

Competitive Physics Informed Networks



(/pdf?id=z9SIj-IM7tn)

Qi Zeng (/profile?id=~Qi_Zeng2), Yash Kothari (/profile?id=~Yash_Kothari1), Spencer H Bryngelson (/profile?id=~Spencer_H_Bryngelson1), Florian Tobias Schaefer (/profile?id=~Florian_Tobias_Schaefer1)

Published: 01 Feb 2023, Last Modified: 21 Oct 2023 ICLR 2023 poster Readers:

Everyone Show Bibtex Show Revisions (/revisions?id=z9SIj-IM7tn)

Keywords: Physics informed learning, multi-agent games, Lagrange multipliers, partial differential equations

TL;DR: We introduce competitive physics informed networks where two neural networks solve a partial differential equation by playing a zero-sum game.

Abstract: Neural networks can be trained to solve partial differential equations (PDEs) by using the PDE residual as the loss function. This strategy is called "physics-informed neural networks" (PINNs), but it currently cannot produce high-accuracy solutions, typically attaining about 0.1% relative error. We present an adversarial approach that overcomes this limitation, which we call competitive PINNs (CPINNs). CPINNs train a discriminator that is rewarded for predicting mistakes the PINN makes. The discriminator and PINN participate in a zero-sum game with the exact PDE solution as an optimal strategy. This approach avoids squaring the large condition numbers of PDE discretizations, which is the likely reason for failures of previous attempts to decrease PINN errors even on benign problems. Numerical experiments on a Poisson problem show that CPINNs achieve errors four orders of magnitude smaller than the best-performing PINN. We observe relative errors on the order of single-precision accuracy, consistently decreasing with each epoch. To the authors' knowledge, this is the first time this level of accuracy and convergence behavior has been achieved. Additional experiments on the nonlinear Schrödinger, Burgers', and Allen-Cahn equation show that the benefits of CPINNs are not limited to linear problems.

Anonymous Url: I certify that there is no URL (e.g., github page) that could be used to find authors' identity.

No Acknowledgement Section: I certify that there is no acknowledgement section in this submission for double blind review.

Supplementary Material: zip (/attachment?id=z9SIj-IM7tn&name=supplementary_material)

Code Of Ethics: I acknowledge that I and all co-authors of this work have read and commit to adhering to the ICLR Code of Ethics

Submission Guidelines: Yes

Please Choose The Closest Area That Your Submission Falls Into: Machine Learning for Sciences (eg biology, physics, health sciences, social sciences, climate/sustainability)

Community Implementations: 4 code implementations (<https://www.catalyzex.com/paper/arxiv:2204.11144/code>)

Revealed to Qi Zeng, Yash Kothari, Spencer H Bryngelson, Florian Tobias Schaefer

Published: 01 Feb 2023, Last Modified: 21 Oct 2023 ICLR 2023 poster

Resubmission: Yes

Student Author: Yes

Add

Withdraw

Reply Type: Author: Visible To:

23 Replies

Hidden From:

[–] Some questions and suggestions

Shuhao Cao (/profile?id=~Shuhao_Cao1)

02 Feb 2023 ICLR 2023 Conference Paper5528 Public Comment Readers: 

Everyone Show Revisions (/revisions?id=jcazKcZjxB)

Comment:

Very interesting paper.

- It would be nice to see an ablation study to isolate the quantitative contributions of different components attributing to the accuracy improvement (a) the condition number change ($\kappa \rightarrow \sqrt{\kappa}$), (b) ACGD vs old schoolers (such as proximal) for saddle point problems, (c) CPINN and PINN using the same number of layers and width (in the current paper, CPINN is either wider or deeper, which translates to easier optimization according to lottery ticket).
- How a vanilla feedforward NN representation of "solution" can be written as (9) in the form of a basis expansion in terms of *parameters* should be elaborated in my humble opinion.

[–] Thank you for your interest!

ICLR 2023 Conference Paper5528 Authors Florian Tobias Schaefer (/profile?id=~Florian_Tobias_Schaefer1) (privately revealed to you)

02 Feb 2023 ICLR 2023 Conference Paper5528 Official Comment Readers: 

Everyone Show Revisions (/revisions?id=drtd9Na8Gs)

Comment:

Thank you for your interest in our work!

Regarding your first point: We determined the architectures for both PINNs and CPINNs by trying a wide range of architectures for both methods and picking whichever is best. The reason we use different architectures for the two methods is that they were their respective best-performing ones. Figure 6 provides an ablation study when using different optimizers. We observe that the classical extra gradient method on CPINNs, just like CGD, improves over PINN with SGD. These methods all use fixed and isotropic learning rates and are therefore natural to compare against one another. When introducing adaptive learning rates (XAdam, ACGD, Adam), we see that this improves performance on all methods but extra gradient, as we observe extra adam to be unstable. Our conclusion from these results is that CPINN can improve over PINN also when using other saddle point methods, such as extragradient. However, CGD is able to maintain stable training even with adaptive learning rates (ACGD), which allows it to overall achieve the best performance.

