Linear Convergence in Bundle Progressive Hedging and its link to Proximal Decomposition

XVIIth Conference on Stochastic Programming

Théo Molfessis, École Polytechnique, Palaiseau, France

Joint work with F. Atenas, C. Sagastizábal and M. Solodov
at IMPA, Rio de Janeiro, Brazil

The multizonal optimization problem

Based on P. Mahey, J. Koko, and A. Lenoir, Decomposition methods for a spatial model for long-term energy pricing problem, in Math Meth Oper Res (2017) 85:137–153

$$\min_{(q,f)} \sum_{z\in Z} G_z(\boldsymbol{q}_z) + \sum_{e\in E} L_e(\boldsymbol{f}_e)$$
$$B\boldsymbol{q}_t + A\boldsymbol{f}_t = \boldsymbol{d}_t, \quad t = 1, \ldots, T-1$$
$$x_{z,t+1} = x_{zt} - u_{zt} + i_{zt} \quad \forall z, t$$
$$q_{z,t} \in \mathcal{P}_{z,t},\ f_{e,t} \in F_{e,t} \quad \forall z, t, e$$

where :

usage of storage

production

storage

$$\mathcal{P}_{z,t} = \left\{0 \le p_{z,t} \le P^{max}_{z,t},\ 0 \le u_{z,t} \le U^{max}_{z,t},\ X^{min}_{z,t} \le x_{z,t} \le X^{max}_{z,t}\right\}$$
$$F_{e,t} = \left\{0 \le f_{e,t} \le \kappa_{et}\right\}.$$

flow

Dissociating in and outgoing flows allows to write

$$\min_{q,f,\phi} \sum_{z} Q_z(\boldsymbol{f}_z, \boldsymbol{\phi}_z) \quad (\boldsymbol{f}, \boldsymbol{\phi}) \in \mathcal{A} \coloneqq \{(f,\phi)\ ,\ \phi_{ez,t} = f_{ez',t}, \quad \text{for} \quad e = (z', z) \in z'^{+} \cap z^{-}, \quad \forall z, z' \in Z, \forall t \ \}$$

More generally : multistage stochastic problems

Based on F. Atenas's PhD thesis: Proximal decomposition methods for optimization problems with structure, State University of Campinas, SP, Brazil – September 2023.

$$\begin{cases} \min_{x} & f(x) = \sum_{s=1}^{S} p_s f_s(x_s) \\ \text{s.t.} & x_s \in C_s, \quad \text{for all } s = 1, \dots, S, \\ & x \in A \end{cases}$$

