Sharpness, Restart, Acceleration

Vincent Roulet, Alexandre D'Aspremont

SIERRA Team - École Normale Supérieure/CNRS/INRIA/PSL Research University, 75005, Paris, France

Overview

For minimization problems of a convex function f

$$\text{minimize} \quad f(x), \qquad \text{Solutions} \quad x^* \in X_* \subset \mathbb{R}^d,$$

we present *optimal* and *adaptive* **restart** strategies of classical algorithms, under a generic description of the **sharpness** of f around its minimizers.

Sharpness

Sharpness of a function on a set $K \supset X^*$ can be described by $r \geq 1$, $\mu > 0$ s.t. for every $x \in K$,

$$\mu d(x, X^*)^r \leq f(x) - f^* \qquad \text{(Sharp)}$$

where $d(x, X^*)$ is the distance from x to X^*.

Examples :

- Gradient dominated convex functions ($r = 2$)
- Matrix game problems like $\min_x \max_y x^T Ay$ ($r = 1$)
- Real analytic functions ($r \geq 1$) (Łojasiewicz inequality)
- A.k.a. Hölderian error bound

Smoothness

Smoothness of f generally described on a set J by $s \in [1, 2]$, $L > 0$ s.t. for every $x, y \in J$

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2^{s-1}, \qquad \text{(Smooth)}$$

where $\nabla f(x)$ is any sub-gradient of f at x if $s = 1$.

Examples

- Classical assumption for non-smooth problems ($s = 1$)
- Smooth functions ($s = 2$)
- Hölder smooth functions ($s \in (1, 2)$)

Links btw sharpness and smoothness

(Sharp) lower bounds, (Smooth) upper bounds s.t. $s \leq r$.
Convergence depends then on

$$\kappa = L^{\frac{2}{s}}/\mu^{\frac{2}{r}}, \qquad \tau = 1 - s/r \in [0, 1[$$

Scheduled restarts for smooth problems

- Take the accelerated algorithm that starts from x_0 and outputs after t iterations $x = \mathcal{A}(x_0, t)$, s.t.

$$f(x) - f^* \leq 4Ld(x_0, X^*)^2/t^2,$$

- Schedule restarts at times t_k and build sequence

$$x_k = \mathcal{A}(x_{k-1}, t_k).$$

Optimal strategy

Assume f satisfies (Sharp) on a set $K \supset \{x : f(x) \leq f(x_0)\}$, schedule restarts at times

$$t_k = C_{\tau,\kappa} e^{\tau k},$$

then after R restarts and $N = \sum_{k=1}^R t_k$ total iterations,

$$f(x_R) - f^* = O\left(\exp(-\kappa^{-1/2} N)\right), \qquad \text{when } \tau = 0,$$
$$f(x_R) - f^* = O\left(1/N^{2/\tau}\right), \qquad \text{when } \tau > 0.$$

Universal scheduled restarts

- For general problems ($s \in [1, 2]$), use fast universal algorithm requires target ϵ to output $x = \mathcal{U}(x_0, \epsilon, t)$ s.t.

$$f(x) - f^* \leq \epsilon/2 + 4\epsilon^{1-2/s} L^{2/s} d(x_0, X^*)^2/t^{2\rho/s}$$

where $\rho = 3s/2 - 1$ is the optimal rate assuming (Smooth)

- Scheduled restarts take then the form

$$x_k = \mathcal{U}(x_{k-1}, \epsilon_k, t_k)$$

Optimal strategy

Assume f satisfies (Sharp) on a set $K \supset \{x : f(x) \leq f(x_0)\}$, given $\epsilon_0 \geq f(x_0) - f^*$, schedule restarts by

$$t_k = C_{\tau,\kappa,\rho,\epsilon_0} e^{\tau k}, \qquad \epsilon_k = e^{-\rho k} \epsilon_0$$

Then after R restarts and $N = \sum_{k=1}^R t_k$ total iterations, when $\tau = 0$,

$$f(x_R) - f^* = O\left(\exp(-\kappa^{-s/(2\rho)} N)\right) \qquad \text{when } \tau = 0$$
$$f(x_R) - f^* = O\left(1/N^{\rho/\tau}\right) \qquad \text{when } \tau > 0$$

