
Replace with your title

Yao-Liang Yu

School of Computer Science
University of Waterloo
Waterloo, ON, N2L 3G1
yaoliang.yu@uwaterloo.ca

report due: July 7.

Abstract

1 Put here a brief summary of the project: what is it about and what are the main
2 results. Consider put a link to your code for reproducibility (if applicable). Be
3 concise and to the point.

4 1 Introduction

5 In this section you are going to present a brief background and motivation of your project. Why is
6 it interesting/significant? Consider summarizing the entire paper in one overarching figure, such as
7 Figure 1.

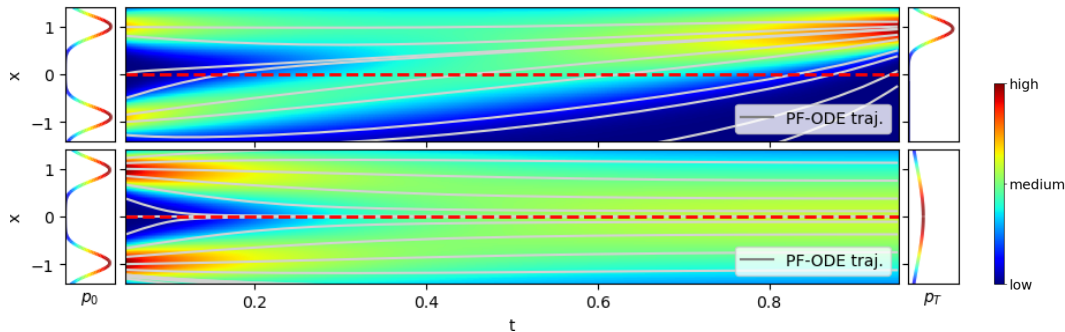


Figure 1: The evolution of p_t driven by diffusion processes where the data distribution p_0 is invariant under flipping with respect to the origin. We also plot the PF-ODE trajectories to visualize the transition direction of $p_t(x)$. The upper plot has $f(x, t) = \frac{1-x}{1-t}$ and $g(t) = 1$. The lower is VP-SDE with $\alpha_t = 1 - t$. For both processes, $T = 0.95$.

8 2 Related Works

9 Perform a reasonably thorough review of relevant literature. Has your problem, or one of similar
10 nature, been considered before? By whom? What are the differences or limitations (if any)?

11 3 Main Results

12 In the following, describe the background of your project, formulate your problem precisely (math-
13 ematically), and present the main findings (often backed up by experiments, proofs, figures and

Table 1: Model Comparison on 28x28x1 Rotated MNIST (Group C4). * indicates author-reported values.

Model	FID↓				Inv-FID↓	$\Delta\hat{x}_0 \downarrow$
	1%	5%	10%	100%	100%	100%
SPDiff	5.97	3.05	3.47	2.81	2.21	0.2997
SPDiff+WT	5.80	3.34	3.57	3.50	2.20	0.0004
SPDiff+OC	6.10	3.09	3.45	2.82	2.12	0.0002
SPDiff+Reg	5.42	3.69	2.83	2.75	2.09	0.1806
SPDiff+Reg+OC	5.64	3.67	2.86	2.64	2.07	0.0002
SP-GAN	149*	99*	88*	81*	—	—
SP-GAN (Reprod.)	16.59	11.28	9.02	10.95	19.92	—

tables). If need be, consider having a separate background section, experiment section or discussion section.

Please always give proper citations to prior work or results. Be precise and concise. Pay some attention to the organization and layout of the entire paper: the smoother it reads and the more visually appealing and neat it is, the better (sounds superficial but remember: we are selling our work to very busy and impatient peers). Add variety (table, curves, bar graph, scatter plot, violin plot, pseudocode, etc.) and report statistical deviation (over at least 3~5 runs).

I expect the report to be less than (\leq) **4 pages** (references excluded).

Algorithm 1: Stochastic variance reduced proximal gradient

Input: $\mathbf{w}_0 \in \text{dom } f$

```

1 for  $k = 0, 1, 2, \dots$  do
2    $\mathbf{g}_k \leftarrow \frac{1}{n} \sum_{i=1}^n \nabla \ell_i(\mathbf{w}_k)$  // compute full gradient at epoch  $k$ 
3    $\mathbf{w}_{k,0} \leftarrow \mathbf{w}_k$ 
22 4   for  $t = 0, \dots, m-1$  do
5     randomly draw  $i_t = i$  with probability  $p_i$ 
6      $\mathbf{g}_{k,t} \leftarrow \mathbf{g}_k - \frac{1}{np_{i_t}} \nabla \ell_{i_t}(\mathbf{w}_k) + \frac{1}{np_{i_t}} \nabla \ell_{i_t}(\mathbf{w}_{k,t})$  // amortized gradient
7      $\mathbf{w}_{k,t+1} \leftarrow \text{P}_r^{\eta k}(\mathbf{w}_{k,t} - \eta_k \mathbf{g}_{k,t})$  // stochastic proximal gradient
8    $\mathbf{w}_{k+1} \leftarrow \frac{1}{m} \sum_{t=1}^m \mathbf{w}_{k,t}$  // in practice, can also do  $\mathbf{w}_{k+1} \leftarrow \mathbf{w}_{k,m}$ 

```

4 Conclusion

What have we learned? What limitations or directions do you think are worth exploring in the future?

