r \left(x_s^{k+1/2} - x_s^{k+1} \right).$$

Convergence

For each $s = 1, \ldots, S$, assume that each function f_s is lsc convex.

If the problem has a nonempty set of minimizers, then $\{x^k\}$ converges to a primal solution $x^\star$, and $\{w^k\}$ converges to a dual solution $w^\star$, such that the distance from each primal-dual iterate to the saddle point (x^*, w^*)

$$\left\| \begin{pmatrix} x^k \\ \frac{1}{r} w^k \end{pmatrix} - \begin{pmatrix} x^\star \\ \frac{1}{r} w^\star \end{pmatrix} \right\|_S^2$$

converges to 0 monotonically.

What can be improved?

- PHA's convergence proof is based on the Proximal Point algorithm for a (primal-dual) maximal monotone operator.

What can be improved?

- PHA's convergence proof is based on the Proximal Point algorithm for a (primal-dual) maximal monotone operator.
- The dual update

$$w_s^{k+1} = w_s^k + r\left(x_s^{k+1/2} - \mathbb{E}\left[x^{k+1/2}\right]\right)$$

is an inexact gradient step for the dual function with fix stepsize r.

What can be improved?

- PHA's convergence proof is based on the Proximal Point algorithm for a (primal-dual) maximal monotone operator.
- The dual update

$$w_s^{k+1} = w_s^k + r\left(x_s^{k+1/2} - \mathbb{E}\left[x^{k+1/2}\right]\right)$$

  is an inexact gradient step for the dual function with fix stepsize r. Monotonicity depends on r being sufficiently small.
- Small r slows down convergence (dual term of stopping test becomes large).

Bundle methods

Proximal bundle methods

Consider $\min_w H(w)$, for convex H.

Proximal bundle methods

Consider $\min_w H(w)$, for convex H.

$$\text{Proximal point: } w^{k+1} = \operatorname{prox}_{1/r} H(w^k) := \arg\min_w \left\{ H(w) + \frac{1}{2r}\|w - w^k\|^2 \right\}$$

Proximal bundle methods

Consider $\min_w H(w)$, for convex H.

$$\text{Proximal point: } w^{k+1} = \operatorname{prox}_{1/r} H(w^k) := \arg\min_w \left\{ H(w) + \frac{1}{2r}\|w - w^k\|^2 \right\}$$

- $\operatorname{prox}_{1/r} H$ might be difficult to evaluate.

Proximal bundle methods

Consider $\min_w H(w)$, for convex H.

$$\text{Proximal point: } w^{k+1} = \operatorname{prox}_{1/r} H(w^k) := \operatorname*{arg\,min}_w \left\{ H(w) + \frac{1}{2r}\|w - w^k\|^2 \right\}$$

- $\operatorname{prox}_{1/r} H$ might be difficult to evaluate. Simpler step?
- Use a convex model H^k of H, such that $H^k \leq H$, and variable r.

Proximal bundle methods

Consider $\min_w H(w)$, for convex H.

$$\text{Proximal point: } w^{k+1} = \text{prox}_{1/r} H(w^k) := \arg\min_{w} \left\{ H(w) + \frac{1}{2r}\|w - w^k\|^2 \right\}$$

- $\text{prox}_{1/r} H$ might be difficult to evaluate. Simpler step?
- Use a convex model H^k of H, such that $H^k \leq H$, and variable r.

$$\text{Approximate Proximal point: } w^{k+1} = \text{prox}_{1/r_k} H^k(w^k)$$

Model H^k for H: definition

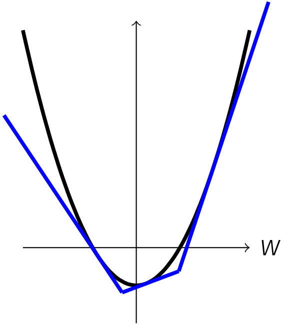
For an index set $B^k \supseteq \{1, \ldots, k\}$ of past iterations:

$$H^k(x) = \max_{i \in B^k} \left\{ H(w^i) + \langle g^i, w - w^i \rangle \right\}$$

where $g^i \in \partial H(w^i)$, that is

$$\forall w, \quad H(w) \geq H(w^i) + \langle g^i, w - w^i \rangle$$

- This model approximates H from below.
- Along iterations, the model is enriched with new linearizations, defining tighter models.



