Optimal Parameters for the Polyak's Momentum Algorithm on Strongly Convex and Smooth Functions

Introduction

We consider the Heavy Ball method with friction (HBF) for minimizing a μ-strongly convex and L-smooth function $f : \mathbb{R}^n \to \mathbb{R}$, where the method is described by the following iteration:

$$\begin{cases} x_0,\, x_1 \in \mathbb{R}^n, \\ x_{k+1} = x_k - \alpha \nabla f(x_k) + \beta(x_k - x_{k-1}), \end{cases}$$

where $\alpha > 0$ is the step size and $\beta \geq 0$ is the momentum parameter.

Our objective is to determine the optimal parameters α and β that minimize the convergence rate for minimizing a μ-strongly convex and L-smooth function.

Assumptions

- **μ-Strong Convexity**: For all $x, y \in \mathbb{R}^n$,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|y - x\|^2.$$

- **L-Smoothness**: For all $x, y \in \mathbb{R}^n$,

$$\|\nabla f(y) - \nabla f(x)\| \leq L\|y - x\|.$$

- **Condition Number**: $\kappa = \frac{L}{\mu} \geq 1$.

Spectral Radius Analysis for Optimal Parameters

Quadratic Case and Error Propagation

We begin by analyzing the Heavy Ball method applied to a quadratic function:

$$f(x) = \frac{1}{2}x^\top Q x - b^\top x + c,$$

where Q is a symmetric positive definite matrix with eigenvalues in the range $[\mu, L]$. The gradient is:

$$\nabla f(x) = Qx - b.$$

Let x^* be the minimizer of f, i.e., $x^* = Q^{-1}b$. Define the error $e_k = x_k - x^*$. The iteration becomes:

$$e_{k+1} = (I - \alpha Q)e_k + \beta(e_k - e_{k-1}).$$

We now express the error as a linear system:

$$\begin{pmatrix} e_{k+1} \\ e_k \end{pmatrix} = \begin{pmatrix} (I - \alpha Q + \beta I) & -\beta I \\ I & 0 \end{pmatrix} \begin{pmatrix} e_k \\ e_{k-1} \end{pmatrix}.$$

Let $z_k = \begin{pmatrix} e_k \\ e_{k-1} \end{pmatrix}$, and the iteration matrix becomes:

$$z_{k+1} = Mz_k,$$

where M is the iteration matrix:

$$M = \begin{pmatrix} (I - \alpha Q + \beta I) & -\beta I \\ I & 0 \end{pmatrix}.$$

Eigenvalue Analysis and Characteristic Equation

We now focus on the spectral properties of the iteration matrix M. For each eigenvalue λ_i of Q, the characteristic equation of M becomes:

$$\xi^2 - (1 - \alpha\lambda_i + \beta)\xi + \beta = 0.$$

The discriminant of this quadratic equation is:

$$\Delta = (1 - \alpha\lambda_i + \beta)^2 - 4\beta.$$

Setting the Discriminant to Zero for Equal Roots

To minimize the spectral radius of the iteration matrix, we set the discriminant to zero. This ensures that the two roots are equal, reducing the oscillations in the convergence behavior and ensuring optimality. Therefore, we set:

$$(1 - \alpha\lambda_i + \beta)^2 = 4\beta.$$

Solving this gives:

$$1 - \alpha\lambda_i + \beta = 2\sqrt{\beta},$$

or equivalently,

$$1 - \alpha\lambda_i = 2\sqrt{\beta} - \beta.$$

Optimal Parameters for $\lambda_{\min} = \mu$ and $\lambda_{\max} = L$

We apply this condition to both $\lambda_{\min} = \mu$ and $\lambda_{\max} = L$. By solving this system of equations, we find that the optimal parameters are:

$$\alpha = \frac{4}{(\sqrt{L} + \sqrt{\mu})^2}, \quad \beta = \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^2, \quad \text{where} \quad \kappa = \frac{L}{\mu}.$$

Verification of Spectral Radius Minimization

With these values of α and β, the spectral radius of the iteration matrix is minimized, ensuring the fastest possible convergence for the Heavy Ball method. The convergence rate is geometric with the rate:

$$\rho(M) = \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}.$$

Conclusion

By choosing the step size and momentum parameters as:

$$\alpha = \frac{4}{(\sqrt{L} + \sqrt{\mu})^2}, \quad \beta = \left(\frac{\sqrt{L} - \sqrt{\mu}}{\sqrt{L} + \sqrt{\mu}}\right)^2,$$

the Heavy Ball method achieves optimal convergence for minimizing a μ-strongly convex and L-smooth function, with the convergence rate given by the spectral radius:

$$\rho(M) = \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}.$$