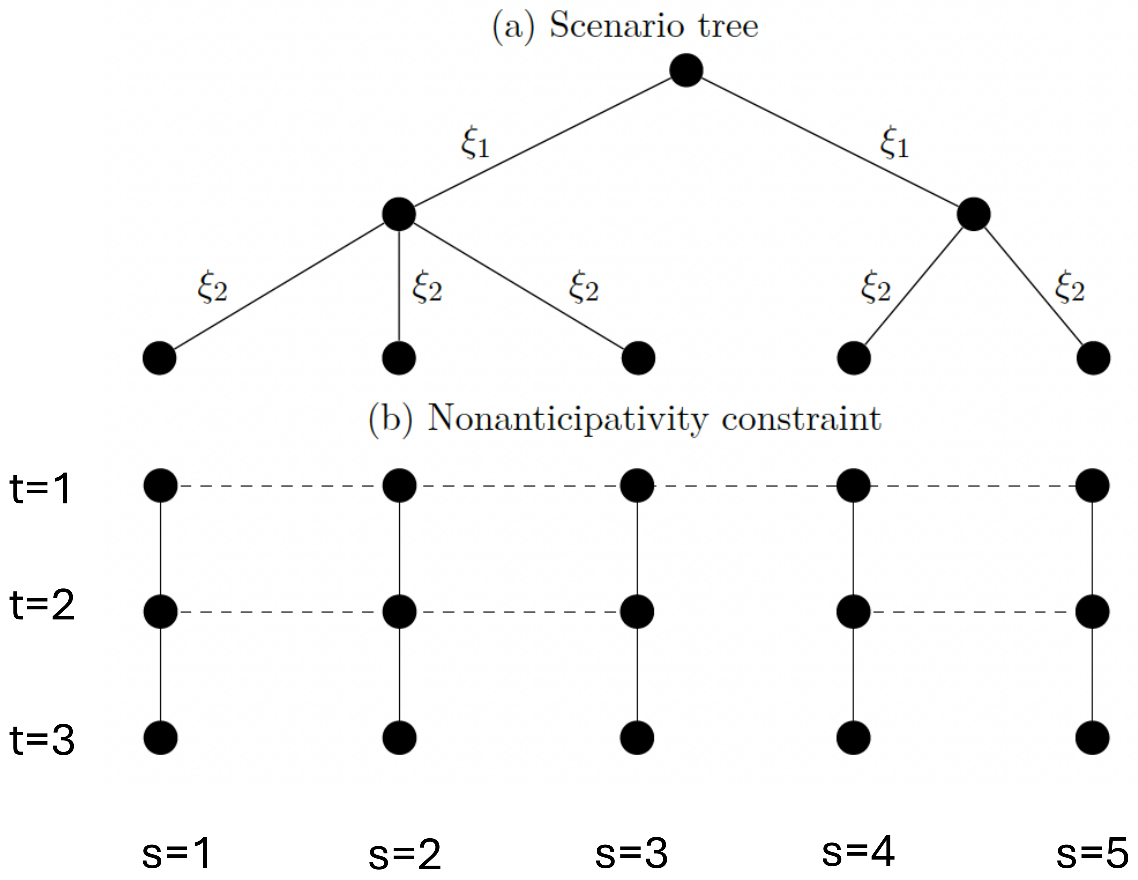
convex compact

subspace (multistage), easy projection

$$P_A\,[x]_t = \mathbb{E}_{[t-1]}[x_t].$$

$$\begin{cases} \min_{w} & h(w) \\ \text{s.t.} & w \in A^{\perp} \end{cases} \quad \text{with} \quad \langle u, v \rangle_S = \sum_{s=1}^{S} p_s u_s \cdot v_s\,.$$

$$h(w) = \sum_{s=1}^{S} p_s h_s(w_s)\,, \ \text{where } h_s(w_s) = \max_{x_s \in C_s} (-\mathcal{L}_s)(x_s, w_s)\,.$$



Progressive Hedging

Introduced in Rockafellar, R. T., and Wets, R. J-B, "Scenarios and policy aggregation in optimization under uncertainty." Mathematics of Operations Research 16 (1991), 119–1473

fixed !

Proximal step : For all s,

$$x_s^{k+1/2} = \text{argmin}_{x_s \in \mathbb{R}^n} f_s(x_s) + x_s \cdot w_s^k + \frac{t}{2}\|x_s - x_s^k\|^2$$

and

$$w^{k+1/2} = w^k + t(x^{k+1/2} - x^k).$$

Projection step: $x_{k+1} = P_{A,S}(x^{k+1/2})$, $w_{k+1} = P_{A^\perp,S}(w^{k+1/2})$ and $k \leftarrow k+1$.

blind ?

Proximal Decomposition

Introduced in Mahey P, Oualibouch S, Pham DT (1995) Proximal decomposition on the graph of a maximal monotone operator. SIAM J Optim 5:454–466

Proximal step : For all s,

$$x_s^{k+1/2} = \text{argmin}_{x_s \in \mathbb{R}^n} f_s(x_s) - x_s \cdot w_s^k + \frac{t}{2}\|x_s - x_s^k\|^2$$

and

$$w^{k+1/2} = w^k + t(x^{k+1/2} - x^k).$$

Projection step: $x_{k+1} = P_{A,\cancel{S}}(x^{k+1/2})$, $w_{k+1} = P_{A^\perp,\cancel{S}}(w^{k+1/2})$ and $k \leftarrow k+1$.

➡ When A is a subspace and distribution is uniform, PD and PH are the same !

Bundle methods

Based on *Correa, R., Lemaréchal, C. Convergence of some algorithms for convex minimization. Mathematical Programming 62, 261–275 (1993)*

Goal : Minimize a convex and finite-valued function f over the whole of X

Having only a Black Box : given $x \in X$, the value $f(x)$ and some $g = g(x) \in \partial f(x)$

Step 1 : Choose a convex function $\varphi^k : \mathcal{H} \longrightarrow \mathbb{R}$, and compute

$$y^{k+1} = \text{argmin}_y \varphi^k(y) + \frac{1}{2t_k}||y - x^k||^2.$$

Step 2 : Compute $f(y^{k+1})$. If a good decrease if obtained, namely if

$$f(x^k) - f(y^{k+1}) \geq m[f(x^k) - \varphi^k(y^{k+1})], \quad m \in (0,1)$$

then set $x^{k+1} = y^{k+1}$. Otherwise, $x^{k+1} = x^k$.

