25S 01.5: SVRG and Visualization of Variational Inequality Algorithms

JunHyun Kim

June 2025

1 Introduction

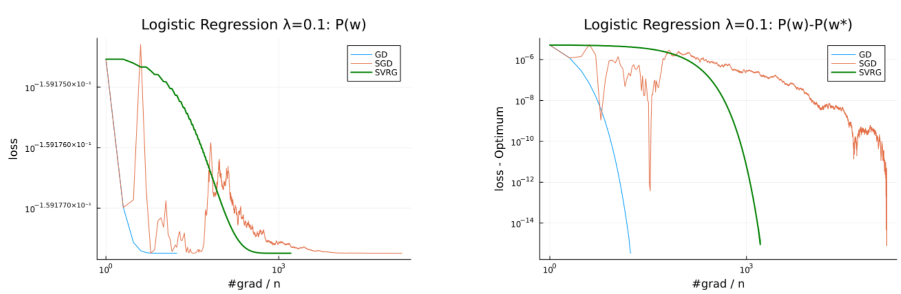
This report presents two experiments: (1) comparing optimization methods for logistic regression on a small synthetic dataset (100×100), and (2) reproducing a variational inequality (VI) visualization from the BGR paper to analyze convergence behavior of three VI solvers.

2 Logistic Regression (100×100)

Logistic regression with L2-regularization was implemented on a small dataset of shape 100×100, and three optimization methods GD, SGD, and SVRG were compared.

The objective function is

$$f(w) = \frac{1}{n}\sum_{i=1}^{n} \log\left(1 + \exp(-b_i a_i^\top w)\right) + \frac{\lambda}{2}\|w\|^2, \quad \lambda = 0.1$$



(a) Loss vs. #grad/n

(b) Loss - Optimum vs. #grad/n

Figure 1: Log-log plots of logistic regression on a 100×100 dataset. Comparison of GD, SGD, and SVRG.

- **GD:** Shows the most stable and fastest convergence throughout the entire training. As the full gradient is inexpensive to compute in this small scale problem, GD is particularly effective in this setting.
- **SGD:** Converges the fastest in the early iterations, but due to high variance, it exhibits a noticeable spike in the loss before continuing to decrease. This instability results in the slowest overall convergence.
- **SVRG:** Effectively reduces variance and avoids the fluctuations observed in SGD. It achieves smoother and more stable convergence, although it does not outperform GD in this specific setting.

3 Visualization of Variational Inequality Algorithms

This part reproduces Figure 3 from the BGR paper using the `beyond_golden_ratio` code. The goal is to visualize the convergence trajectories of three VI solvers on a bilinear saddle-point problem known as the polar game.

Algorithms Compared

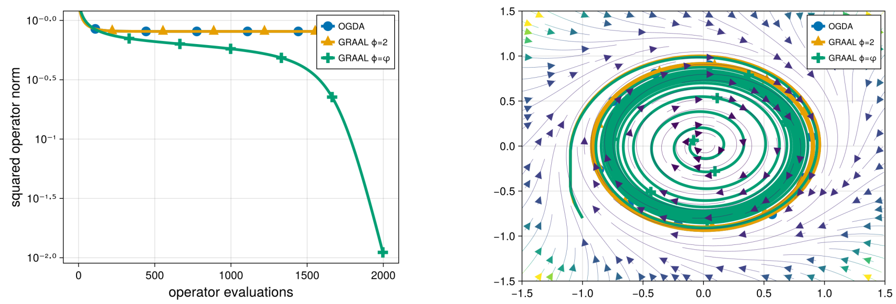
- Optimistic Gradient Descent-Ascent (OGDA)
- GRAAL with constant $\phi = 2$
- GRAAL with $\phi = \varphi$ (golden ratio)

Implementation

The experiment is run using the following setup:

```
VI, params = polar_game(1/3, 2000)

algorithms = [
    algorithm(ogda, "OGDA"),
    algorithm(golden_ratio, "GRAAL phi=2", (; phi=2.0)),
    algorithm(golden_ratio, "GRAAL phi=golden", (; phi=(1 + sqrt(5))/2)
        )
]

data = []

for algo in algorithms
    cb = Callback(label=algo.label)
    algo.method(VI, params, cb; algo.params...)
    push!(data, cb)
end

fig = plot_iterates(data, VI, params.path)
display(fig)
```

The function `plot_iterates` in `plotting-makie.jl` generates the streamplot with algorithm trajectories by reshaping the iterates and plotting them over the VI's vector field. The resulting visualization exactly reproduces Figure 3 from the BGR paper.



(a) Squared operator norm vs. operator evaluations

(b) Streamplot and iterates

Figure 2: Polar game experiment with weak Minty solution ($\rho = \frac{1}{3}$). Comparison of OGDA, GRAAL $\phi = 2$, and GRAAL $\phi = \varphi$. GRAAL with φ shows clear convergence, while other methods stagnate.

- OGDA shows almost no progress, remaining in a circular trajectory due to the rotational structure of the problem.
- GRAAL with $\phi = 2$ also stagnates and fails to reduce the operator norm.
- GRAAL with $\phi = \varphi$ (the golden ratio) achieves clear and consistent convergence, successfully reducing the operator norm.

4 Conclusion

Gradient descent converged the fastest on the logistic regression task, while SGD suffered from variance instability. SVRG achieved stable and reliable performance.

In the VI experiment, only GRAAL with $\phi = \varphi$ converged effectively, while OGDA and GRAAL with $\phi = 2$ stagnated.

These results illustrate how variance reduction improves stochastic optimization, and how proper parameter choices significantly affect convergence in variational inequality algorithms.