Zero-order methods for Saddle Point

by Antyshev Tikhon

2022

Establishing goals

- We consider following stochastic saddle point problem:

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x, y)$$

  where $f(x, y) \stackrel{\text{def}}{=} \mathbb{E}_{\xi} f(x, y, \xi)$.
- However, we only possess zero-order oracle information. Derivatives either do not exist or are not available.

Gradient Approximation

- We assume that we can call a noise-corrupted zeroth-order oracle: $\phi(x, y, \xi) \stackrel{\text{def}}{=} f(x, y, \xi) + \delta(x, y)$
- To approximate gradient we pick two vectors from Euclidian sphere:$\mathbf{e}_x, \mathbf{e}_y$ and define $\mathbf{e} = \begin{pmatrix} \mathbf{e}_x \\ -\mathbf{e}_y \end{pmatrix}$

$$g(x, y, \mathbf{e}) = \frac{d_x + d_y}{2\tau} \left(\phi\left((x, y) + \tau\mathbf{e}; \xi\right) - \phi\left((x, y) - \tau\mathbf{e}; \xi\right)\right) \cdot \mathbf{e}$$

Support Vector Machine SP

- The Lagrangian of the SVM is:

$$\mathcal{L}(w, \lambda) = \frac{1}{2}\|w\|^2 + \sum_{i=1}^{l} \lambda_i (1 - M_i(w))$$

- This results in a SP:

$$\min_{w} \max_{\lambda} \mathcal{L}(w, \lambda)$$

Mirror Descent algorithm

- To solve the optimization task we will be using a zeroth-order modification of the Mirror Descent:

Algorithm 1 Zeroth-order SMD

Input: iteration number N,
$z^1 \leftarrow \arg\min_{z\in\mathcal{Z}} d(z)$
for $k = 1, \ldots, N$ **do**
  Sample $\boldsymbol{e}^k, \xi^k$ independently
  Initialize $\gamma_k \to \frac{\mathcal{D}}{M}\sqrt{\frac{2}{N}}$ with M defined by (6) or (7)
  Calculate $g(z^k, \xi^k, \boldsymbol{e}^k)$ via (3)
  $z^{k+1} \leftarrow \mathrm{Prox}_{z^k}\left(\gamma_k g(z^k, \xi^k, \boldsymbol{e}^k)\right)$
end for
Output: $\hat{z}^N \leftarrow \left(\sum_{i=1}^N \gamma_k\right)^{-1} \sum_{k=1}^N \gamma_k z^k$

Figure: The 0-order MD algorithm

The Experiment

- The test is performed on a synthetic low-dimensional data.

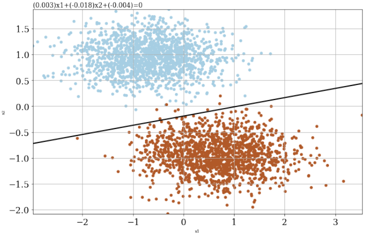

Figure: Visualized results

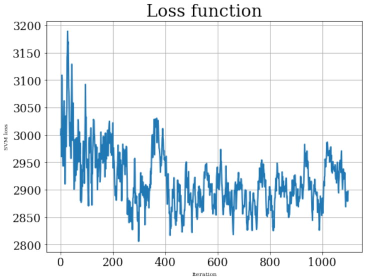

Figure: Loss

Conclusion

- As a result we demonstrated, that zeroth order methods are capable of solving ML SP tasks.
- There are still more testing to be done in comparing Mirror Descent to Mirror Prox algorithm.
- It should also be tried on large classification task.

References

Darina Dvinskikh, Vladislav Tominin, Yaroslav Tominin, and Alexander Gasnikov. Gradient-free optimization for non-smooth saddle point problems under adversarial noise. 2022. doi: 10.48550/ARXIV.2202.06114. URL https://arxiv.org/abs/2202.06114.

Alexander Gasnikov, Anton Novitskii, Vasilii Novitskii, Farshed Abdukhakimov, Dmitry Kamzolov, Aleksandr Beznosikov, Martin Takáč, Pavel Dvurechensky, and Bin Gu. The power of first-order smooth optimization for black-box non-smooth problems. 2022. doi: 10.48550/ARXIV.2201.12289. URL https://arxiv.org/abs/2201.12289.