Implementation of Adaptive Step-Size for Convex Optimization

Ali Ghasemzadeh - 401106339
Armin Khosravi - 401105872
Donya Jafari - 401101524

February 4, 2025

1 Introduction

This report presents the implementation of the adaptive step-size optimization algorithm proposed in *Adaptive Accelerated Composite Minimization*. The algorithm considers different cases based on acceleration and composite structures. The convergence details and theoretical analysis are provided in the original paper.

2 Formulas

The key formulas used in the adaptive step-size method are as follows:

$$x_{k+1} = x_k + \lambda_k d_k,$$

where

$$\varphi_k(\lambda) := f(x_k + \lambda d_k).$$

For a general optimization problem:

$$F(x) = \min_{x \in X} (f(x)) + h(x),$$

the adaptive step size is determined by:

$$\lambda_k := \arg\max_{\lambda} \lambda \quad \text{such that the inequalities in the table below hold.}$$

The proximal operator is defined as:

$$\mathrm{prox}_h(x) = \arg\min_u \left(h(u) + \frac{1}{2}\|u - x\|^2 \right),$$

and the gradient mapping is:

$$G^f_{\lambda h}(x) := \frac{1}{\lambda} \left(x - \mathrm{prox}_{\lambda h}(x - \lambda \nabla f(x)) \right).$$

Below is a table summarizing the step size rules and convergence rates for different settings:

Algorithm	**Problem**	**Stepsize Rule**
Non-accelerated	Composite	$\varphi(2\lambda) \leq \varphi(\lambda) - \lambda(G^f_{\lambda h}(x), \nabla f(x)) + \frac{\lambda}{2}\|G^f_{\lambda h}(x)\|^2$
Non-accelerated	Smooth	$\varphi(2\lambda) \leq \varphi(\lambda) + \frac{\lambda}{2}\varphi'(0)$
Accelerated	Composite	$\varphi(2\lambda) \leq \varphi(\lambda) - \lambda(G^f_{\lambda h}(x), \nabla f(x)) + \frac{\lambda}{2}\|G^f_{\lambda h}(x)\|^2$
Accelerated	Smooth	$\varphi(2\lambda) \leq \varphi(\lambda) + \frac{\lambda}{2}\varphi'(0)$

3 Problem List and Implementation

The following problems were implemented and tested:

1. Logistic Regression
2. Quadratic programming
3. Log-Sum-Exp

4. Approximate Semidefinite Programming for the Max-Cut problem
5. ℓ1-Regularized least square
6. ℓ1-Constrained least square
7. ℓ1-Regularized logistic regression
8. ℓ1-Regularized least square (acc)
9. ℓ1-Regularized logistic regression (acc)

For each problem, we provide two plots: one showing the function value convergence and another depicting step-size evolution.

3.1 Non-Accelerated Cases

The first set of experiments evaluates the non-accelerated versions of the algorithm across all seven problems.

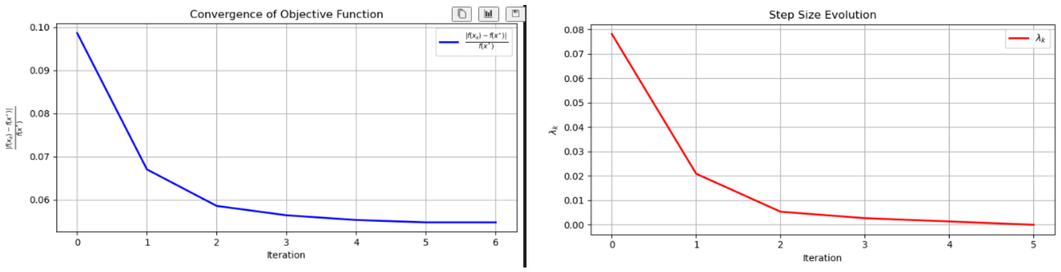

Figure 1: Results for Logistic Regression (Non-Accelerated)

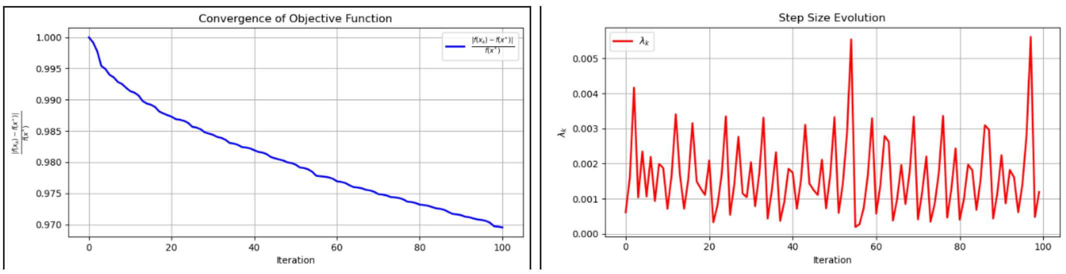

Figure 2: Results for Quadratic Programming (Non-Accelerated)