Bundle Progressive Hedging

Based on F. Atenas's PhD thesis: Proximal decomposition methods for optimization problems with structure, State University of Campinas, SP, Brazil – September 2023.

Algorithm 3: Bundle Progressive Hedging

Initialization : $k = 0, (x_0, w_0) \in A \times A^{\perp}, 0 < t_{min} \leq t_0 \leq t_{max}, m, l \in]0,1[$, and $TOL > 0$.

Proximal step : For all s,

(3.9)
$$w_s^{k+1/2} = \text{argmin}_{w_s \in \mathbb{R}^n} h_s(w_s) + w_s \cdot x_s^k + \frac{2}{t_k}|w_s - w_s^k|^2$$

and

(3.10)
$$x^{k+1/2} = x^k + \frac{1}{t_k}(w^{k+1/2} - w^k).$$

Stopping test : Compute the model descent :

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

If $\delta_{mod}^k < TOL$: return (x^k, w^k).

Projection step: $x_{k+1} = P_{A,S}(x^{k+1/2})$, $u_{k+1} = P_{A^{\perp},S}(w^{k+1/2})$.

Descent test : Compute the real descent :

$$\delta_{ex}^k = h(w^k) - h(u^{k+1}).$$

If $\delta_{ex}^k \geq m \cdot \delta_{mod}^k$: we have a serious step, $w^{k+1} = u^{k+1}$, and $t_{k+1} = \max(t_{min}, \min(t_{max}, t_k/l))$.

Else : we have a null step, $w^{k+1} = w^k$, and $t_{k+1} = \max(t_{min}, \min(t_{max}, l \cdot t_k))$.

$k \leftarrow k + 1$

(3.9) and (3.10) can also be written

$$x_s^{k+1/2} = argmin_{x_s} f_s(x_s) + w_s^k x_s + \frac{t_k}{2}||x_s - x_s^k||^2$$

$$w^{k+1/2} = w^k + t_k(x^{k+1/2} - x^k).$$

Same can be done for Proximal Decomposition

Step size can be changed !

Cost: One more dual functional evaluation to be made

Convergence analysis of BPH :previous results

Based on F. Atenas's PhD thesis: Proximal decomposition methods for optimization problems with structure, State University of Campinas, SP, Brazil – September 2023.

If finite termination (and $TOL = 0$) : last primal and dual iterates are optimal

set of dual solutions

If infinitely many serious steps + error bound : $d(w, S) \leq \alpha ||x||$ for $x \in \partial h(w) \cap B(0, \epsilon)$

- dual functional values at serious steps converge towards optimum monotonically, linearly
- dual serious (intermediate) iterates converge towards optimum linearly
- primal serious (intermediate) iterates converge subsequentially towards optimum

If infinite tail of null steps : (with stabilizing step size)

- last serious dual iterate W is optimal (and $u^k, w^{k+\frac{1}{2}}$ converge towards it)
- primal (intermediate) iterates converge subsequentially towards optimum … can we say more?

Convergence analysis of BPH : strongly convex function

Projected gradient (PG) algorithm : $y^{k+1} = P_A(y^k - \alpha_k \nabla f(y^k))$

Last serious dual iterate, optimal

Here primal iterates are given by PG on the Moreau envelope of $\phi(x) = f(x) + \langle W, x \rangle$

set of primal solutions

Projection error bound (PEB): $d(x, \bar{S}) \leq l||x - P_A(x - \frac{1}{t}\nabla f(x))||$

Theorem[1]: a function is strongly convex iff its Moreau envelope is strongly convex

Proposition[2]: PEB holds for strongly convex function

Theorem[3]: if PEB holds , then iterates of PG converge R-linearly

1 : *Theorem 2.1 from Zhi-Quan Luo and Paul Tseng. Error bounds and convergence analysis of feasible descent methods: a general approach. Annals of Operations Research, 46(1):157–178, 1993.*

2 : *Proposition 2.19 from C. Planiden and X. Wang. Strongly convex functions, moreau envelopes, and the generic nature of convex functions with strong minimizers. SIAM Journal on Optimization, 26(2):1341–1364, 2016.*

