

684 A Convergence analysis

685 In this section, we derive EigenPro4.0-Exact (Algorithm 2), a precursor to EigenPro4.0. However, this
 686 version does not scale efficiently. In Section 3, we enhance its scalability by introducing stochastic
 687 approximations, resulting in EigenPro4.0 (Algorithm 1).

688 Recall that the derivatives of the loss function, as defined in (13), lie in the span of the training data,
 689 denoted as \mathcal{X} . However, these derivatives cannot directly update the model, which resides in the span
 690 of the model centers, \mathcal{Z} . To address this, we first fit the labels within the \mathcal{X} and then project the
 691 solution into the \mathcal{Z} . This process is repeated iteratively on the residual labels until convergence, as
 692 outlined in Algorithm algorithm 2

Algorithm 2 EigenPro 4-Exact

Require: Data (X, \mathbf{y}) , centers Z
 1: $\tilde{\mathbf{y}}_0 = \mathbf{y}$
 2: **for** $t = 1, 2, \dots$ **do**
 3: $\boldsymbol{\alpha}_t = K^{-1}(X, X)\tilde{\mathbf{y}}_t$
 4: $K(\cdot, Z)\boldsymbol{\beta}_t = \text{proj}_{\mathcal{Z}}(K(\cdot, X)\boldsymbol{\alpha}_t)$
 5: $\tilde{\mathbf{y}}_{t+1} = \mathbf{y} - K(X, Z)\boldsymbol{\beta}_t$
 6: **end for**

693 The following proposition provides the fixed point analysis for this algorithm.

694 **Proposition 1.** Consider any dataset X, \mathbf{y} and a choice of model centers Z , with a kernel function
 695 $K : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$. Assume that $K(X, X)$ and $K(Z, X)$ are full Rank. Then, Algorithm 2 converges
 696 to the following solution:

$$\hat{f} = K(\cdot, Z) \left(K(Z, X)K^{-1}(X, X)K(X, Z) \right)^{-1} K(Z, X)K^{-1}(X, X)\mathbf{y}. \quad (21)$$

697 Furthermore, if $\mathbf{y} = K(X, Z)\boldsymbol{\beta}^* + \boldsymbol{\xi}$, where $\boldsymbol{\xi}$ is a vector of independent centered random noise with
 698 $\mathbb{E}[\xi_i^2] = \sigma^2$, then

$$\lim_{t \rightarrow \infty} \mathbb{E}[\boldsymbol{\beta}_t] = \boldsymbol{\beta}^*, \quad \lim_{t \rightarrow \infty} \frac{\mathbb{E}[\|\boldsymbol{\beta}_t - \boldsymbol{\beta}^*\|^2]}{\sigma^2} = \text{tr} \left(\left(K(Z, X)K^{-1}(X, X)K(X, Z) \right)^{-2} K(Z, X)K^{-2}(X, X)K(X, Z) \right).$$

699 *Proof.* We begin by expressing Algorithm 2 recursively and substituting $\text{proj}_{\mathcal{Z}}$ with the expression
 700 in (18). Recall that $f_t = K(\cdot, Z)\boldsymbol{\beta}_t$ with base case $\boldsymbol{\beta}_0 = 0$. The update rule for $\boldsymbol{\beta}_t$ is given by:

$$\boldsymbol{\beta}_t = K^{-1}(Z, Z)K(Z, X)K^{-1}(X, X)(\mathbf{y} - K(X, Z)\boldsymbol{\beta}_{t-1}) + \boldsymbol{\beta}_{t-1}. \quad (22)$$