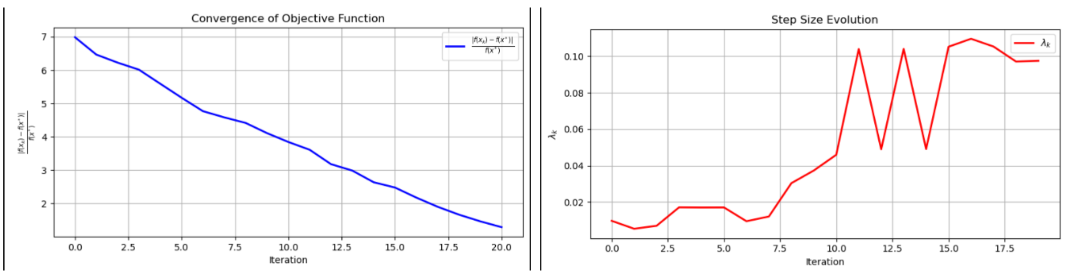

Figure 3: Results for Log-Sum-Exp (Non-Accelerated)

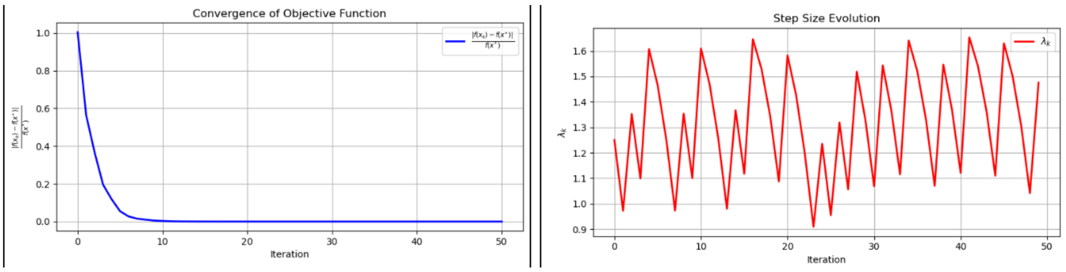

Figure 4: Results for Max-Cut problem (Non-Accelerated)

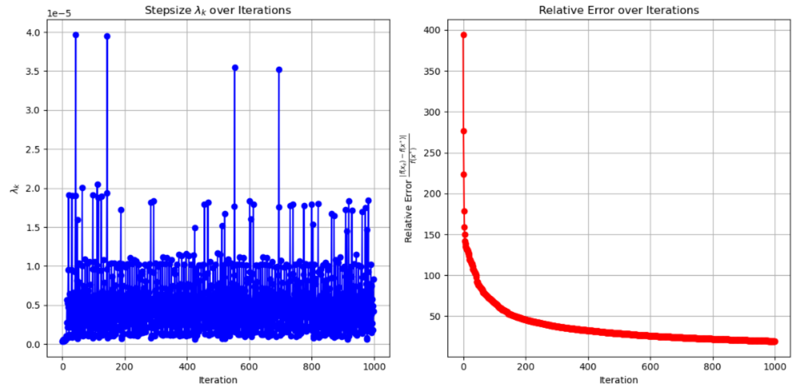

Figure 5: Results for $\ell 1$-Regularized least square (Non-Accelerated)

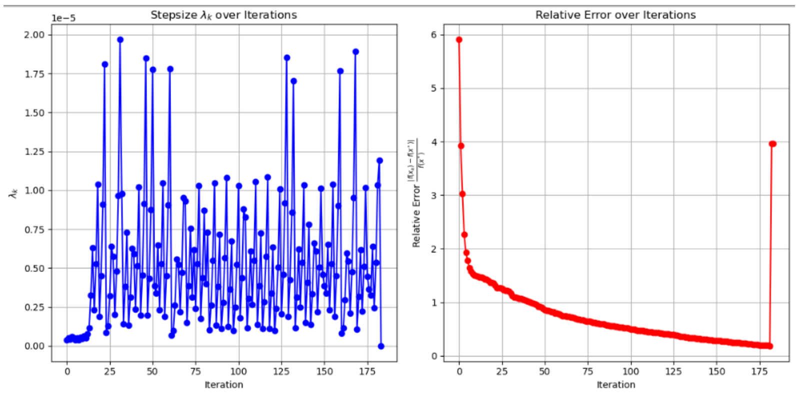

Figure 6: Results for $\ell 1$-Constrained least square (Non-Accelerated)

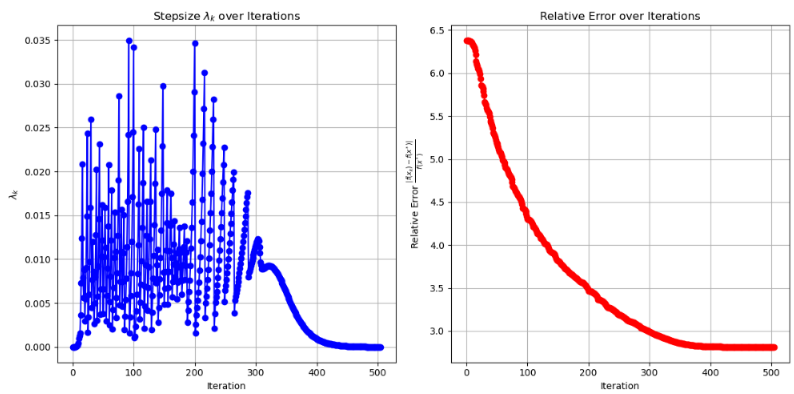

Figure 7: Results for $\ell 1$-Regularized logistic regression (Non-Accelerated)

3.2 Accelerated Cases

Acceleration was applied only to two of the convex problems. The results are presented below.

