A scaled ε–subgradient method

Alessandro Benfenati[‡], Silvia Bonettini[†], Valeria Ruggiero[†]

[†]Department of Mathematics and Computer Science, University of Ferrara, [‡]Université Paris–Est LIGM

ε–Subgradient Method: a Scaled Version

The optimization problem reads as

$$\min_x f(x) + \Phi(x)$$

- $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convex, proper, l.s.c function (possible non–differentiable)
- Φ convex, proper, l.s.c function; $\mathrm{dom}(\Phi) \subset \mathrm{dom}(f)$

Scaled ε subgradient method: generalization of the Forward Backward algorithm

$$x^{k+1} = \mathrm{prox}_{\alpha_k \Phi, D_k^{-1}} \left(x^k - \alpha_k D_k u^k\right) \qquad (1)$$

α_k chosen as in the classical ε–subgradient method (e.g., constant stepsize, Ermoliev series).

Assumption:

$$\lim_{k\to\infty} \varepsilon_k = 0$$

- $u^k \in \partial_{\varepsilon_k} f(x^k)$ for some $\varepsilon_k \geq 0$
- α_k is a positive stepsize
- D_k is a symmetric positive definite matrix with bounded eigenvalues
- $\|y\|_{D^{-1}} = y^\top D^{-1} y$
- e.g. $\Phi = i_X$, with $X \subset \mathrm{dom}(f)$, $X \neq \emptyset$, closed, convex set

Convergence Results

Assume that both the ε–subgradients of f and Φ and the eigenvalues of D_k are bounded. Let set $f^* = \inf_{x\in\mathbb{R}^n} (f(x) + \Phi(x))$ and define X^* as the set of solutions; under the assumptions on (1)

$$\lim_{k\to\infty} \epsilon_k = 0, \quad \sum_{k=0}^{\infty} \alpha_k = \infty, \sum_{k=0}^{\infty} \alpha_k^2 < \infty, \quad \sum_{k=0}^{\infty} \epsilon_k \alpha_k < \infty \qquad (2)$$

one has

a) $\liminf_{k\to\infty} (f(x^k) + \Phi(x^k)) = f^*$;

b) If $\{x^k\}$ is bounded, there exists a limit point of it belonging to X^*;

c) If $X^* \neq \emptyset$ $\lim_{k\to\infty} x^k = x^* \in X^*$ and

$$\lim_{k\to\infty} (f(x^k) + \Phi(x^k)) = f^* = f(x^*)$$

d) If $X^* = \emptyset$, $\{x^k\}$ is unbounded.

The convergence rate is quite pessimistic $\left(\sim (\sum \alpha_k)^{-1}\right)$,but the numerical experience shows that the actual performance of the scaled method overcomes the non scaled version.