Proximal bundle methods: main idea

$$w^{k+1} = \mathrm{prox}_{1/r_k} H^k(\hat{w}^k)$$

for a center point $\hat{w}^k$.

Proximal bundle methods: main idea

$$w^{k+1} = \mathrm{prox}_{1/r_k} H^k(\hat{w}^k)$$

for a center point $\hat{w}^k$.

Does w^{k+1} give sufficient descent for H?

$$H(w^{k+1}) - H(\hat{w}^k) \leq m(H^k(w^{k+1}) - H(\hat{w}^k)), \quad m \in (0,1)$$

Proximal bundle methods: main idea

$$w^{k+1} = \mathrm{prox}_{1/r_k} H^k(\hat{w}^k)$$

for a center point $\hat{w}^k$.

Does w^{k+1} give sufficient descent for H?

$$H(w^{k+1}) - H(\hat{w}^k) \leq m(H^k(w^{k+1}) - H(\hat{w}^k)), \quad m \in (0,1)$$

Yes: serious step $\rightarrow$ $\hat{w}^{k+1} = w^{k+1}$ – No: null step $\rightarrow$ $\hat{w}^{k+1} = \hat{w}^k$.

Proximal bundle methods: main idea

$$w^{k+1} = \mathrm{prox}_{1/r_k} H^k(\hat{w}^k)$$

for a center point $\hat{w}^k$.

Does w^{k+1} give sufficient descent for H?

$$H(w^{k+1}) - H(\hat{w}^k) \leq m(H^k(w^{k+1}) - H(\hat{w}^k)), \quad m \in (0,1)$$

Yes: serious step $\rightarrow$ $\hat{w}^{k+1} = w^{k+1}$ — No: null step $\rightarrow$ $\hat{w}^{k+1} = \hat{w}^k$.

In both cases, use $H-$information at w^{k+1} to improve the model H^{k+1}.

Proximal bundle methods: main idea

$$w^{k+1} = \text{prox}_{1/r_k} H^k(\hat{w}^k)$$

for a center point $\hat{w}^k$.

Does w^{k+1} give sufficient descent for H?

$$H(w^{k+1}) - H(\hat{w}^k) \leq m(H^k(w^{k+1}) - H(\hat{w}^k)), \quad m \in (0,1)$$

Yes: serious step $\rightarrow$ $\hat{w}^{k+1} = w^{k+1}$ — No: null step $\rightarrow$ $\hat{w}^{k+1} = \hat{w}^k$.

In both cases, use $H-$information at w^{k+1} to improve the model H^{k+1}.

How can we apply these ideas to the PHA?

Dual bundle method for $H = h + i_{\mathcal{N}^\perp}$

$$w^{k+1/2} = \mathrm{prox}_{1/r_k} H^k(\hat{w}^k)$$

for a center point $\hat{w}^k$.

$$w^{k+1/2} = \mathrm{prox}_{1/r_k} H^k(\hat{w}^k)$$

for a center point $\hat{w}^k$.

Dual Projection: $w^{k+1} = P_{\mathcal{N}^\perp}(w^{k+1/2})$.

Dual bundle method for $H = h + i_{\mathcal{N}^\perp}$

$$w^{k+1/2} = \mathrm{prox}_{1/r_k} H^k(\hat{w}^k)$$

for a center point $\hat{w}^k$.

Dual Projection: $w^{k+1} = P_{\mathcal{N}^\perp}(w^{k+1/2})$.

Does w^{k+1} give sufficient descent for H?

Descent test: $H(w^{k+1}) - H(\hat{w}^k) \leq m(H^k(w^{k+1}) - H(\hat{w}^k)), \quad m \in (0,1)$

Dual bundle method for $H = h + i_{\mathcal{N}^\perp}$

$$w^{k+1/2} = \text{prox}_{1/r_k} H^k(\hat{w}^k)$$

for a center point $\hat{w}^k$.

Dual Projection: $w^{k+1} = P_{\mathcal{N}^\perp}(w^{k+1/2})$.

Does w^{k+1} give sufficient descent for H?