701 Let us define the matrices:

$$B := K^{-1}(Z, Z)K(Z, X)K^{-1}(X, X), \quad C := BK(X, Z) - I,$$

702 which allows us to rewrite the recursion more succinctly:

$$\begin{aligned} \boldsymbol{\beta}_t &= B(\mathbf{y} - K(X, Z)\boldsymbol{\beta}_{t-1}) + \boldsymbol{\beta}_{t-1} \\ &= B\mathbf{y} - C\boldsymbol{\beta}_{t-1} = B\mathbf{y} - CB\mathbf{y} + C^2\boldsymbol{\beta}_{t-2} \\ &\quad \vdots \\ &= \left(\sum_{i=0}^{t-1} (-1)^i C^i \right) B\mathbf{y}. \end{aligned} \quad (23)$$

703 As the number of iterations tends to infinity, we can define the infinite series sum:

$$S := \sum_{i=0}^{\infty} (-1)^i C^i.$$

704 Observe that:

$$S + CS = I.$$

705 Substituting the definition $C = BK(X, Z) - I$ and $B = K^{-1}(Z, Z)K(Z, X)K^{-1}(X, X)$, we
706 have:

$$K^{-1}(Z, Z)K(Z, X)K^{-1}(X, X)K(X, Z)S = I.$$

707 Thus, this simplifies to:

$$S = \left(K(Z, X)K^{-1}(X, X)K(X, Z) \right)^{-1} K(Z, Z).$$

708 Therefore, the final solution converges to:

$$\hat{f} = K(\cdot, Z) \left(K(Z, X)K^{-1}(X, X)K(X, Z) \right)^{-1} K(Z, X)K^{-1}(X, X)\mathbf{y}. \quad (24)$$

709 Substituting $\mathbf{y} = K(X, Z)\beta^* + \xi$ readily completes the second claim.

710

□

line in Algorithm 1	computation	flops
7	$K(X_m, Z)\alpha - \mathbf{y}_t$	mp
7	$K(X_m, Z_{\text{tmp}})\alpha_{\text{tmp}}$	$m^2(t - kT - 1)$
7	$K(X_m, X_s)\alpha_s$	ms
9, 10	$\mathbf{h}_1 := \mathbf{F}^\top K(X_s, X_m)\mathbf{g}_m \in \mathbb{R}^q$	$ms + sq$
9	$\mathbf{F}\mathbf{h}_1$	sq
10	$K(Z, X_m)\mathbf{g}_m$	mp
10	$\mathbf{M}\mathbf{h}_1$	pq

Table 2: Computational cost analysis of Algorithm 1 for processing batch t for $kT < t \leq (k+1)T$ for some $k \in \mathbb{N}$.

The cost of processing batch t without the post-processing adds up to $2mp + 2ms + 2sq + pq + m^2(t - kT - 1)$ flops.

B Computational complexity comparison

We assume that EigenPro4 is processing T batches of data at once before running the post-processing step of projection. Here we show we calculated the optimal value of T .

Cost for processing t^{th} batch of data. For a some $k \in \mathbb{N}$, let $kT < t \leq (k+1)T$. See Table 2.

Cost of processing T batches of data before post-processing The total cost for processing T batches $t = kT + 1$ to $t = (k+1)T$ before the projection is the sum of the above

$$T(2mp + 2ms + 2sq + pq) + m^2 \sum_{t=kT+1}^{(k+1)T} (t - kT - 1) = T(2mp + 2ms + 2sq + pq) + m^2 \frac{T(T-1)}{2} \quad (25)$$

Average of processing T batches of data with post-processing Assuming the post processing involves T_{ep2} epochs of EigenPro 2, the average cost of processing T batches is

$$\frac{T(2mp + 2ms + 2sq + pq) + m^2 \frac{T(T-1)}{2} + p^2 T_{\text{ep2}}}{T} \quad (26)$$

A simple calculation shows that

$$T^* = \frac{p}{m} \sqrt{2T_{\text{ep2}}} \quad (27)$$

minimizes the average time above. The average cost of processing a batch is thus

$$2mp(1 + \sqrt{2T_{\text{ep2}}}) + 2ms + 2sq + pq \quad (28)$$

Illustration for delayed projection for $T = 4$.

