Worst-Case Analysis of First-Order Optimization Methods involving Linear Operators

Nizar Bousselmi, Julien M. Hendrickx, François Glineur

ICTEAM, UCLouvain, Belgium

Goal: Obtaining guarantees on methods applied to $\min_x f(x) + g(Ax)$

- Automatic computation of worst-case performance of a method
- Better understanding, tuning of methods
- Selection of best method

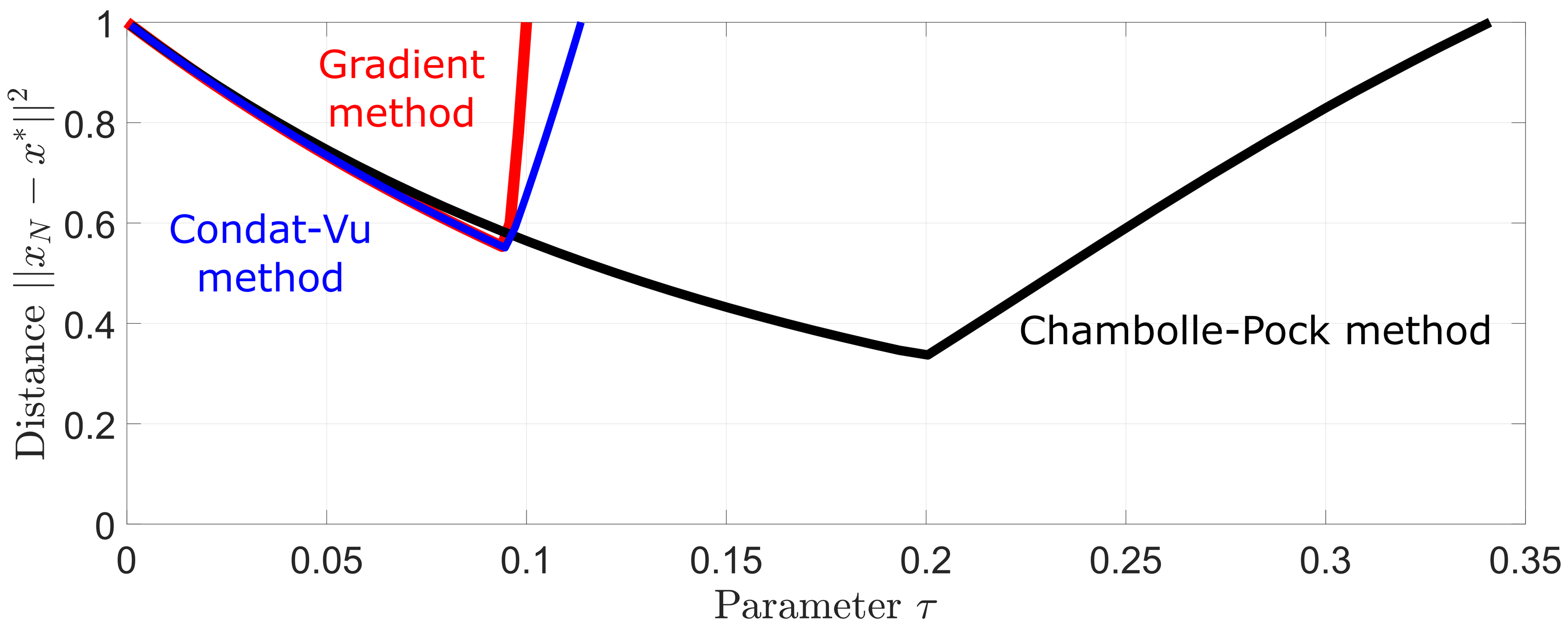

Fig. 1 Comparison by **PEP** of three methods (f and g smooth convex)