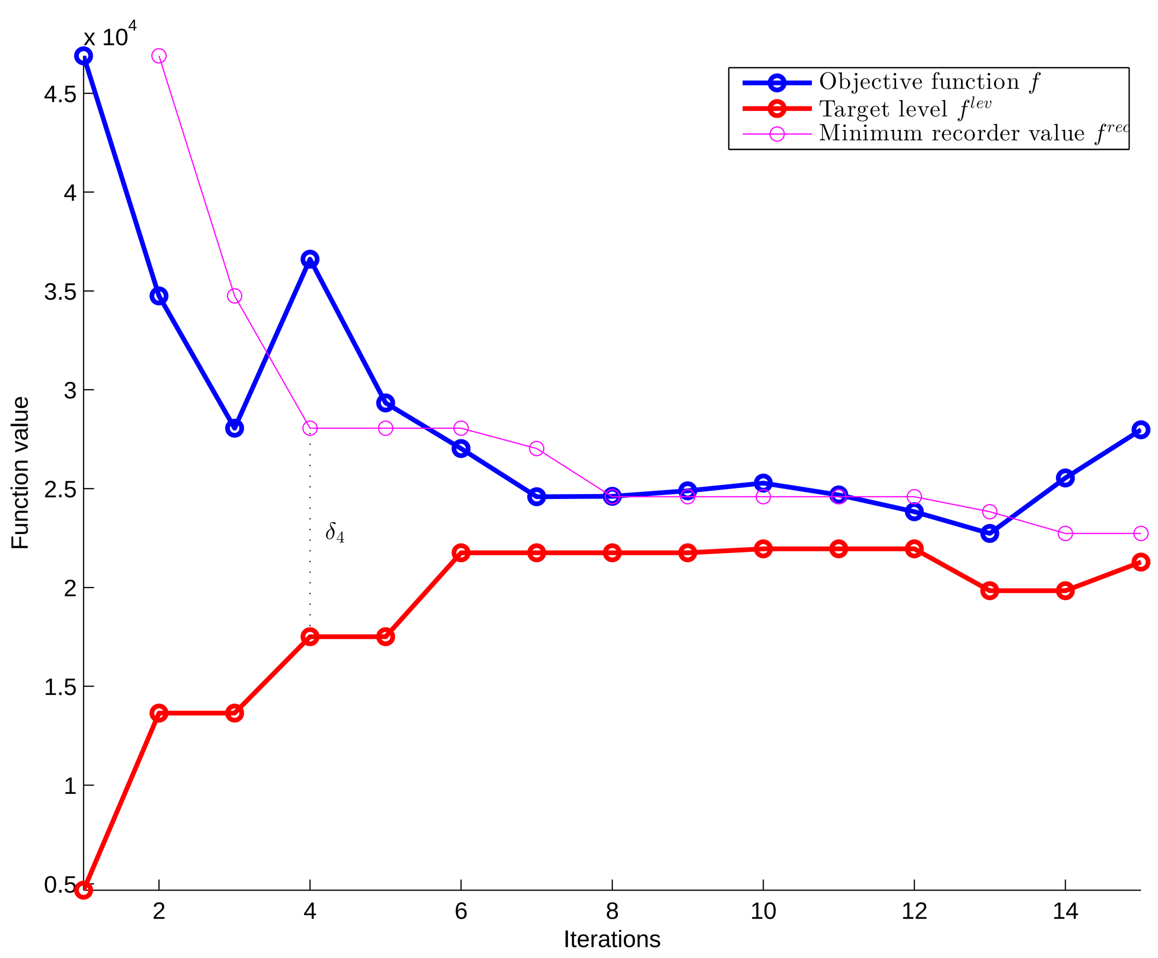
A choice for α_k

Dynamic rule

$$\alpha_k = \frac{f(x^k) - f_k}{\|u^k\|^2} \text{ or } \alpha_k = \frac{f(x^k) - f_k}{\max\{1, \|u^k\|^2\}}$$

Assumption: ε-subgradients of f and Φ bounded.

Inspired by the Polyak rule, f_k is an estimation of f^*: a *level algorithm* (Goffin 99) is employed to obtain such an estimation.



Application: Image Restoration with Poisson Noise

Let consider a blurred image affected by Poisson noise: the aim is to restore the image by solving

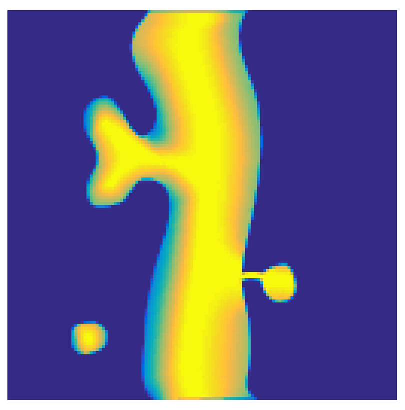
$$\min_x f(x) + \Phi(x) \equiv f_0(x) + f_1(Ax) + \Phi(x)$$

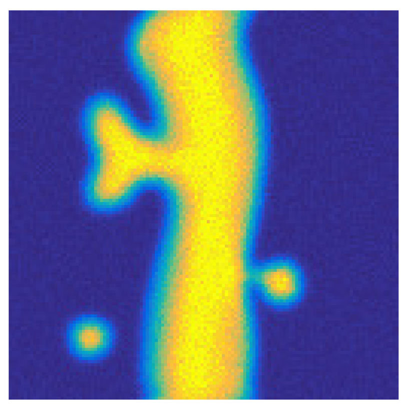
$f_0(x)$ is the generalized Kullback–Leibler divergence:

$$f_0(x) = \sum_{i=1}^{n} g_i \log \frac{g_i}{(Hx)_i + b} + (Hx)_i + b - g_i$$

- g is the blurred and noisy image;
- b is a constant background term;
- H is the linear blurring operator.

$$\Phi(x) = i_{\{x\in\mathbb{R}^n | x_i \geq 0\}}, \quad f_1(Ax) = \beta \sum_{i=1}^{n} \|A_i x\|, \; A_i \in \mathbb{R}^{2\times n} \text{ (Total Variation)}$$

x^* 

g 

Micro Test problem:

128×128, $\max(x^*) = 690$

$\frac{\|g - x^*\|}{\|x^*\|} = 0.1442$

$\beta = 0.0477$

The Scaled Primal Dual Hybrid Algorithm (SPDHG) reads as

$$\begin{aligned} y^{k+1} &= \mathrm{prox}_{\tau_k f_1^*, Id}(y^k + \tau_k A x^k) \\ u^k &= d^k + A^\top y^{k+1} \\ x^{k+1} &= \mathrm{prox}_{\alpha_k \Phi, D_k^{-1}}(x^k - \alpha_k D_k u^k) \end{aligned}$$

$$\begin{aligned} u^k &= U_k - V_k, \quad U_k \geq 0, V_k > 0 \\ D_k &= \min\left\{L_k^{-1}, \max\left\{x^k / V_k, L_k\right\}\right\} \\ L_k &= \sqrt{1 + \gamma_k}. \end{aligned}$$

Assume that $d^k = \nabla f(x^k)$, $A^\top y^{k+1} \in \partial_{\varepsilon_k} \Phi(x^k)$ and the eigenvalues of D_k are bounded.