Descent test: $H(w^{k+1}) - H(\hat{w}^k) \leq m(H^k(w^{k+1}) - H(\hat{w}^k)), \quad m \in (0,1)$

Yes: serious step $\rightarrow \hat{w}^{k+1} = w^{k+1}$ — No: null step $\rightarrow \hat{w}^{k+1} = \hat{w}^k$.

Dual bundle method for $H = h + i_{\mathcal{N}^\perp}$

$$w^{k+1/2} = \text{prox}_{1/r_k} H^k(\hat{w}^k)$$

for a center point $\hat{w}^k$.

Dual Projection: $w^{k+1} = P_{\mathcal{N}^\perp}(w^{k+1/2})$.

Does w^{k+1} give sufficient descent for H?

Descent test: $H(w^{k+1}) - H(\hat{w}^k) \leq m(H^k(w^{k+1}) - H(\hat{w}^k)), \quad m \in (0,1)$

Yes: serious step $\rightarrow \hat{w}^{k+1} = w^{k+1}$ — No: null step $\rightarrow \hat{w}^{k+1} = \hat{w}^k$.

In both cases, use $H-$information at $w^{k+1/2}$ to improve the model H^{k+1}.

Separable model for $H = h + i_{\mathcal{N}^\perp}$

Model $H^k(\cdot) = h^k(\cdot) + \langle x^k, \cdot \rangle$, for each scenario:

- When evaluating $h(w^{k+1/2})$, we find a primal point $x^{k+1/2}$ such that

$$h(w^{k+1/2}) = -L(x^{k+1/2}, w^{k+1/2}), \text{ and } -x^{k+1/2} \in \partial h(w^{k+1/2}).$$

Separable model for $H = h + i_{\mathcal{N}^\perp}$

Model $H^k(\cdot) = h^k(\cdot) + \langle x^k, \cdot \rangle$, for each scenario:

- When evaluating $h(w^{k+1/2})$, we find a primal point $x^{k+1/2}$ such that

$$h(w^{k+1/2}) = -L(x^{k+1/2}, w^{k+1/2}), \text{ and } -x^{k+1/2} \in \partial h(w^{k+1/2}).$$

- The new generated linearization for the dual function h at iteration k is

$$h(w^{k+1/2}) + \langle -x^{k+1/2}, w - w^{k+1/2} \rangle_S$$

Separable model for $H = h + i_{\mathcal{N}^\perp}$

Model $H^k(\cdot) = h^k(\cdot) + \langle x^k, \cdot \rangle$, for each scenario:

- When evaluating $h(w^{k+1/2})$, we find a primal point $x^{k+1/2}$ such that

$$h(w^{k+1/2}) = -L(x^{k+1/2}, w^{k+1/2}), \text{ and } -x^{k+1/2} \in \partial h(w^{k+1/2}).$$

- The new generated linearization for the dual function h at iteration k is

$$h(w^{k+1/2}) + \langle -x^{k+1/2}, w - w^{k+1/2} \rangle_S$$

- The indicator function $i_{\mathcal{N}^\perp}$ is linearized by means of x^k, the expected value of past primal points, noting that $\partial i_{\mathcal{N}^\perp}(w) = N_{\mathcal{N}^\perp}(w) = \mathcal{N}$.

Separable subproblems: dual BPHA

In the BPHA, for each scenario, $w^{k+1/2}$ solves

$$\min_w \left\{ h^k(w) + \langle x^k, w \rangle + \frac{1}{2r_k} |w - \hat{w}^k|^2 \right\}$$

where

- h^k is a piecewise linear model of h, and
- $x^k \in \mathcal{N}$ defines the separable linearization of the indicator function $i_{\mathcal{N}^\perp}$, as $\langle x^k, w \rangle$.
- $\hat{w}^k$ is the last serious step.

Separable subproblems: primal PHA

In the PHA, for each scenario, $x^{k+1/2}$ solves

$$\min_{x} \left\{ f(x) + \langle w^k, x \rangle + \frac{r}{2}|x - x^k|^2 \right\}$$

where

- w^k is a multiplier associated with the relaxation of the nonanticipativity constraint.
- x^k is the last projected primal point.