Methodology: Performance Estimation Problem

- Compute the worst-case performance as an optimization problem

(PEP)

$$\max_{x_0, x^*, f} \ \mathrm{Perf}(x_N, f) \quad \rightarrow \text{Performance measure (e.g. } ||x_N - x^*||^2\text{)}$$

$$f \in \mathcal{F} \quad \rightarrow \text{Function class (e.g. smooth convex functions)}$$

$$x_N = \mathcal{M}(x_0, f) \quad \rightarrow \text{Optimization method (e.g. Gradient method, Chambolle-Pock method)}$$

$$||\nabla f(x^*)||^2 = 0 \quad \rightarrow \text{Optimality}$$

$$||x_0 - x^*||^2 \leq 1 \quad \rightarrow \text{Bounded initial distance}$$

- **Exact convex reformulation** efficiently solvable
- Solution of (PEP) is the **worst-case performance** of **method** $\mathcal{M}$ on **function class** $\mathcal{F}$
- Available in Python library (PEPit) and Matlab toolbox (PESTO)

Main requirement: Interpolation conditions

- **Discretization** of (PEP):

Continuous function $f \longrightarrow$ Discrete variables f_i, g_i : $\begin{cases} f(x_i) &= f_i \\ \nabla f(x_i) &= g_i \end{cases}$

- **Interpolation conditions** guarantee existence of a function consistent with the variables x_i, g_i, f_i
- Example on L-smooth convex function $\mathcal{F}_{0,L}$:

Theorem: Given $(x_1, g_1, f_1), \dots, (x_N, g_N, f_N)$

$$\exists f \in \mathcal{F}_{0,L} : \begin{cases} f(x_i) &= f_i \\ \nabla f(x_i) &= g_i \end{cases} \Leftrightarrow f_i \geq f_j + g_j^T(x_i - x_j) + \tfrac{1}{2L}||g_j - g_i||^2 \quad \forall (i,j)$$

- Analyzing $\min_x f(x) + g(Ax)$ requires **interpolation conditions for linear operators**

Theorem: Given $(x_1, y_1), \dots, (x_N, y_N)$

$$\exists \mu I \preceq Q \preceq LI : y_i = Qx_i \Leftrightarrow \begin{cases} X^TY = Y^TX \\ (Y - \mu X)^T(LX - Y) \succeq 0 \end{cases}$$

where $X = (x_1 \cdots x_N)$ and $Y = (y_1 \cdots y_N)$

Applications: Rate of convergence

1) Optimal step size selection of **Gradient method**:

$$x_{k+1} = x_k - \tau \nabla f(x_k)$$

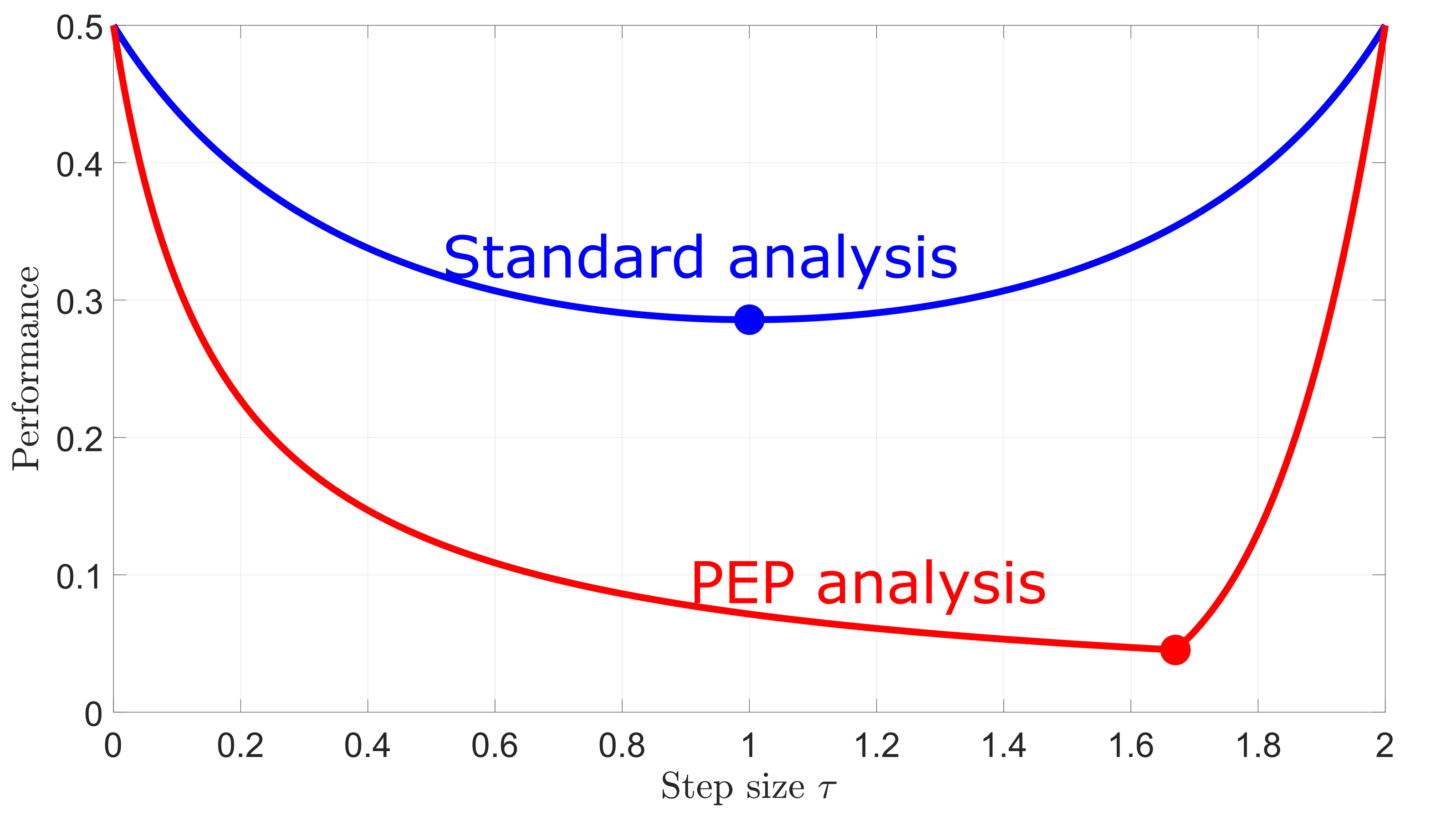

Fig. 2 Worst-case performance of 3 iterations of Gradient method on smooth convex functions: Standard analysis (blue line) and PEP analysis (red line)

- PEP provides tight worst-case performance
- Not tight standard analysis led to sub-optimal step size

2) Direct analysis of different performance measures of the **Chambolle-Pock method**:

$$\begin{cases} x_{k+1} &= \mathrm{prox}_{\tau f(.)}(x_k - \tau A^T u_k) \\ u_{k+1} &= \mathrm{prox}_{\sigma g^*(.)}(u_k + \sigma A(2x_{k+1} - x_k)) \end{cases}$$