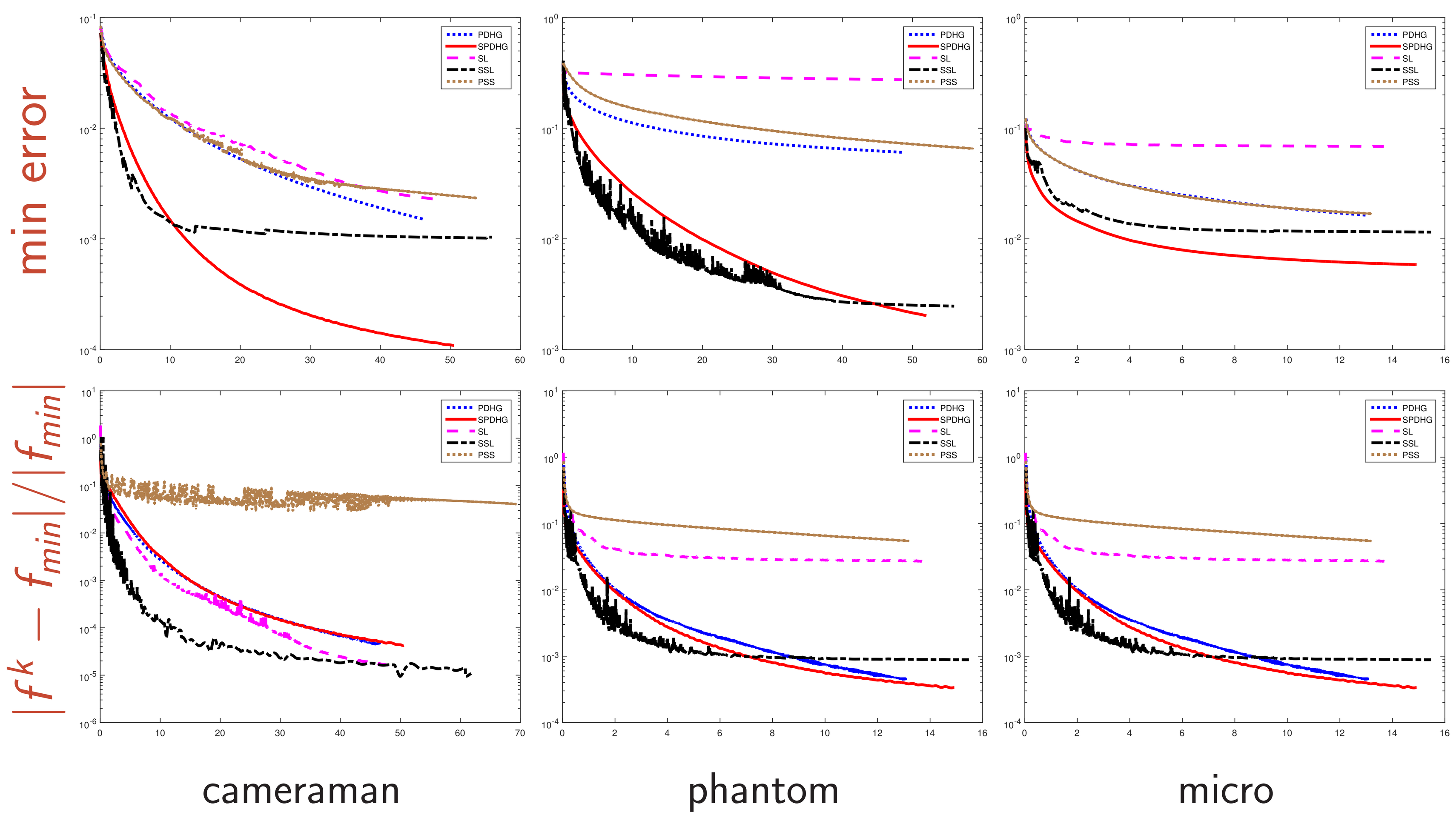
$$\alpha_k = \mathcal{O}\left(k^{-p}\right), \; \tau_k = \mathcal{O}(k^p), \; \gamma_k = \mathcal{O}\left(k^{-q}\right), \; \frac{1}{2} < p \leq 1, \quad q > 1$$

If $\mathrm{diam}(\mathrm{dom}(f_1^*)) < \infty$ then

$$\liminf_{k\to\infty} f(x^k) + \Phi(x^k) = f^*.$$

If the set of the solutions $X^* \neq \emptyset$, $\lim_{k\to\infty} x^k = x^* \in X^*$ and

$$\lim_{k\to\infty} f(x^k) + \Phi(x^k) = f^* = f(x^*).$$



x–axis: time for 3000 iterations.

All the technical details and complete references are available in
S. Bonettini, A. Benfenati, and V. Ruggiero, *Scaling Techniques for ε-Subgradient Methods*, SIAM Journal on Optimization 2016 26:3, 1741-1772