PHA and BPHA comparison

PHA

$x^{k+1/2}$ solves

$$\min_x \left\{ f(x) + \langle w^k, x\rangle + \frac{r}{2}|x - x^k|^2 \right\}$$

$$x^{k+1} = \mathbb{E}\left[x^{k+1/2}\right]$$

$$w^{k+1} = w^k + r\left(x^{k+1/2} - x^{k+1}\right)$$

BPHA

$w^{k+1/2}$ solves

$$\min_w \left\{ h^k(w) + \langle \hat{x}^k, w\rangle + \frac{1}{2r_k}|w - \hat{w}^k|^2 \right\}$$

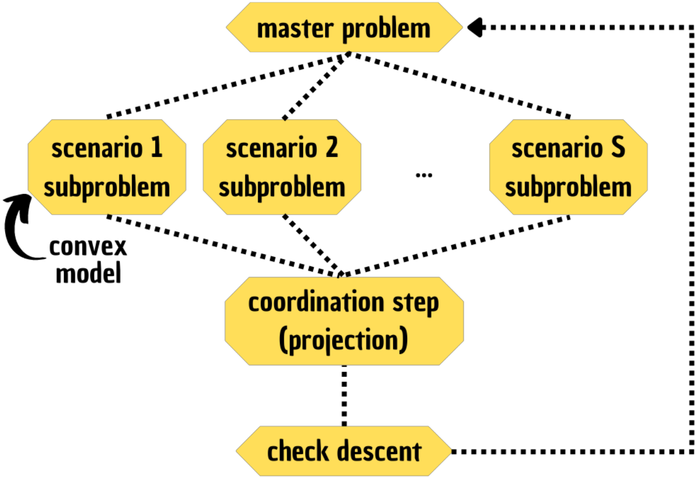
$$x^{k+1} = \mathbb{E}\left[x^{k+1/2}\right]$$

$$w^{k+1} = \hat{w}^k + r_k\left(x^{k+1/2} - x^{k+1}\right)$$

$$h(w^{k+1}) - h(\hat{w}^k) \leq -m\delta^{k+1}$$

Bundle Progressive Hedging

Decomposition for different scenarios + descent



Algorithm in primal terms

- For each scenario, compute $x^{k+1/2}$ in the convex hull of past bundle information.
- Update the intermediate dual point as $w^{k+1/2} = \hat{w}^k + r_k(x^{k+1/2} - x^k)$.
- Project $x^{k+1/2}$ and $w^{k+1/2}$ onto $\mathcal{N}$ and $\mathcal{N}^\perp$, respectively, to obtain x^{k+1} and w^{k+1}.
- Check descent for the dual function to assess adequacy of its model and w^{k+1}. Only when there is sufficient descent, the candidate becomes $\hat{w}^{k+1}$.
- The new generated $H-$information at $w^{k+1/2}$ is used to improve the models.

Primal and dual projection-coordination step

For each scenario,

BPHA

- Primal projection: $x^{k+1} = \mathbb{E}[x^{k+1/2}]$, where $x^{k+1/2}$ is the subproblem solution.
- Dual intermediate point: $w^{k+1/2} = \hat{w}^k + r_k(x^{k+1/2} - x^k)$.
- Dual projection: $w^{k+1} = \hat{w}^k + r_k(x^{k+1/2} - x^{k+1})$.

PHA

- Primal projection: $x^{k+1} = \mathbb{E}[x^{k+1/2}]$, where $x^{k+1/2}$ is the subproblem solution.
- Dual projection: $w^{k+1} = w^k + r\left(x^{k+1/2} - x^{k+1}\right)$.

Bundle Progressive Hedging Algorithm

Stopping test

Define the predicted decrease by the model

$$\delta^{k+1} = \mathbb{E}\left[h^k(w^{k+1/2}) + \langle \hat{x}^k, w^{k+1/2}\rangle - h(\hat{w}^k)\right].$$

If $\delta^{k+1} \leq \mathrm{TOL}_\delta$, stop and return x^{k+1}, and $\hat{w}^{k+1}$.