3 : *Theorem 5.3 from Zhi-Quan Luo and Paul Tseng. Error bounds and convergence analysis of feasible descent methods: a general approach. Annals of Operations Research, 46(1):157–178, 1993*

Convergence analysis : general case

We show that if PEB holds and the zeroes of f are isolated, then

- primal iterates and intermediate iterates converge towards an optimum
- primal iterates converge Q-linearly

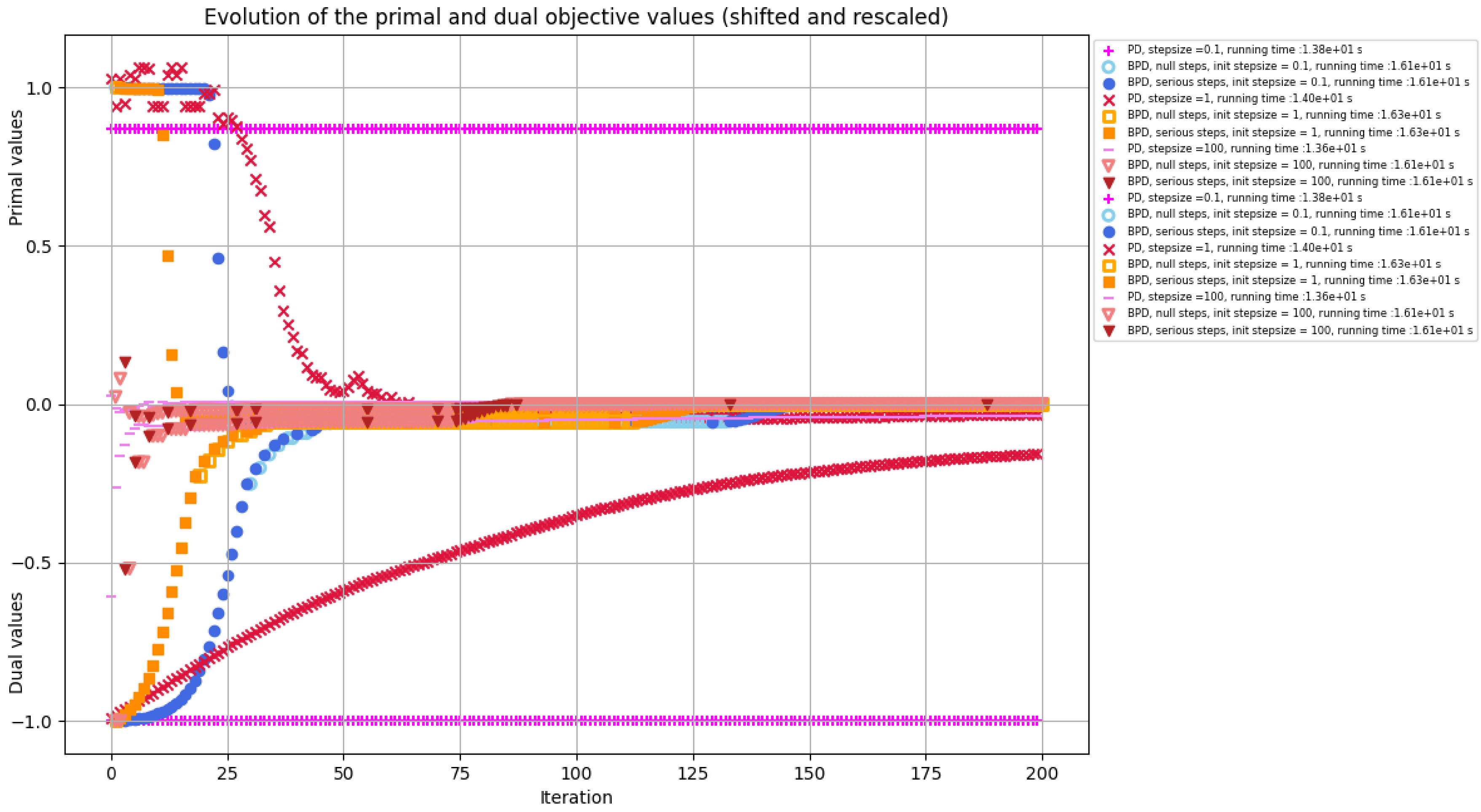
What does it mean ?