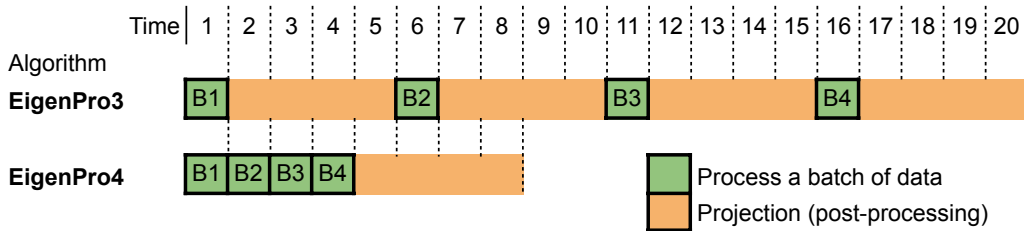


Figure 4: **Design of EigenPro4.** An illustration of how batches of data are processed by the two algorithms. EigenPro3 involves an expensive *projection* step when processing every batch of data. EigenPro4 waits for multiple batches to be processed before running the projection step for all of them together. This reduces the amortized cost for processing each batch.

Comparison between EigenPro 4 and EigenPro 3. Figure 5 shows how EigenPro 4 and EigenPro 3 perform over training iterations. EigenPro 4 accuracy improves between projections and drops after

each projection step. While EigenPro 3 projects at every step, EigenPro 4 maintains comparable accuracy with fewer projections. The left panel of Figure 5 confirms that both methods reach similar final accuracy, while the right panel shows EigenPro 4 significant speed advantage. With continued training, EigenPro 4 accuracy drops from projections become progressively smaller.

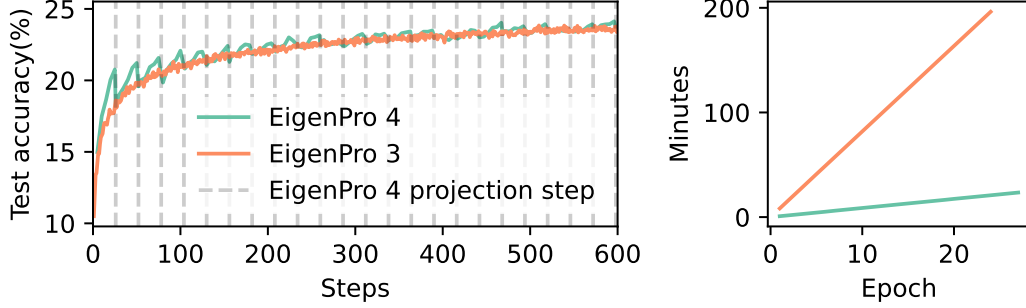


Figure 5: Performance and computational time comparison between EigenPro 4.0 ($T = 11$) and EigenPro 3.0 (equivalent to $T = 1$), highlighting the impact of the projection step on the performance of EigenPro 4.0. The detail of the experiment can be found in Appendix C.

Algorithm	setup	FLOPS per batch*	Memory
EigenPro 4.0	$O(s^2q)$	$2mp(1 + \sqrt{2T_{\text{ep}2}}) + 2ms + 2sq + pq$	$s^2 + p(1 + \sqrt{2T_{\text{ep}2}})$
EigenPro 3.0	$O(s^2q)$	$2mp + p^2T_{\text{ep}2} + 2ms + 2sq + pq$	$s^2 + p$
Falkon	$O(p^3)$	$2mp$	p^2

Table 3: **Comparing complexity of algorithms.** Number of training samples n , number of model centers p , batch size m , Nyström sub-sample size s , preconditioner level q . Here we assumed only a constant number of epochs of EigenPro 2.0 is needed for large scale experiments. Cost of kernel evaluations and number of classes are assumed to be $O(1)$, also it is reasonable to assume $p \gg s \gg q$.

* FLOPS per iteration reported are amortized over multiple batches processed.