Descent test

$$\mathbb{E}[h(w^{k+1}) - h(\hat{w}^k)] \leq m\delta^{k+1}?$$

$$\left.\begin{array}{r l} \text{Yes:} & \text{set } \hat{w}^{k+1} = w^{k+1}, \text{ and define } r_{k+1} \geq r_{\min} \textit{ (serious step)} \\ \text{No:} & \text{set } \hat{w}^{+1} = \hat{w}^k, \text{ and define } r_{k+1} \in [r_{\min}, r_k] \textit{ (null step)} \end{array}\right\} \text{ update the model.}$$

Convergence analysis

Analysis à la Bundle

For each scenario,

- There can be infinitely many serious steps: after finitely many null steps, there is a serious step.

$$\mathcal{K} = \{k : w^{k+1} \text{ is a serious step}\}$$

Analysis à la Bundle

For each scenario,

- There can be infinitely many serious steps: after finitely many null steps, there is a serious step.

$$\mathcal{K} = \{k : w^{k+1} \text{ is a serious step}\}$$

- There is a last serious step $\hat{w} := \hat{w}^{\hat{k}}$, after which there are infinitely many null steps.

$$\mathcal{K} = \{k : k > \hat{k}\}$$

Serious steps

If there are **infinitely many serious steps**, then for each scenario and $k \in \mathcal{K}$:

$$w^{k+1/2} = \hat{w}^k + r_k(x^{k+1/2} - x^k),$$
$$\hat{w}^{k+1} = P_{\mathcal{N}^\perp}(w^{k+1/2}),$$
$$h(\hat{w}^{k+1}) - h(\hat{w}^k) \leq m\mathbb{E}[h^k(w^{k+1/2}) + \langle x^k, w^{k+1/2}\rangle - h(\hat{w}^k)], \qquad \text{for } m \in (0,1).$$

Serious steps

If there are **infinitely many serious steps**, then for each scenario and $k \in \mathcal{K}$:

$$\begin{aligned}
w^{k+1/2} &= \hat{w}^k + r_k(x^{k+1/2} - x^k),\\
\hat{w}^{k+1} &= P_{\mathcal{N}^\perp}(w^{k+1/2}),\\
h(\hat{w}^{k+1}) - h(\hat{w}^k) &\le m\mathbb{E}[h^k(w^{k+1/2}) + \langle x^k, w^{k+1/2}\rangle - h(\hat{w}^k)], \qquad \text{for } m \in (0,1).
\end{aligned}$$

Descent steps of the unifying framework of [ASSS21].

[ASSS21]: F. Atenas, C. Sagastizábal, P. J. Silva, and M. Solodov (2021). "A unified analysis of descent sequences in weakly convex optimization, including convergence rates for bundle methods".

Assumptions

- f is lsc convex.
  Remark: h is always convex.
- Error bound for h : for any $u \geq \inf_{w \in \mathcal{N}^\perp} h(w)$, whenever $w \in \mathcal{N}^\perp$, $h(w) \leq u$, there holds

$$d(w, S) \leq \ell \|x\|$$

  for any nearly nonanticipative x solving the problem that defines $h(w)$.

Serious steps: convergence

Assumptions

- f is lsc convex.
  Remark: h is always convex.
- Error bound for h.
- Constraint qualification to guarantee strong duality.

Serious steps: convergence

Assumptions

- f is lsc convex.
  Remark: h is always convex.
- Error bound for h.
- Constraint qualification to guarantee strong duality.

Convergence is shown extending the analysis in [ASSS21] to deal with intermediate projection steps.

Theorem

- Any limit point of $\{x^k\}_{k\in\mathcal{K}}$ is primal optimal, such that $\{\mathbb{E}[f(x^k)]\}_{k\in\mathcal{K}}$ subsequentially converges to the primal optimal value.

