Information and Advice for Prospective Students

1. General Background

I am a mathematician by training. So in order to work with me, in addition to being extremely independent, you should have a strong background in probability and analysis. Specifically, you should have received an A in ORF 523, ORF 525, ORF 526, and ORF 527, or their equivalents. You will also be expected at some point to learn high dimensional probability at the level of getting an A in ORF 550.

2. My Research

My research these days is focused on the theoretical study of neural networks. The main directions of my work are:

- Understanding the approximation capacity of neural networks [HS17, Han17, DDF$^+$19, DHP21]
- Understanding the behavior of neural networks at the start of training [HR18, Han18, HN19b, HR19, HR, HN19a, Han21a, HP21, RYH21, HJR21]
- Understanding generalization for neural networks [RYH21, HS, HN19a, Han21b]

The questions I am interested in tend to be strongly rooted in understanding how neural networks work in practice and how to make them better. Mathematically, my work typically uses tools from stochastic processes, random matrix theory, high dimensional probability, combinatorics, random geometry, and functional analysis.

3. Advising Style

My advising style emphasizes your need to be broadly interested in theoretical machine learning and mathematics. This means that rather than giving you a specific problem to work on, I will help you choose a direction of study by providing the input Papers to the advising algorithm listed in the next section. This algorithm is meant to get you up to speed in an area. Once it is successfully executed, I'll help you look for a problem to work on. In the meantime, if you're potentially interested in working me, you should participate (i.e. attend and present at) my informal seminar on deep learning theory, which runs every week. You should also attend my ORF 543 course on deep learning theory.

4. Advising Algorithm

```
def BorisAdvising(Papers) :

    CurrentUnderstanding = 0
    TargetUnderstanding = 1

    While interested == True and CurrentUnderstanding < BroadlyKnowledgeable :

        Learning(CurrentUnderstanding, TargetUnderstanding, Papers)
        Result = Presentation(Time = 20min, Venue = AtBoard,
                     CurrentUnderstanding, TargetUnderstanding)

        If Result == True :
            CurrentUnderstanding += 1
            TargetUnderstanding += 1
            Presentation(Time = 50min, Venue = LearningSeminar,
                         CurrentUnderstanding, TargetUnderstanding)

    def Learning(CurrentUnderstanding, TargetUnderstanding, Papers) :

        While CurrentUnderstanding < TargetUnderstanding :

            Read(Papers)
            Read(References)

    def Presentation(Time, Venue, CurrentUnderstanding, TargetUnderstanding) :

        If Venue == AtBoard :
            While CurrentUnderstanding < TargetUnderstanding :
                Read(Papers)
                Read(References)

                return CurrentUnderstanding >= TargetUnderstanding
```

References

[DDF$^+$19] Ingrid Daubechies, Ronald A. DeVore, Simon Foucart, Boris Hanin, and Guergana Petrova. Nonlinear approximation and (deep) relu networks. *arXiv:1905.02199*, 2019.

[DHP21] Ronald DeVore, Boris Hanin, and Guergana Petrova. Neural network approximation. *Acta Numerica*, 30:327–444, 2021.

[Han17] Boris Hanin. Universal function approximation by deep neural nets with bounded width and relu activations. *arXiv preprint arXiv:1708.02691*, 2017.

[Han18] Boris Hanin. Which neural net architectures give rise to exploding and vanishing gradients? In *Advances in Neural Information Processing Systems*, 2018.

[Han21a] Boris Hanin. Random neural networks in the infinite width limit as gaussian processes. *arXiv preprint arXiv:2107.01562*, 2021.

[Han21b] Boris Hanin. Ridgeless interpolation with shallow relu networks in 1d is nearest neighbor curvature extrapolation and provably generalizes on lipschitz functions. *arXiv preprint arXiv:2109.12960*, 2021.

[HJR21] Boris Hanin, Ryan Jeong, and David Rolnick. Deep relu networks preserve expected length. *arXiv preprint arXiv:2102.10492*, 2021.

[HN19a] Boris Hanin and Mihai Nica. Finite depth and width corrections to the neural tangent kernel. *ICLR 2020 and arXiv:1909.05989*, 2019.

[HN19b] Boris Hanin and Mihai Nica. Products of many large random matrices and gradients in deep neural networks. *Communications in Mathematical Physics (in Press). arXiv:1812.05994*, 2019.

[HP21] Boris Hanin and Grigoris Paouris. Non-asymptotic results for singular values of gaussian matrix products. *Geometric and Functional Analysis*, 31(2):268–324, 2021.

[HR] Boris Hanin and David Rolnick. Deep relu networks have surprisingly few activation patterns. *NeurIPS 2019*.

[HR18] Boris Hanin and David Rolnick. How to start training: The effect of initialization and architecture. In *Advances in Neural Information Processing Systems*, pages 571–581, 2018.

[HR19] Boris Hanin and David Rolnick. Complexity of linear regions in deep networks. *ICML*, 2019.

[HS] Boris Hanin and Yi Sun. Data augmentation as stochastic optimization. *arXiv:2010.11171*.

[HS17] Boris Hanin and Mark Sellke. Approximating continuous functions by relu nets of minimal width. *arXiv preprint arXiv:1710.11278*, 2017.

[RYH21] Daniel A Roberts, Sho Yaida, and Boris Hanin. The principles of deep learning theory. *arXiv preprint arXiv:2106.10165*, 2021.