Adaptive strategies

Parameters μ, r often unknown, adaptivity is crucial

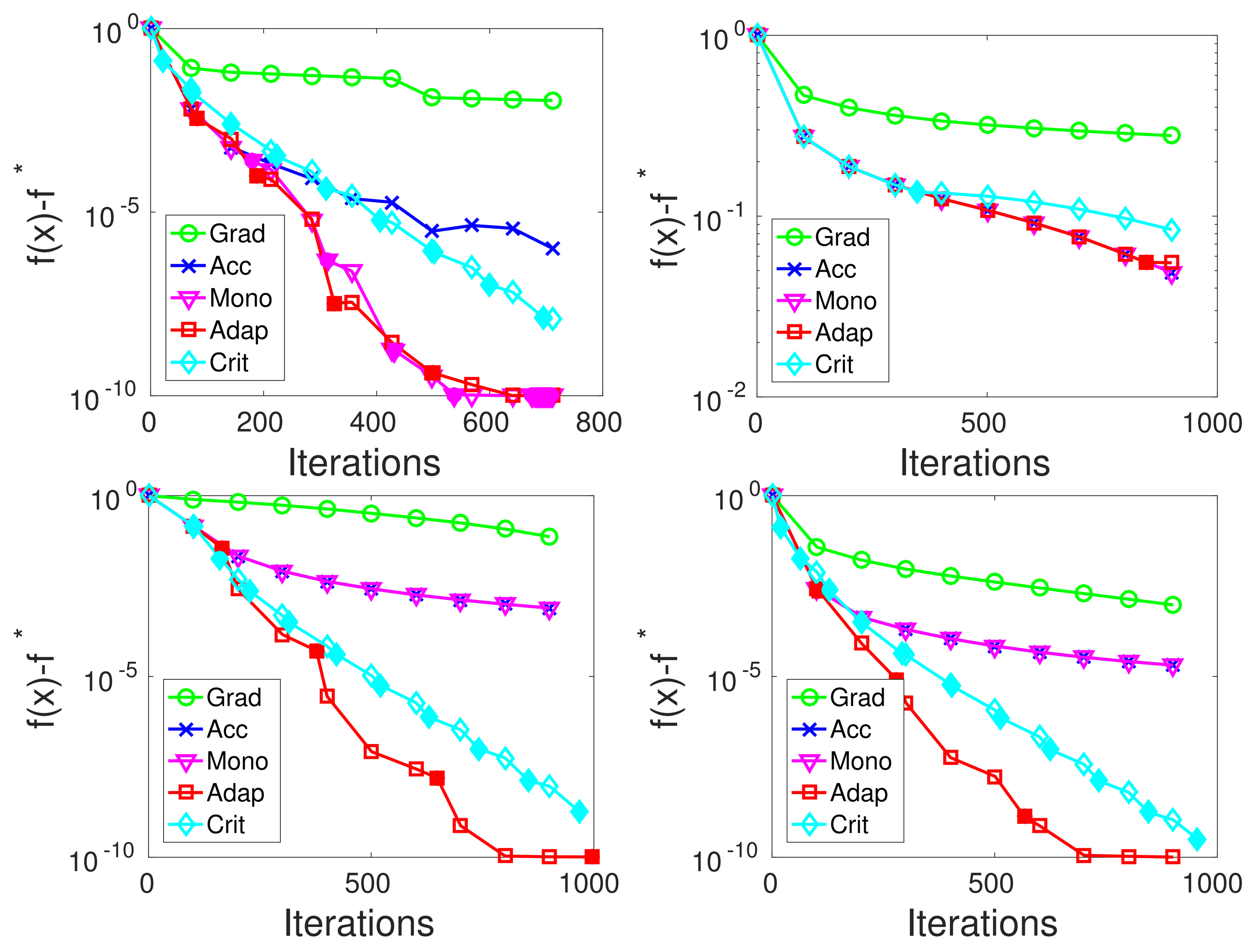
Log-scale grid search on possible schedules

- works for smooth pbs, fixed budget of iterations N
- suboptimal by a factor 4 at a cost of $\log_2(N)^2$ search

Restart with termination criterion if f^* is known

- restart the algorithm each time the gap has decreased by a constant factor
- gives optimal bounds for general convex problems

Numerical Experiments



From top to bottom and left to right : least squares, logistic, dual SVM and LASSO for classification on *Sonar* data set.

We use gradient descent (Grad), accelerated gradient (Acc), restart heuristic enforcing monotonicity (Mono), best schedule found by grid-search (Adap) and schedule with gap criterion (Crit). Large dots represent the restart iterations

Generalizations

Analysis relies only on convergence rates, generalizes then to composite problems and/or non-Euclidean settings.