We found the dual optimal... but we don't know it yet! (stopping test is a sufficient condition for optimality, not a necessary one)

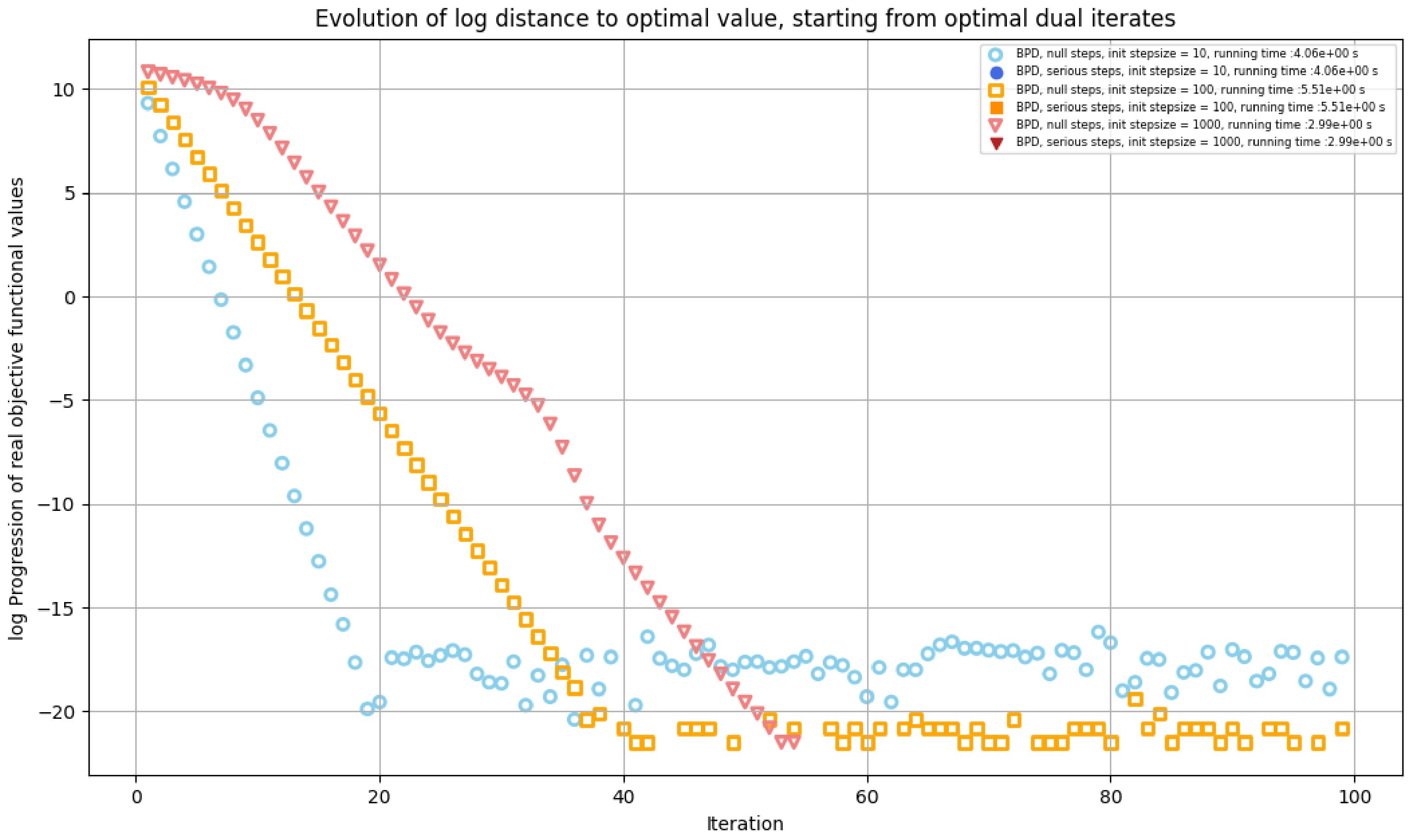
We proved that it takes a "linear time" to realize it !

⚠ This is a convergence rate analysis… not one of arithmetic complexity !

Simulations on the multizonal optimization problem



Simulations on the multizonal optimization problem (2)



Conclusion… what's next?

Extend the convergence results to non subspace constraint sets

Allowing to use a model of the objective function itself

Why? Extension to not so easy subproblems

How? With a *super-serious* descent test ? More soon!

Thank you for your time and attention 😌