Theorem

- Any limit point of $\{x^k\}_{k\in\mathcal{K}}$ is primal optimal, such that $\{\mathbb{E}[f(x^k)]\}_{k\in\mathcal{K}}$ subsequentially converges to the primal optimal value.
- Both $\{\hat{w}^k\}_{k\in\mathcal{K}}$ and $\{w^{k+1/2}\}_{k\in\mathcal{K}}$ converge to a dual optimal solution $w^\star$ with linear rate: there exists $q \in (0,1)$, such that for all sufficiently large k,

$$\|\hat{w}^k - w^\star\| \leq cq^k, \quad \|w^{k+1/2} - w^\star\| \leq cq^k.$$

Serious steps: convergence

Theorem

- Any limit point of $\{x^k\}_{k\in\mathcal{K}}$ is primal optimal, such that $\{\mathbb{E}[f(x^k)]\}_{k\in\mathcal{K}}$ subsequentially converges to the primal optimal value.
- Both $\{\hat{w}^k\}_{k\in\mathcal{K}}$ and $\{w^{k+1/2}\}_{k\in\mathcal{K}}$ converge to a dual optimal solution $w^\star$ with linear rate: there exists $q\in(0,1)$, such that for all sufficiently large k,

$$\|\hat{w}^k - w^\star\| \leq cq^k, \quad \|w^{k+1/2} - w^\star\| \leq cq^k.$$

- $\{h(\hat{w}^k)\}_{k\in\mathcal{K}}$ converges to the dual optimal value $h^\star$ with linear rate: there exists $q\in(0,1)$, such that for all sufficiently large k,

$$h(\hat{w}^{k+1}) - h^\star \leq q(h(\hat{w}^k) - h^\star)$$

Tail of null steps

Theorem

If there is **a last serious step** $\hat{w}$ at iteration $\hat{k}$, the stepsizes stabilize and a CQ holds, then

- Any limit point of $\{x^k\}_{k\in\mathcal{K}}$ is primal optimal, such that $\{\mathbb{E}[f(x)]\}_{k\in\mathcal{K}}$ subsequentially converges to the primal optimal value.

Tail of null steps

Theorem

If there is **a last serious step** $\hat{w}$ at iteration $\hat{k}$, the stepsizes stabilize and a CQ holds, then

- Any limit point of $\{x^k\}_{k\in\mathcal{K}}$ is primal optimal, such that $\{\mathbb{E}[f(x)]\}_{k\in\mathcal{K}}$ subsequentially converges to the primal optimal value.
- Both $\{w^k\}_{k\in\mathcal{K}}$ and $\{w^{k+1/2}\}_{k\in\mathcal{K}}$ converge to some $w^* \in \mathcal{N}^\perp$.

Tail of null steps

Theorem

If there is **a last serious step** $\hat{w}$ at iteration $\hat{k}$, the stepsizes stabilize and a CQ holds, then

- Any limit point of $\{x^k\}_{k\in\mathcal{K}}$ is primal optimal, such that $\{\mathbb{E}[f(x)]\}_{k\in\mathcal{K}}$ subsequentially converges to the primal optimal value.
- Both $\{w^k\}_{k\in\mathcal{K}}$ and $\{w^{k+1/2}\}_{k\in\mathcal{K}}$ converge to some $w^* \in \mathcal{N}^\perp$.
- Furthermore, w^* coincides with the last serious step $\hat{w}$. It is also its own proximal-step for $h + i_{\mathcal{N}^\perp}$:

$$w^* = \mathrm{prox}_{\frac{1}{t^*}}(h + i_{\mathcal{N}^\perp})(w^*) \iff w^* \text{ is critical (minimizer).}$$

Contributions

- The BPHA proposes a dual-based approach to solve stochastic optimization problems separately for different scenarios, using a model to approximate the dual function.
- The quality of the approximation is measured by testing descent.
- The price to pay is an extra evaluation of the dual function per iteration.
- In exchange, we gain:
    - Variable stepsizes.
    - Linear rate of convergence.
    - Stopping test.

Future research

- Nonconvex setting: the current approach provides lower bounds (duality gap) $\rightarrow$ use generalized Augmented Lagrangians (sharp).
- Inexact Bundle Progressive Hedging: allow inexact subproblem solutions.
- Numerical experiments: how do we tune the parameters?

QUESTIONS?

F. Atenas, C. Sagastizábal (2022). "A bundle-like approach to induce monotonicity in the progressive hedging algorithm". Working paper.

F. Atenas, C. Sagastizábal, P. J. Silva, and M. Solodov (2021). "A unified analysis of descent sequences in weakly convex optimization, including convergence rates for bundle methods". Submitted.

