

View Reviews

Paper ID

8246

Paper Title

Precise expressions for random projections: Low-rank approximation and randomized Newton

Reviewer #1

Not Submitted

Reviewer #2

Questions

1. Summary and contributions: Briefly summarize the paper and its contributions.

This paper proposes a semi-explicit expression for the expected value of a random projection matrix involved in sketching. More precisely, if S is the sketch matrix for A the data matrix, then in order to investigate the error of sketching one needs to control $\|P_{orth} - I\|_F$, where P is the orthogonal projection on the subspace spanned by the rows of SA . The main result of the paper is that $\mathbb{E}\|P_{orth} - I\|_F = O(\sqrt{\gamma} \|A\|_F)$ up to error terms depending on the stable rank of A , and where γ is solution of some equation. Some applications of this results are given in the field of randomized iterative optimization, and the validity of the main result is verified by several numerical experiments.

2. Strengths: Describe the strengths of the work. Typical criteria include: soundness of the claims (theoretical grounding, empirical evaluation), significance and novelty of the contribution, and relevance to the NeurIPS community.

The theoretical result is new, and has very interesting corollaries for randomized iterative optimization schemes (namely, trivializing the convergence proofs). Another application of the result is implicit regularization: the paper shows that Theorem 1 can be used to study the bias of the ℓ_2 regularized solution for least squares.

The exposition of the theoretical results is very neat, and quite understandable.

Some care was also given in demonstrating numerically the pertinence of the main result in simple, understandable cases.

3. Weaknesses: Explain the limitations of this work along the same axes as above.

In my opinion, the main caveat is the non explicitness of γ . While this is seriously addressed by the paper in Section 4, I feel like I still do not understand well the behavior of γ after reading the paper.

4. Correctness: Are the claims and method correct? Is the empirical methodology correct?

As far as I can tell, the proof of the theoretical results are correct.

One minor thing: line 87, the paper states that " γ increases at least linearly as a function of k ." Even if the proof is not complicated, I think it deserves to be explained to help the reader get intuition on γ . Simple

bounds, and maybe a graph, would also help.

5. Clarity: Is the paper well written?

The paper is very well-written.

6. Relation to prior work: Is it clearly discussed how this work differs from previous contributions?

The prior work is appropriately cited. Maybe it would just be better to cite Hanson and Wright when Hanson-Wright inequality is used in Lemma 2 even though the version from Zajkowski (2018) is used.

7. Reproducibility: Are there enough details to reproduce the major results of this work?

Yes

8. Additional feedback, comments, suggestions for improvement and questions for the authors:

This is just a suggestion, but I think that Section 1.3 should be expanded. The current version of the paper goes a bit quickly there for readers that are not familiar with randomized iterative optimization, double descent, or both topics. It is especially hard to see from where the equation line 136 comes from. This could be achieved easily by removing one of the two proof sketches of Theorem 2 (page 5): one of them is enough.

Another suggestion: since the closed-form expressions of γ are obtained in the $n, \epsilon \rightarrow +\infty$ limit, it looks possible to consider centered i.i.d. data and use Marcenko-Pastur to approximate the sum in (4), therefore obtaining another expression for γ in this case.

9. Please provide an "overall score" for this submission.

8: Top 50% of accepted NeurIPS papers. A very good submission; a clear accept.

10. Please provide a "confidence score" for your assessment of this submission.

4: You are confident in your assessment, but not absolutely certain. It is unlikely, but not impossible, that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work.

11. Have the authors adequately addressed the broader impact of their work, including potential negative ethical and societal implications of their work?

Yes

Reviewer #3

Questions

1. Summary and contributions: Briefly summarize the paper and its contributions.

This paper is concerned with sketching of matrices, where the authors use subgaussian sketching matrices. The authors derive accurate predictions for the residual projection matrix. This allows them to make accurate prediction for the low-rank approximation error.

Furthermore, the authors conduct simulations, which confirm the predictions of their theory.

2. Strengths: Describe the strengths of the work. Typical criteria include: soundness of the claims (theoretical grounding, empirical evaluation), significance and novelty of the contribution, and relevance to the NeurIPS community.

This work characterizes the low-rank approximation error precisely with respect to the singular values of the sketched matrix. Since their guarantee depends precisely on the distribution of the existing singular values, this goes significantly beyond existing literature, where typically worst-case guarantees are discussed.

This makes this paper an important contribution in the sketching literature.

3. Weaknesses: Explain the limitations of this work along the same axes as above.

-Theorem 1 and Theorem 2 only hold with high probability according to the proofs. This is missing in the statement of the both theorems. The authors should add this to the statement.

-In order to quantify the low-rank approximation error the authors use the quantity $E\|A-AP\|^2$. Since this is central to the paper, it would be great if the authors could add more explanation and motivation to it (and maybe mention other notions).

4. Correctness: Are the claims and method correct? Is the empirical methodology correct?

The claims are sound and the arguments in the main part are correct.

5. Clarity: Is the paper well written?

The paper is clearly written.

6. Relation to prior work: Is it clearly discussed how this work differs from previous contributions?

Prior work is clearly discussed.

7. Reproducibility: Are there enough details to reproduce the major results of this work?

Yes

8. Additional feedback, comments, suggestions for improvement and questions for the authors:

What is the high-level motivation of relying on the rank-one update formula in the proof (see Lemma 1)?