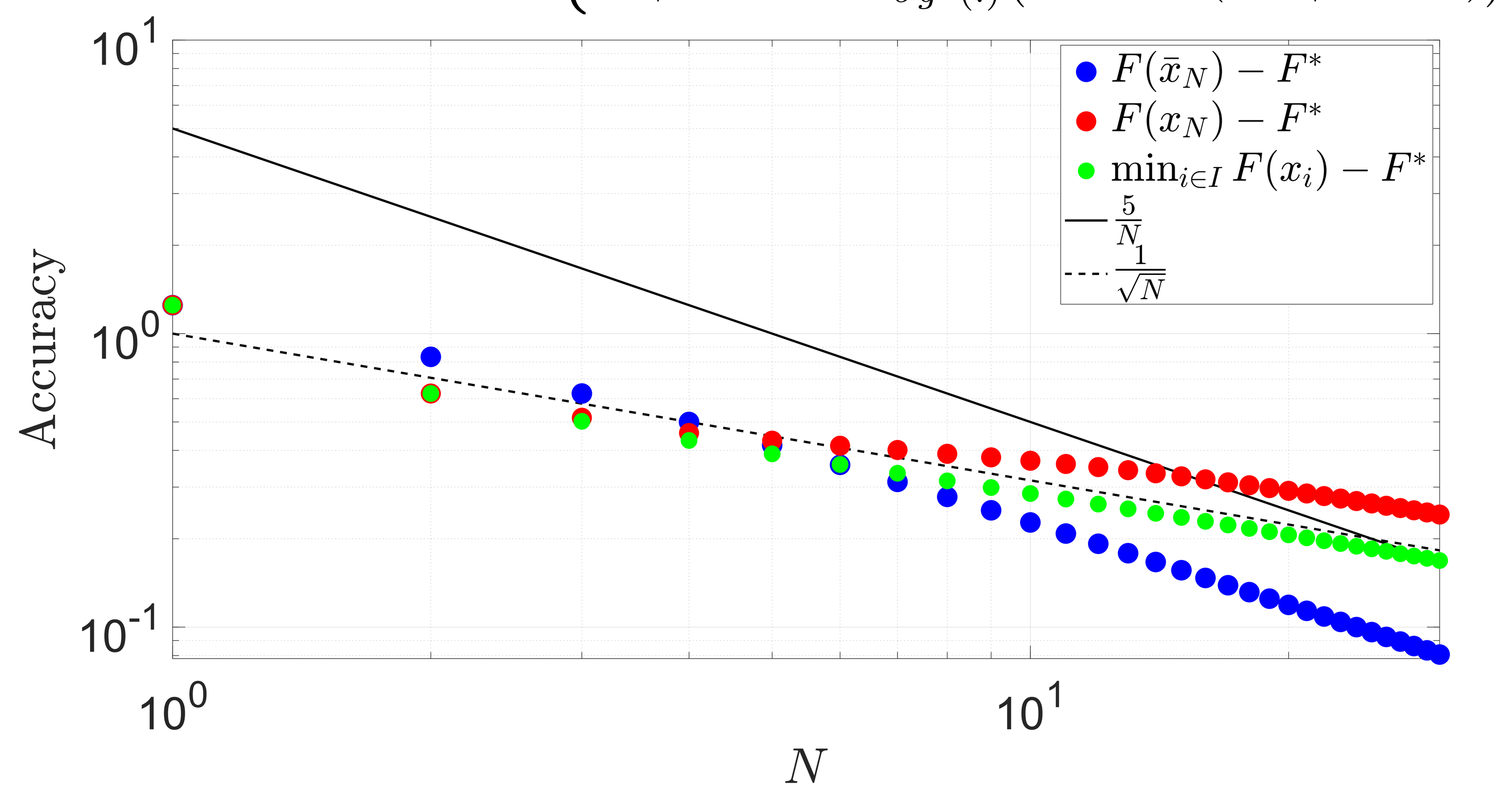

Fig. 3 Worst-case performance of Chambolle-Pock method in terms of number of iterations N for different performance measures: value accuracy of the average (blue dots), last (red dots) and best (green dots) iterations

- PEP easily unifies the analysis of different frameworks

3) Other methods with linear operators

- Alternating Direction Method of Multipliers (ADMM)
- Condat-Vu method
- Primal-Dual Three-Operator Splitting
- Proximal Alternating Predictor-Corrector

References

N. Bousselmi, J. M. Hendrickx, and F. Glineur, *"Interpolation conditions for linear operators and applications to performance estimation problems"* arXiv preprint arXiv:2302.08781, 2023.