Task 2: Adaptive Stepsizes – Solution

June 10, 2025

1 Task 2: Adaptive Stepsizes

1.1 Solution for Improving the Analysis for a Convex Quadratic Function

Task: Improve the convergence analysis of the methods presented in the paper for the specific case of a convex quadratic function of the form $f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$.

1.1.1 Key Properties of Quadratic Functions

For a convex quadratic function $f(x) = \frac{1}{2}x^T Ax - b^T x$, where $A \in \mathbb{R}^{n\times n}$ is a symmetric positive definite (SPD) matrix, we have the following crucial properties:

- The gradient is linear: $\nabla f(x) = Ax - b$.
- The Hessian (second derivative) is constant: $\nabla^2 f(x) = A$.
- The function is strongly convex. If μ and L are the smallest and largest eigenvalues of A respectively, then $f(x)$ is μ-strongly convex and its gradient is L-Lipschitz.
- The unique minimizer x^* satisfies $Ax^* = b$.

1.1.2 Simplification of Adaptive Step Size Rule

The adaptive step size λ_k in Algorithm 1 of the paper is defined as:

$$\lambda_k = \min\left\{\sqrt{1+\theta_{k-1}}\lambda_{k-1}, \frac{\|x_k - x_{k-1}\|}{2\|\nabla f(x_k) - \nabla f(x_{k-1})\|}\right\}$$

For a quadratic function, we can simplify the second term. Let $e_k = x_k - x^*$ be the error at iteration k. Then $\nabla f(x_k) = Ax_k - b = A(x_k - x^*) = Ae_k$. The difference in gradients can be expressed as:

$$\nabla f(x_k) - \nabla f(x_{k-1}) = (Ax_k - b) - (Ax_{k-1} - b) = A(x_k - x_{k-1})$$

Substituting this into the second term for λ_k:

$$\frac{\|x_k - x_{k-1}\|}{2\|\nabla f(x_k) - \nabla f(x_{k-1})\|} = \frac{\|x_k - x_{k-1}\|}{2\|A(x_k - x_{k-1})\|}$$

Let $v = x_k - x_{k-1}$. Then this term becomes $\frac{\|v\|}{2\|Av\|}$. Since A is SPD, we know that for any non-zero vector v:

$$\lambda_{\min}(A)\|v\|^2 \le \langle Av, v\rangle \le \lambda_{\max}(A)\|v\|^2$$

and also

$$\lambda_{\min}(A)\|v\| \le \|Av\| \le \lambda_{\max}(A)\|v\|$$

Therefore, the term $\frac{\|v\|}{2\|Av\|}$ is bounded:

$$\frac{1}{2\lambda_{\max}(A)} \le \frac{\|v\|}{2\|Av\|} \le \frac{1}{2\lambda_{\min}(A)}$$

This provides an explicit range for the second part of the step size choice, depending directly on the eigenvalues of A.

1.2 Solution for Applying New Stepsizes in Solving Linear Systems

Task: Use these new (improved) rules for step size selection in the context of solving linear systems $Ax = b$.

1.2.1 Algorithm Integration

The problem of solving a linear system $Ax = b$ is equivalent to minimizing the quadratic function $f(x) = \frac{1}{2}x^T Ax - b^T x$. Therefore, the adaptive gradient descent algorithm from the paper can be directly applied. The iterative scheme becomes:

$$\begin{aligned}
x_{k+1} &= x_k - \lambda_k(Ax_k - b) \\
\lambda_k &= \min\left\{\sqrt{1 + \theta_{k-1}}\lambda_{k-1}, \frac{\|x_k - x_{k-1}\|}{2\|A(x_k - x_{k-1})\|}\right\} \\
\theta_k &= \frac{\lambda_k}{\lambda_{k-1}}
\end{aligned}$$

with $x_0 \in \mathbb{R}^d$, $\lambda_0 > 0$, $\theta_0 = +\infty$.