C Experiments Results

C.1 Computational resources used

This work used the Extreme Science and Engineering Discovery Environment (XSEDE) [23]. We used machines with NVIDIA-V100, NVIDIA-A100 and NVIDIA-A40 GPUs, with a V-RAM up to 1.3 T, and 8x cores of Intel(R) Xeon(R) Gold 6248 CPU @ 2.50GHz with a RAM of 100 GB. Note that we had 1.3T of RAM for just one experiment CIFAR5M*, for the rest of experiments we were constrained with 400G of RAM.

C.2 Datasets

We perform experiments on these datasets: (1) CIFAR10, [11], (2) CIFAR5M, [17], (3) ImageNet, (4) Webvision, [12], and (5) librispeech.

CIFAR5M. In our experiments, we utilized both raw and embedded features from the CIFAR5M data-set. The embedded features were extracted using a MobileNetv2 model pre-trained on the ImageNet data-set, obtained from *timm* library [25]. We indicate in our results when pre-trained features were used by adding an asterisk (*) to the corresponding entries.

ImageNet. In our experiments, we utilized embedded features from the ImageNet data-set. The embedded features were extracted using a MobileNetv2 model pre-trained on the ImageNet dataset, obtained from *timm* library [25]. We indicate in our results when pre-trained features were used by adding an asterisk (*) to the corresponding entries.

Webvision. In our experiments, we utilized embedded features from the Webvision data-set. The embedded features were extracted using a ResNet-18 model pre-trained on the ImageNet dataset, obtained from *timm* library [25]. Webvision data set contains 16M images in 5k classes. However, we only considered the first 2k classes.

Librispeech. Librispeech [18] is a large-scale (1000 hours in total) corpus of 16 kHz English speech derived from audio books. We choose the subset train-clean-100 and train-clean-300 (5M samples) as our training data, test-clean as our test set. The features are got by passing through a well-trained acoustic model (a VGG+BLSTM architecture in [8]) to align the length of audio and text. It is doing a 301-wise classification task where different class represents different uni-gram [10]. The implementation of extracting features is based on the ESPnet toolkit [24].

C.3 Experiments details

Figure 1 This experiment used CIFAR5M* data set, where embedding has been generated using a pre-trained mobile-net network mentioned earlier. This is the only experiment that we had access to 1.3T of VRAM. We set the bandwidth to 5.0 and use 1k Nystrom samples with preconditioning level of size 100. We used float16 for this experiment.

Figure 5 This experiment has been run over Webvision data set with extracted embedding through Resnet18. The model size here is set to 100k number of centers. The bandwidth used is 5.0, 1k Nystrom samples with preconditioning level of size 100. We used float16 for this experiment.

Figure 3 We follow the setting in [1]. The bandwidth used here is 20 for Librispeech and Webvision and 16 for ImageNet. Here again we used extracted feature of these datasets mentioned earlier. The precision used here is float32. with 10k Nystrom samples with preconditioning level of size 1000.

Table 1 For all datasets here we used bandwidth of 5.0 with 1k Nystrom samples with preconditioning level of size 100. We used float16 for all dataset except for Librispeech where we used float32. Further, we note that *Falcon* latest library ran out of GPU memory for model sizes larger than 256000 number of centers that is the reason we could not run it for 256000. And as mentioned for model sizes 521000 and above the algorithm has inherent quadratic scaling with respect to model size and we ran out of VRAM. In the plot we refer to both of these memory issues as OOM.

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [Yes]

Justification: We demonstrate the efficiency of our method across multiple datasets and scales: Figures 1 and 2 illustrate performance trends, and Table 1 provides an extensive quantitative evaluation, showing speedups of up to $\times 600$.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: We provide a theoretical convergence analysis for the exact version of our method in Appendix A. The full algorithm is based on an efficient approximation of this exact version. Although a formal theoretical analysis of the approximation is not included, we validate its effectiveness extensively through experiments on diverse datasets and scales. Establishing a complete theoretical understanding of the approximation involves nontrivial technical challenges, which we consider an important direction for future work beyond the scope of this paper. We addressed the limitation in 3.6.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.

- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: The theoretical result is presented in Appendix A, where we include all necessary assumptions and provide a complete proof.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We provide all necessary details for reproducing our experiments in Appendix C. We will release our GitHub repository upon acceptance and have also included a .zip file version of the code in supplementary for the review process.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case

of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.

- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We use publicly available datasets and provide all necessary details for reproducing our experiments in Appendix C. The code for our algorithm is included in the supplemental material for review process.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: We provide all the necessary details for reproducing our experiments in Appendix C

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: Our experimental setup ensures stability and reproducibility by using the same training/test splits, center selections, and random seeds across all methods. As is typical in kernel methods, there is no randomness from model or weight initialization, and the only minor source of stochasticity arises from optimization, which is consistently controlled. Because all methods are evaluated under identical conditions, traditional error bars would not provide meaningful additional information. We validate the effectiveness and robustness of our method through extensive experiments on a wide range of datasets, consistently demonstrating significant speedups with comparable or better performance.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: We provide all necessary details in Appendix C

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: We followed the NeuIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [NA]

Justification: We believe there is no immediate societal impact, either positive or negative, as this work focuses on foundational algorithmic improvements to kernel methods. The contributions are methodological in nature and not tied to any specific application domain or deployment scenario. As such, we mark this question as not applicable.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out

1006 that a generic algorithm for optimizing neural networks could enable people to train
1007 models that generate Deepfakes faster.

- 1008 • The authors should consider possible harms that could arise when the technology is
1009 being used as intended and functioning correctly, harms that could arise when the
1010 technology is being used as intended but gives incorrect results, and harms following
1011 from (intentional or unintentional) misuse of the technology.
- 1012 • If there are negative societal impacts, the authors could also discuss possible mitigation
1013 strategies (e.g., gated release of models, providing defenses in addition to attacks,
1014 mechanisms for monitoring misuse, mechanisms to monitor how a system learns from
1015 feedback over time, improving the efficiency and accessibility of ML).

1016 11. Safeguards

1017 Question: Does the paper describe safeguards that have been put in place for responsible
1018 release of data or models that have a high risk for misuse (e.g., pretrained language models,
1019 image generators, or scraped datasets)?

1020 Answer: [NA]

1021 Justification: This work does not involve the release of pretrained models, generative systems,
1022 or scraped datasets that pose a high risk of misuse. Therefore, no specific safeguards are
1023 necessary, and we mark this question as not applicable.

1024 Guidelines:

- 1025 • The answer NA means that the paper poses no such risks.
- 1026 • Released models that have a high risk for misuse or dual-use should be released with
1027 necessary safeguards to allow for controlled use of the model, for example by requiring
1028 that users adhere to usage guidelines or restrictions to access the model or implementing
1029 safety filters.
- 1030 • Datasets that have been scraped from the Internet could pose safety risks. The authors
1031 should describe how they avoided releasing unsafe images.
- 1032 • We recognize that providing effective safeguards is challenging, and many papers do
1033 not require this, but we encourage authors to take this into account and make a best
1034 faith effort.

1035 12. Licenses for existing assets

1036 Question: Are the creators or original owners of assets (e.g., code, data, models), used in
1037 the paper, properly credited and are the license and terms of use explicitly mentioned and
1038 properly respected?

1039 Answer: [Yes]

1040 Justification: To the best of our knowledge, all datasets and codebases used in this work
1041 have been properly cited and used in accordance with their licenses and terms of use.

1042 Guidelines:

- 1043 • The answer NA means that the paper does not use existing assets.
- 1044 • The authors should cite the original paper that produced the code package or dataset.
- 1045 • The authors should state which version of the asset is used and, if possible, include a
1046 URL.
- 1047 • The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- 1048 • For scraped data from a particular source (e.g., website), the copyright and terms of
1049 service of that source should be provided.
- 1050 • If assets are released, the license, copyright information, and terms of use in the
1051 package should be provided. For popular datasets, paperswithcode.com/datasets
1052 has curated licenses for some datasets. Their licensing guide can help determine the
1053 license of a dataset.

