The non-monotonocity effect of accelerated optimization methods

Marina Danilova Boris Polyak

Moscow Institute of Physics and Technology,
Department of Control and Applied Mathematics

Institute of Control Sciences RAS,
Laboratory of Adaptive and Robust Systems

June 28, 2018

Introduction

Accelerated first-order algorithms

- Conjugate gradient
- Heavy-ball
- Nesterov's accelerated gradient

Mathematical Challenges

- Accelerated methods converge **non-monotonically**.
- Asymptotic estimates obtained for them can give a **distorted representation** of the method behavior.
- **How to implement** the algorithm for real problems correctly?

Contents

Stating the problem

Optimization problem

$$\min_{x \in R^n} \; f(x)$$

- $x \in R^n, \; f(x) : R^n \to R, \; f(x) \in \mathscr{F}^{1,1}_{L,\mu}$
- $\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$
  $L > 0$ - Lipschitz constant
- $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2}\mu\|x - y\|^2$
  $\mu > 0$ - constant of strong convexity
- $\varkappa = \frac{L}{\mu}$ - condition number
- quadratic case $f(x) = \frac{1}{2}\langle Ax, x \rangle + \langle b, x \rangle, \; \mu I \preceq A \preceq LI$

Stating the problem

Heavy ball method (Polyak 1964)

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

α, β - method parameters

Known results

- local convergence: $\alpha \in \left(0, \frac{2(1+\beta)}{L}\right)$, $\beta \in [0, 1)$
- optimal parameters: $\alpha^* = \frac{1}{(\sqrt{L}+\sqrt{\mu})^2}$, $\beta^* = \left(\frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}\right)^2$
- best convergence rate: $q^* = \sqrt{\beta^*}$
- asymptotic estimate: $\|x_k - x^*\| = O(q^k)$

Non-monotonicity effect

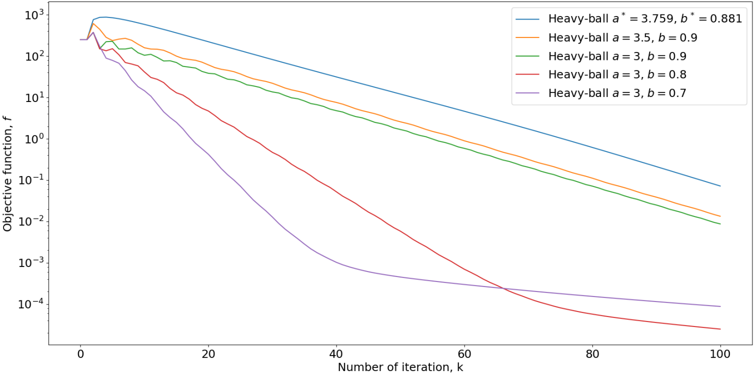

Figure 1: Convergence of Heavy-ball method with different estimates of $\alpha,\ \beta,\ A \in R^{n\times n},\ n = 10^3,\ \varkappa = 10^3$.

Aim and Objectives

Heavy-ball method

1. analysis of non-monotonic behavior
2. construction of the Lyapunov function
3. implementation of Heavy-ball method for the State estimation problem in power systems

Non-monotonic behavior

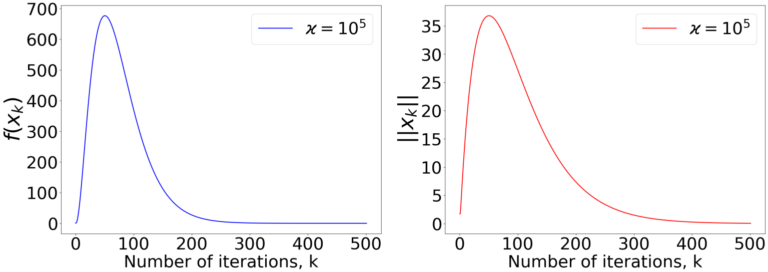

Figure 2: Dependence of $f(x_k)$ and $\|x_k\|$ on the number of iterations k.