References

- W. Hare and C. Sagastizábal (2010). "A redistributed proximal bundle method for nonconvex optimization". SIAM Journal on Optimization, 20(5), 2442-2473.
- W. Hare and C. Sagastizábal (2009). "Computing proximal points of nonconvex functions". Mathematical Programming, 116(1), 221-258.
- Rockafellar, R. (2018). "Solving stochastic programming problems with risk measures by progressive hedging". In: Set-Valued and Variational Analysis 26.4, pp. 759–768.
- Rockafellar, R. and R. J.-B. Wets (1991). "Scenarios and policy aggregation in optimization under uncertainty". In: Mathematics of operations research 16.1, pp. 119–147.

Solve separate primal QPs

For each scenario $s = 1, \ldots, S$, find

$$\alpha_s^k = \operatorname*{arg\,min}_{\alpha_s \in \Delta^{B_s^k}} \left\{ (F_s^k)^\top \alpha_s + (\hat{w}_s^k - t_k x^k)^\top X_s^k \alpha_s + \frac{r_k}{2} \alpha_s (X_s^k)^\top X_s^k \alpha_s \right\}$$

where B_s^k is the simplex associated with the set of indices B_s^k.

For each scenario $s = 1, \ldots, S$, define

$$x_s^{k+1/2} = \sum_{j \in B_s^k} \alpha_{s,j}^k x_s^j$$

Dual reformulation

Solve separate primal QPs

For each $s = 1, \ldots, S$, the intermediate dual points satisfy

$$w_s^{k+1/2} = \arg\min_{w_s} \left\{ h_s^k(w_s) + (x^k)^\top w_s + \frac{1}{2r_k}|w_s - \hat{w}_s^k|^2 \right\},$$

Dual reformulation

Solve separate primal QPs

For each $s = 1, \ldots, S$, the intermediate dual points satisfy

$$w_s^{k+1/2} = \arg\min_{w_s} \left\{ h_s^k(w_s) + (x^k)^\top w_s + \frac{1}{2r_k} |w_s - \hat{w}_s^k|^2 \right\},$$

where

$$h_s^k(w_s) = \max_{j \in B_s^k} \left\{ -f_s \left(x_s^{j-1/2} \right) - (x_s^j)^\top w_s \right\}.$$

is the lower convex model of h_s.

Dual reformulation

Solve separate primal QPs

For each $s = 1, \ldots, S$, the intermediate dual points satisfy

$$w_s^{k+1/2} = \arg\min_{w_s} \left\{ h_s^k(w_s) + (x^k)^\top w_s + \frac{1}{2r_k} |w_s - \hat{w}_s^k|^2 \right\},$$

and α_s^k corresponds to the Lagrangian multipliers of

$$\begin{cases} \min\limits_{w_s, r_s} & r_s + (x^k)^\top w_s + \dfrac{1}{2r_k} |w - \hat{w}_s^k|^2 \\ \text{s.t.} & r_s \geq h_s^k(w_s), \quad j \in B_s^k \end{cases}$$

Descent test

If

$$\sum_{s=1}^{S} p_s[h_s(w_s^{k+1}) - h_s(\hat{w}_s^k)] \leq m \sum_{s=1}^{S} p_s[h_s^k(w_s^{k+1/2}) + (x^k)^\top w_s^{k+1/2} - h_s(\hat{w}_s^k)],$$

then set $\hat{w}^{k+1} = w^{k+1}$, and define $r_{k+1} \geq r_{\text{min}}$. *(serious step).*

Otherwise, set $\hat{w}^{k+1} = \hat{w}^k$, and define $r_{k+1} \in [r_{\text{min}}, r_k]$. *(null step).*

Scenario subproblems

For each scenario:

BPHA

$$\alpha^k = \arg\min_{\alpha \in \Delta^k} \left\{ \sum_{j \in B^k} \alpha_j f(x^j, y^j) + (\hat{w}^k - t_k x^k)^\top X^k \alpha + \frac{r_k}{2} \alpha (X^k)^\top X^k \alpha \right\}$$

$$x^{k+1/2} = \sum_{j \in B^k} \alpha^k x^{j-1/2}$$

PHA

$$x^{k+1/2} = \arg\min \left\{ f(x,y) + (w^k)^\top x + \frac{r}{2} |x - x^k|^2 \right\}$$