5 Some possible readings

- Bauschke, H. H., M. N. Bui, and X. Wang (2019). “Applying FISTA to optimization problems (with or) without minimizers”. *Mathematical Programming*, vol. 184, pp. 349–381.
- Bubeck, S. and R. Eldan (2019). “The Entropic Barrier: Exponential Families, Log-Concave Geometry, and Self-Concordance”. *Mathematics of Operations Research*, vol. 44, no. 1, pp. 264–276.
- Carderera, A., M. Besançon, and S. Pokutta (2021). “Simple steps are all you need: Frank-Wolfe and generalized self-concordant functions”. In: *Advances in Neural Information Processing Systems*.
- Chen, T., Y. Sun, and W. Yin (2021). “Closing the Gap: Tighter Analysis of Alternating Stochastic Gradient Methods for Bilevel Problems”. In: *Advances in Neural Information Processing Systems*.
- Cheng, H.-B., L.-T. Cheng, and S.-T. Yau (2005). “Minimization with the affine normal direction”. *Communications in Mathematical Sciences*, vol. 3, no. 4, pp. 561–574.
- Cohen, J., S. Kaur, Y. Li, J. Z. Kolter, and A. Talwalkar (2021). “Gradient Descent on Neural Networks Typically Occurs at the Edge of Stability”. In: *International Conference on Learning Representations*.
- Crawshaw, M., M. Liu, F. Orabona, W. Zhang, and Z. Zhuang (2022). “Robustness to Unbounded Smoothness of Generalized SignSGD”. In: *Advances in Neural Information Processing Systems (NeurIPS)*.
- Dagréou, M., P. Ablin, S. Vaiter, and T. Moreau (2022). “A framework for bilevel optimization that enables stochastic and global variance reduction algorithms”. In: *Advances in Neural Information Processing Systems*.
- Davis, D. and D. Drusvyatskiy (2019). “Stochastic Model-Based Minimization of Weakly Convex Functions”. *SIAM Journal on Optimization*, vol. 29, no. 1, pp. 207–239.
- Davis, D., D. Drusvyatskiy, K. J. MacPhee, and C. Paquette (2018). “Subgradient Methods for Sharp Weakly Convex Functions”. *Journal of Optimization Theory and Applications*, vol. 179, pp. 962–982.
- Diakonikolas, J. and M. I. Jordan (2021). “Generalized Momentum-Based Methods: A Hamiltonian Perspective”. *SIAM Journal on Optimization*, vol. 31, no. 1, pp. 915–944.
- Dvurechensky, P., K. Safin, S. Shtern, and M. Staudigl (2023). “Generalized self-concordant analysis of Frank-Wolfe algorithms”. *Mathematical Programming*, vol. 198, pp. 255–323.
- Foret, P., A. Kleiner, H. Mobahi, and B. Neyshabur (2021). “Sharpness-aware Minimization for Efficiently Improving Generalization”. In: *International Conference on Learning Representations*.
- Gupta, V., T. Koren, and Y. Singer (2018). “Shampoo: Preconditioned Stochastic Tensor Optimization”. In: *Proceedings of the 35th International Conference on Machine Learning*, pp. 1842–1850.
- Khanh, P. D., H.-C. Luong, B. Mordukhovich, and D. B. Tran (2024). “Fundamental Convergence Analysis of Sharpness-Aware Minimization”. In: *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.
- Kwon, J., D. Kwon, S. Wright, and R. D. Nowak (2023). “A Fully First-Order Method for Stochastic Bilevel Optimization”. In: *Proceedings of the 40th International Conference on Machine Learning*.
- Li, H., J. Qian, Y. Tian, A. Rakhlin, and A. Jadbabaie (2023). “Convex and Non-convex Optimization Under Generalized Smoothness”. In: *Thirty-seventh Conference on Neural Information Processing Systems*.

- 69 Liu, R., Y. Liu, S. Zeng, and J. Zhang (2021). “Towards Gradient-based Bilevel Optimization with
70 Non-convex Followers and Beyond”. In: *Advances in Neural Information Processing Systems*.
- 71 Long, P. M. and P. L. Bartlett (2024). “Sharpness-Aware Minimization and the Edge of Stability”.
72 *Journal of Machine Learning Research*, vol. 25, no. 179, pp. 1–20.
- 73 Manton, J. H. (2020). “Geometry, Manifolds, and Nonconvex Optimization: How Geometry Can
74 Help Optimization”. *IEEE Signal Processing Magazine*, vol. 37, no. 5, pp. 109–119.
- 75 Morwani, D., I. Shapira, N. Vyas, E. Malach, S. M. Kakade, and L. Janson (2025). “A New Per-
76 spective on Shampoo’s Preconditioner”. In: *The Thirteenth International Conference on Learning*
77 *Representations*.
- 78 Sow, D., K. Ji, and Y. Liang (2022). “On the Convergence Theory for Hessian-Free Bilevel Algo-
79 rithms”. In: *Advances in Neural Information Processing Systems*.
- 80 Wan, Y., W.-W. Tu, and L. Zhang (2022). “Online Frank-Wolfe with Arbitrary Delays”. In: *Advances*
81 *in Neural Information Processing Systems*.
- 82 Zhang, J., T. He, S. Sra, and A. Jadbabaie (2020). “Why Gradient Clipping Accelerates Training: A
83 Theoretical Justification for Adaptivity”. In: *International Conference on Learning Representa-*
84 *tions*.
- 85 Zhao, Y., H. Zhang, and X. Hu (2022). “Penalizing Gradient Norm for Efficiently Improving Gen-
86 eralization in Deep Learning”. In: *Proceedings of the 39th International Conference on Machine*
87 *Learning*, pp. 26982–26992.

88 **Acknowledgement**

89 Thank people who have helped or influenced you in this project.