"Peak effect"

- $A = \mathrm{diag}\left(1, 10^5\right),\ x_0 = x_1 = (1, 1)^T$
- α^*, β^* - optimal parameters

Non-monotonic behavior

Form of a linear difference equation

$$x_{k+1} = x_k\left((1+\beta)I - \alpha A\right) - \beta x_{k-1}$$

$$Ae_i = \lambda_i e_i$$

$$x^i_{k+1} = ax^i_k + bx^i_{k-1}$$

$$a = (1+\beta-\alpha\lambda_i), \quad b = -\beta$$

Proposition ("Peak effect")

Assume that $f(x) = \frac{1}{2}\left(Ax, x\right),\ \mu I \preceq A \preceq LI$,
where $\mu,\ L$ - strong convexity and Lipschitz constants. There are initial conditions $\|x_0\| \leq 1,\ \|x_1\| \leq 1,\ x_0, x_1 \in R^n$, which lead to a peak effect in Heavy-ball method with optimal parameters α^*, β^*:

$$\max_k ||x_k|| \geq \frac{\sqrt{\varkappa}}{2\mathrm{e}}.$$

Non-monotonic behavior

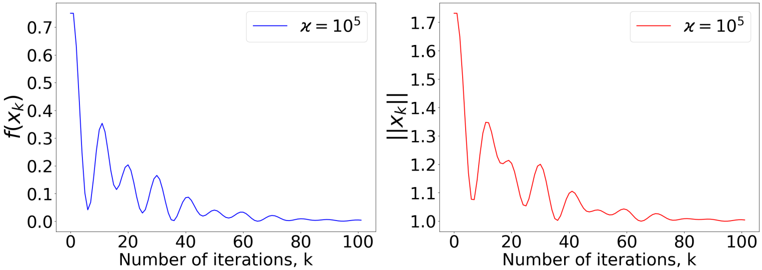

Figure 3: Dependence of $f(x_k)$ and $\|x_k\|$ on the number of iterations k.

- $A = \mathrm{diag}\left(1, 10^5\right), \; x_0 = x_1 = (1,1)^T$
- $\alpha \neq \alpha^*, \beta \neq \beta^*$ - non-optimal parameters

Heavy-ball method (Continuous case)

Continuous case

$$\ddot{x} + a\dot{x} + b\nabla f(x) = 0$$

- $\dot{x} = y$
  $\dot{y} = -ay - b\nabla f(x)$
- $a, b > 0$

Lyapunov function (total energy)

$$V(x, y) = f(x) + \frac{1}{2b}\|y\|^2, \quad \dot{V}(x, y) \leq 0$$

Upper bound

$$\|x(t) - x^*\| \leq \sqrt{\varkappa}\|x(0) - x^*\|$$

Construction of the Lyapunov function

Heavy-ball method (Discrete case)

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

$\alpha, \beta > 0$ - method parameters

Theorem (Lyapunov function)

Assume that $f \in \mathscr{F}_L^{1,1}$ and that $\alpha \in \left(0, \frac{1}{L}\right)$, $\beta \in \left[0, \sqrt{(1-\alpha L)}\right]$. Then for any initial conditions $x_0,\ x_1 \in R^n$ the following function

$$V(x_k) = f(x_k) + \frac{1-\alpha L}{2\alpha}\|x_k - x_{k-1}\|^2$$

is a Lyapunov function for the discrete case of the Heavy-ball method

$$V(x_k) \leq V(x_{k-1}).$$

Monotonic behavior

Lyapunov function

$$V(x_k) = f(x_k) + \frac{1-\alpha L}{2\alpha}\|x_k - x_{k-1}\|^2$$

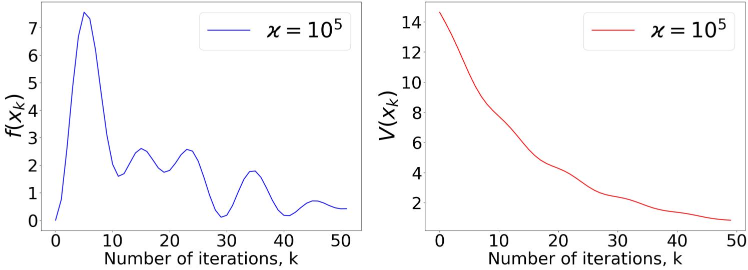

Figure 4: The behavior of objective function $f(x_k)$ and Lyapunov function $V(x_k)$ depending on the number of iterations k.