- 1054 • For existing datasets that are re-packaged, both the original license and the license of
1055 the derived asset (if it has changed) should be provided.
- 1056 • If this information is not available online, the authors are encouraged to reach out to
1057 the asset's creators.

1058 **13. New assets**

1059 Question: Are new assets introduced in the paper well documented and is the documentation
1060 provided alongside the assets?

1061 Answer: [Yes]

1062 Justification: We introduce new code implementing our algorithm, which we include as
1063 a .zip file in the supplemental material for the review process. The code is documented
1064 with instructions for reproduction and GitHub repository will be released publicly upon
1065 acceptance.

1066 Guidelines:

- 1067 • The answer NA means that the paper does not release new assets.
- 1068 • Researchers should communicate the details of the dataset/code/model as part of their
1069 submissions via structured templates. This includes details about training, license,
1070 limitations, etc.
- 1071 • The paper should discuss whether and how consent was obtained from people whose
1072 asset is used.
- 1073 • At submission time, remember to anonymize your assets (if applicable). You can either
1074 create an anonymized URL or include an anonymized zip file.

1075 **14. Crowdsourcing and research with human subjects**

1076 Question: For crowdsourcing experiments and research with human subjects, does the paper
1077 include the full text of instructions given to participants and screenshots, if applicable, as
1078 well as details about compensation (if any)?

1079 Answer: [NA]

1080 Justification: This work does not involve any crowdsourcing or research with human subjects,
1081 so this question is not applicable.

1082 Guidelines:

- 1083 • The answer NA means that the paper does not involve crowdsourcing nor research with
1084 human subjects.
- 1085 • Including this information in the supplemental material is fine, but if the main contribu-
1086 tion of the paper involves human subjects, then as much detail as possible should be
1087 included in the main paper.
- 1088 • According to the NeurIPS Code of Ethics, workers involved in data collection, curation,
1089 or other labor should be paid at least the minimum wage in the country of the data
1090 collector.

1091 **15. Institutional review board (IRB) approvals or equivalent for research with human
1092 subjects**

1093 Question: Does the paper describe potential risks incurred by study participants, whether
1094 such risks were disclosed to the subjects, and whether Institutional Review Board (IRB)
1095 approvals (or an equivalent approval/review based on the requirements of your country or
1096 institution) were obtained?

1097 Answer: [NA]

1098 Justification: This work does not involve any research with human subjects, and therefore
1099 no IRB or equivalent approval was necessary.

1100 Guidelines:

- 1101 • The answer NA means that the paper does not involve crowdsourcing nor research with
1102 human subjects.
- 1103 • Depending on the country in which research is conducted, IRB approval (or equivalent)
1104 may be required for any human subjects research. If you obtained IRB approval, you
1105 should clearly state this in the paper.
- 1106 • We recognize that the procedures for this may vary significantly between institutions
1107 and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the
1108 guidelines for their institution.
- 1109 • For initial submissions, do not include any information that would break anonymity (if
1110 applicable), such as the institution conducting the review.

1111 16. Declaration of LLM usage

1112 Question: Does the paper describe the usage of LLMs if it is an important, original, or
1113 non-standard component of the core methods in this research? Note that if the LLM is used
1114 only for writing, editing, or formatting purposes and does not impact the core methodology,
1115 scientific rigorousness, or originality of the research, declaration is not required.

1116 Answer: [NA]

1117 Justification: This work does not involve the use of large language models as an important,
1118 original, or non-standard component of the core methodology. Any use of language models
1119 was limited to minor writing assistance and does not impact the scientific contribution of the
1120 paper.

1121 Guidelines:

- 1122 • The answer NA means that the core method development in this research does not
1123 involve LLMs as any important, original, or non-standard components.
- 1124 • Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>)
1125 for what should or should not be described.