Regarding your second point: We do not claim that a generic neural network representation of a solution of a PDE can be meaningfully written in the summation form described in the paper. We only point out that the classical petrov-galerkin methods can be interpreted as a very peculiar "neural network architecture", thus allowing us to relate some intuition from classical numerical analysis to (C)PINNs.

[–] Paper Decision

ICLR 2023 Conference Program Chairs

20 Jan 2023 ICLR 2023 Conference Paper5528 Decision Readers: 

Everyone Show Revisions (/revisions?id=3W4A3JugPP)

Decision: Accept: poster

Metareview: Summary, Strengths And Weaknesses:

This paper introduces competitive physics-informed networks where two neural networks solve a partial differential equation by playing a zero-sum game. The formulation is built using a weak form. The authors claim that this weak form have smaller condition number.

The paper originally received diverse reviews, and we had a dedicated discussion among all the reviewers. The author response also helped a lot. Eventually, as a consensus, we believe this paper has more strengths than weaknesses, and it would be good to accept it to the conference.

Summary Of AC-reviewer Meeting:

QL5E:

I think this paper is different from <https://arxiv.org/pdf/2009.04544.pdf> (<https://arxiv.org/pdf/2009.04544.pdf>) in the following sense:

- the cpinn objective to solve $Ax = b$ is $\min \max \text{weight}(Ax-b)$, *the self-active pink is* $\min \max ||\text{weight}(Ax-b)||^2$
- the later formulation still suffers from the double A problem as stated in the paper
- In adaptive pinn, all the weight is positive, but in cpinn the weight can have negative,
- the introduction of two approach are different, in this paper, they intro the weight to approximate $||Ax-b||^2$ using only a single A

In my opinion, we can't state that both method introduce an adversarial weight thus they are similar. The mathematics behind the two papers differs.

At the same time, this paper have a nice discussion on the intuition behind the acceleration of optimization, although some of them contradicted with my own previous paper. I'm happy to see debate in this area. The claims made in author's rebuttal also convince me the algorithm will work in some cases.

hCa9:

I read the authors' responses and others' reviews. I hold my position that this paper is 'marginally above the acceptance threshold'. The paper proposes a rather straightforward augment to PINN, which I am not particularly excited about. The execution of the ideas looks OK to me. Considering that this probably reflects the hard work of a Ph.D. student, I lean towards accepting the paper as an encouragement.

aueW:

I updated the comments two days ago and wanna see the replies from the authors. The paper achieves good accuracy, and I want to check if good results are obtained by their special design. The performance should be achieved by the difference that Reviewer QL5E mentioned, not by other reasons. To check it, some experiments of using neural networks in the adaptive PINN paper may be good proof. More importantly, there is an issue that the authors compare the results between PINN with different optimizers, and I want to clarify if it is fair. Hence I proposed several concerns in the updated review.

There is nothing wrong with Reviewer QL5E, and we all admit the difference. If it is the difference making it different, I would appreciate the novelty. There is no problem if the paper ends up being accepted as encouragement, but I want to see how they would reply to my concerns.

aueW:

The authors provided a detailed response to my latest comments and I think most of my previous concerns have been addressed.

Justification For Why Not Higher Score:

There are still concerns on the technical novelty - the augment to PINN is quite straightforward, although the execution of the idea looks good.

Justification For Why Not Lower Score:

After the discussion, the reviewers believe the ICLR audience should be interested in seeing this paper.


Note From PC:

if the above contains the word "oral" or "spotlight" please see: "oral" presentation means -> notable-top-5% and "spotlight" means -> notable-top-25%. As stated in our emails, we are disassociating presentation type from AC recommendations

[–] Response to all reviewers

ICLR 2023 Conference Paper5528 Authors Florian Tobias Schaefer (/profile?id=~Florian_Tobias_Schaefer1) (privately revealed to you)

15 Nov 2022 (modified: 16 Nov 2022) ICLR 2023 Conference Paper5528 Official

Comment Readers:  Everyone Show Revisions (/revisions?id=Sti2vkM--J4)

Comment:

We thank all reviewers for reading our manuscript. We were delighted that the reviewers commented positively on the quality of our results and their exposition. We especially thank reviewers aueW and QL5E for their valuable criticism that will help us improve the paper.

In the remainder of this post, we will respond to what we perceive as the main points of critique. We provide more detailed responses to each reviewer in a separate post.