Results

Lyapunov function

$$V(x_k) = f(x_k) + \frac{1-\alpha L}{2\alpha}\|x_k - x_{k-1}\|^2$$

- The conditions for the parameters:

$$0 \ < \ \alpha \ < \ \frac{1}{L} \qquad 0 \ \leq \ \beta \ \leq \ \sqrt{1-\alpha L}$$

- it is not necessary to know the constant of strong convexity μ
- adaptive algorithm without knowledge of the Lipschitz constant L

Results

Theorem (global convergence)

Assume that $f \in \mathscr{F}^{1,1}_{\mu,L}$ и $0 \leq \mu \leq L$ and that
$\alpha \in \left(0, \frac{1}{L}\right)$, $\beta \in \left[0, \sqrt{(1-\alpha L)(1-\alpha\mu)}\right]$.
Then, the Heavy-ball method converges linearly for any initial conditions $x_0 = x_1 \in R^n$:

$$\|x_k - x^*\| \leq \sqrt{\varkappa} q^k \|x_0 - x^*\|,$$

where $q = (1 - \alpha\mu)$.

Practical part

State estimation of power system

Calculating an approximation for
the unknown state variables in the system obtained from **imperfect measurements**.

State variables

- V - voltage magnitude
- θ - voltage phase angle

State estimation

$$z = \begin{pmatrix} z_1 \\ ... \\ z_m \end{pmatrix} \text{ - measurements } (P, Q, V, \theta)$$

$$x = \begin{pmatrix} x_1 \\ ... \\ x_n \end{pmatrix} \text{ - state variables } (V, \theta)$$

$$h(x) = \begin{pmatrix} h_1(x_1, ...x_n) \\ ... \\ h_m(x_1, ...x_n) \end{pmatrix} \text{ - non-linear functions (nodal equations)}$$

The objective function to be minimized is

$$\min_x J(x) = \min_x \sum_{i=1}^{m} \frac{(z_i - h_i(x))^2}{\sigma_i^2}$$

σ^2 - i-th measurement variance .

Non-monotonicity effect

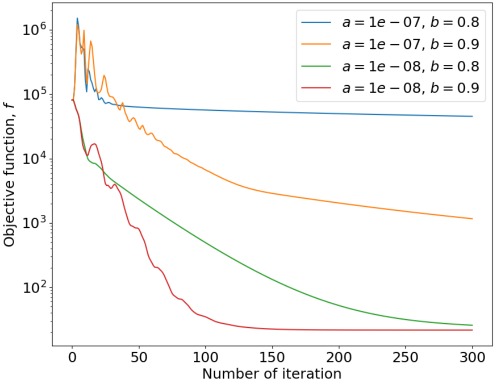

Figure 5: Convergence of HB method with different estimates of α, β for IEEE 14-Bus power system.

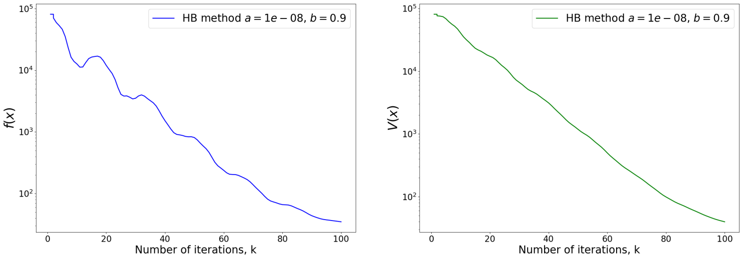

Figure 6: The behavior of objective function $f(x_k)$ and Lyapunov function $V(x_k)$ with parameters $\alpha = 1e - 08,\ \beta = 0.9$ for IEEE 14-Bus power system.

Future research

- Improving the Lyapunov function
- Developing an adaptive algorithm
- Considering the Nesterov's accelerated gradient method

References

- Boris Polyak, Marina Danilova and Anastasiya Kulakova. **Non-asymptotic Behavior of Multi-Step Iterative Methods.** *ICDEA 2018.*
- Boris Polyak. **Some methods of speeding up the convergence of iteration methods.** *USSR Computational Mathematics and Mathematical Physics*, 4 : 5 : 1 – 17, *1964.*
- Boris Polyak and Pavel Shcherbakov. **Lyapunov functions: An optimization theory perspective.** *IFAC-Papers OnLine*, 50.1 : 7456 – 7461, *2017.*
- Yurii Nesterov. **A method for unconstrained convex minimization problem with the rate of convergence $O(1/k2)$.** *Soviet Mathematical Doklady*, 27 : 372 – 376, *1983.*
- Pontus Giselsson and Stephen Boyd. **Monotonicity and restart in fast gradient methods.** *In Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on. 2014.*

Thank you for your attention!