9. Please provide an "overall score" for this submission.

8: Top 50% of accepted NeurIPS papers. A very good submission; a clear accept.

10. Please provide a "confidence score" for your assessment of this submission.

2: You are willing to defend your assessment, but it is quite likely that you did not understand central parts of the submission or that you are unfamiliar with some pieces of related work. Math/other details were not carefully checked.

11. Have the authors adequately addressed the broader impact of their work, including potential negative ethical and societal implications of their work?

Yes

Reviewer #4

Questions

1. Summary and contributions: Briefly summarize the paper and its contributions.

This paper analyzes the effect of random projections on the spectrum of the matrix. Specifically, the authors show that, given a matrix A and a sketch S , the difference of the singular values of $A^T A$ and $A^T S^T S A$ is concentrated around the expectation. The authors also show a number of applications to linear regression, low rank approximation, and further problems and algorithms arising in numerical linear algebra.

2. Strengths: Describe the strengths of the work. Typical criteria include: soundness of the claims (theoretical grounding, empirical evaluation), significance and novelty of the contribution, and relevance to the NeurIPS community.

While I consider such a paper out of scope of NeurIPS (an 8 page paper with a proofs in the supplementary material is not the format that a subject like this needs), I do like these questions and consider the application of random projections to problems a worthwhile contribution.

The analysis seems rigorous, though I did not check it in detail (more on that below).

3. Weaknesses: Explain the limitations of this work along the same axes as above.

I unfortunately see papers like these often submitted to conferences that are not purely theoretical computer science, such as ICML, NeurIPS, AAAI, IJCAI and related conferences. While I wish that these methods would receive more widespread application and use, the related work done in TCS is often covered extremely poorly. At best, the authors might be merely sloppy (which in of itself already warrants a reject), at worst one has to wonder if this is deliberate, given the frequency with which this occurs.

To this paper at hand: the fact that random projections preserve subspaces has been known for quite some time. I will list relevant references further below, but citing lecture notes and a communications of the acm letters (mainly authored by one or two people) is not adequate related work, given all the references the authors missed.

Specifically, it is well known that for any fixed subspace of rank k , a Gaussian, or a Rademacher (or a multitude of other distributions the authors do not mention) matrix of target dimension $O(k/\epsilon^2)$ preserves the subspace. One could argue, since the authors consider the stable rank instead of the rank that this offers more useful bounds, but this has also been considered.

Technically, this papers does little more than write out the moments. While this is not how the results are usually presented in TCS, writing out the moments is in my opinion a different perspective rather than a new result. I also do not wish to argue this point overly much; had the authors compared their results with TCS literature and argued why these bounds are more useful; I would have likely given a high grade, if the arguments were convincing. But avoiding the discussion completely is simply not adequate.

I would like to emphasize that this is not merely about missing the odd reference, which is sometimes (and unfortunately) used as a justification of a lazy reviewer to shoot down a paper. This misses essentially an entire line of research, and until the authors treat the related work with the appropriate care, I'm afraid that this paper should not be published. A number of these results are also surveyed in David Woodruff's book, which the authors do cite. I would encourage them to go over that book again and compare their results with those included in that book. Nevertheless, there have also been further results published since. Quite simply, if the next submission does not contain at least 10 new citations, then it is insufficient.

Sarlos FOCS 2005. Improved Approximation Algorithms for Large Matrices via Random Projections (the first paper to show that JL-transforms preserve subspaces)

Kane, Nelson JACM 2014 (originally published 2012). Sparser Johnson-Lindenstrauss Transforms (general analysis of subspace embeddings and the moments necessary)

Clarkson, Woodruff STOC 2013. Low rank approximation and regression in input sparsity time (a landmark paper that showed how to make sketching-based algorithms for numerical linear algebra very efficient)

Nelson, Nguyen FOCS 2013. OSNAP: Faster Numerical Linear Algebra Algorithms via Sparser Subspace Embeddings (a generalization of the CW12 result)

Cohen, Elder, Musco, Musco, Persu, STOC 2015. Dimensionality Reduction for k-Means Clustering and Low Rank Approximation (arguably the strongest guarantee for preserving subspaces, along with an analysis of what the authors

call the residual projection matrix)

Cohen, Nelson, Woodruff ICALP 2016. Optimal Approximate Matrix Product in Terms of Stable Rank (Main result here is a generalization of theorem 1).

Finally, this is NOT a sufficient list, rather one should regard these papers as necessary. I am also quite confident that further results will have appeared since 2016 in various TCS conferences (and maybe also the odd ICML, NeurIPS paper). Most of these papers give results for low rank approximation and linear regression, so the claim that the results presented here are new is doubtful, without further discussion. What I would consider adequate related work would likely be 2-3 times as many citations.

4. Correctness: Are the claims and method correct? Is the empirical methodology correct?

Seems solid.

The reason I didn't check the proofs is because the results presented here are, in one form or another, either known results from TCS literature, or easy corollaries.

5. Clarity: Is the paper well written?

yes

6. Relation to prior work: Is it clearly discussed how this work differs from previous contributions?

absolutely inadequate.

7. Reproducibility: Are there enough details to reproduce the major results of this work?

Yes

9. Please provide an "overall score" for this submission.

2: A strong reject.

10. Please provide a "confidence score" for your assessment of this submission.

5: You are absolutely certain about your assessment. You are very familiar with the related work.

11. Have the authors adequately addressed the broader impact of their work, including potential negative ethical and societal implications of their work?

Yes