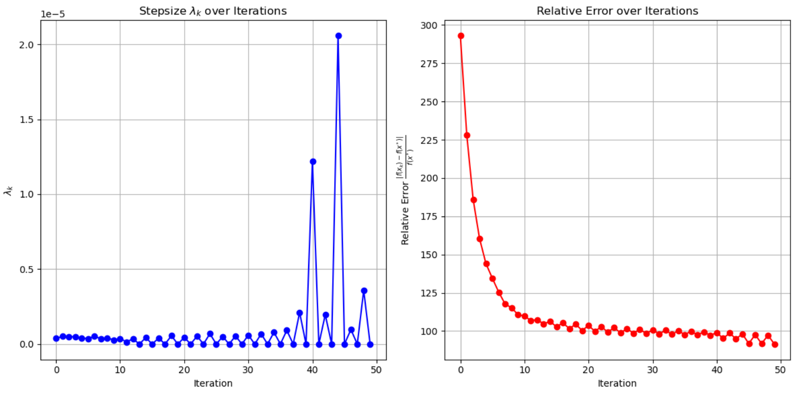

Figure 8: Results for ℓ1-Regularized Least Square (Accelerated)

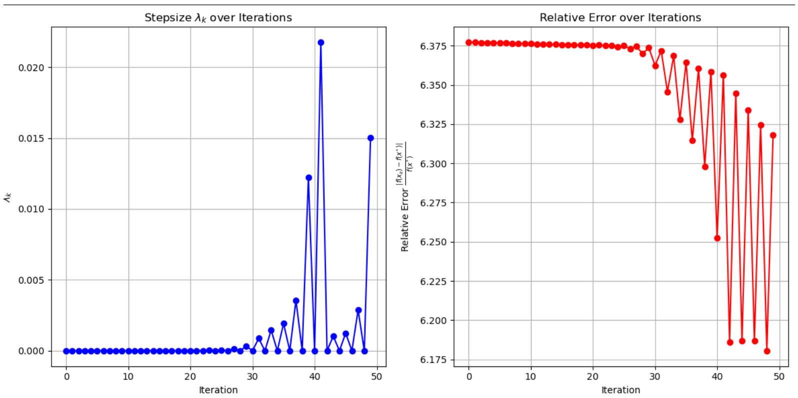

Figure 9: Results for ℓ1-Regularized logistic regression (Accelerated)

4 Initialization and Backtracking

The initial value of the step size, λ_0, is determined by:

$$\lambda_0 = 2\left(\frac{f(x_{k-1}) - f(x_k)}{\|\nabla f(x_k)\|^2}\right).$$

Backtracking is then used to find the largest step size λ such that the condition holds. For the smooth case, the direction is given by the gradient, and for the composite case, the direction is $-G$.

For accelerated composite cases, we used:

$$\begin{cases} \beta_k = \dfrac{1+\sqrt{1+4\beta_{k-1}^2}}{2},\ \gamma_k = \dfrac{1-\beta_k}{\beta_{k+1}},\ x_0 = x_1,\ \beta_1 = 0, \\ \lambda_k = \underset{\lambda\in\mathbb{R}}{\arg\max}\left\{\lambda \mid \phi(2\lambda) \leq \phi(\lambda) - \lambda\langle G^f_{\lambda h}(x), \nabla f(x)\rangle + \dfrac{\lambda}{2}\|G^f_{\lambda h}(x)\|^2,\ \lambda \leq \lambda_{k-1}\right\}, \\ d_k = -\gamma_k\left(\dfrac{\lambda_{k-1}}{\lambda_k}d_{k-1} - (1+\dfrac{1}{\gamma_k})G^f_{\lambda_k h}(x_k) + \dfrac{\lambda_{k-1}}{\lambda_k}\nabla G^f_{\lambda_{k-1}h}(x_{k-1})\right). \end{cases} \tag{Alg. 2}$$

Figure 10: Accelerated update rule (there is a typo in the gradient of G. we only need G here)

5 Results and Discussion

The results indicate that the adaptive step-size method effectively optimizes the convex problems. The impact of acceleration is observed in faster convergence for the selected problem.

References

[1] M. Mohajerin Esfahani, et al., "Adaptive Step-Size Methods for Convex Optimization," 2024.