- Reviewer aueW is concerned with the relationship of our work to the prior work that uses an adaptive method to choose the weights of the different PINN loss terms. This line of work is discussed at the end of section 4. Choosing weights for a positive loss function is fundamentally different from making bets on the sign of the residual. In the former case, the optimal strategy of the maximizing player is to move all weights to infinity. In the latter case, the optimal strategy of the maximizing player depends on the mistakes of the minimizing one. To prevent future misunderstanding, we now make this point after the introduction of CPINNs in Section 2.2.
- Reviewers aueW and QL5E suggest a comparison to weak adversarial networks (WAN) in order to better disentangle the effects of the formulation (CPINN) from the optimizer (ACGD). We have added a comparison with (WAN), trained with both Adam and ACGD, to Figure 6. We were unable to achieve accuracy comparable to CPINN with ACGD despite trying different sets of activation functions, loss functions (dropping the logarithm as suggested in section 3.5 of Zang et al.), and training modes (minibatching vs fixed batch). Nevertheless, we observe that ACGD improves the stability of WAN training. We have also added an additional paragraph to section 4 that highlights the differences between CPINN and WAN.
- Reviewers aueW and QL5E both express concern about the additional complexity introduced by the minimaximization. We regret this situation, because we believe we did account accounted for this complexity in an appropriate way, there was just a textual misunderstanding. In figures 1, 3, 4, 5, and 6, we provide the relative error as a function of not just the iterations, but also of the number of forward passes. The forward passes count the computation of every single gradient or Hessian-Vector product with respect to either of the networks, including those needed by the linear system solve required when computing the CGD update. Thus, they account for the additional computation that the CPINN method requires.

[–] Official Review of Paper5528 by Reviewer aueW

ICLR 2023 Conference Paper5528 Reviewer aueW

27 Oct 2022 (modified: 29 Nov 2022) ICLR 2023 Conference Paper5528 Official

Review Readers:  Everyone Show Revisions (/revisions?id=_HMPI_5FMF)

Summary Of The Paper:

This paper presents an adversarial approach that overcomes this limitation, called competitive PINNs (CPINNs). CPINNs train a discriminator that is rewarded for predicting mistakes the PINN makes. The discriminator and PINN participate in a zero-sum game with the exact PDE solution as an optimal strategy. This approach avoids squaring the large condition numbers of PDE discretizations, which is the likely reason for failures of previous attempts to decrease PINN errors even on benign problems. The numerical results show the accuracy.

Strength And Weaknesses:

Strength: By introducing another neural network as the point weight function and following with a min-max optimization, the solution of the PDEs can be more accurate.

Weakness: Introducing a new neural network besides the solution neural network means much more complex for training and becomes much trickier on different tasks. And the weakness is summarized as follows.

1. there is one similar paper in 2020 as McClenny, Levi, and Ulisses Braga-Neto. "Self-adaptive physics-informed neural networks using a soft attention mechanism." arXiv preprint arXiv:2009.04544 (2020). In that work, they use mask functions for the point weight, and there is no significant difference using a neural network in this work, and the min-max procedure is the same; thus, the novelty is not much. The story is under the structure of the games but still can not change the reality of using a neural network for the point weight adversarial update. No intrinsic difference.
2. The analysis in section 2.3 is not very firm and too simple, sometimes naive, and the analysis of the neural network is just for a linear mapping which should be at least a two-layer neural network if you really want to explain something. As far as I know, in traditional numerical PDEs, the linear combination of the basis is, of course, a very common and efficient way. And it works well even using the vanilla schemes, such as the finite element method. Further, what really makes the so-called neural PDE different is the approximation by nonlinear neural networks, and it is far not enough to use one linear combination to analyze the neural PDE. There is plenty of work for analyzing PINNs, a direct search may help your analysis.
3. The results of the experiments are good but also tricky as a new neural network besides the solution neural network is introduced. More neural networks with a min-max optimization will increase the difficulties in training, and it is already complex compared with the traditional numerical solvers. If you can prove in high dimensional problem, the method still achieves good performance, just like Bao et al. in the weak adversarial neural network (Zang, Yaohua, et al. "Weak adversarial networks for high-dimensional partial differential equations." Journal of Computational Physics 411 (2020): 109409.), it is much better. Further, in the experiments, such as nearly all of the L2 error figures, the L2 relative error curve with respect to the iteration is not stable, but the experiments are just stopped at best. See Figure 4; there is still a very large oscillation. And Figure 5 has a big shock and etc. Also, it is hard to determine which one enhanced the performance, the optimizer or the structure, see Table 1.
4. The terms are a little confusing for those who are not familiar with neural PDEs and may lead misunderstandings. The work is very direct with adversarial updating point weight in the area of Neural PDE, but the terms, such as bets, game, etc., will confuse the readers. A direct statement would make things much easier.

Clarity, Quality, Novelty And Reproducibility:

Clarity: Not clear. The work is very direct with adversarial updating point weight in the area of Neural PDE, but the terms, such as bets, game, etc., will confuse the readers.

Quality: Good. But the stop iteration is not natural. See #3 above for the detail.

Novelty: Weak. See #1 above.

Reproducibility: Good. The authors provide the code, but the method seems hard to use in other cases for requiring careful training.

Summary Of The Review:

Based on the weakness, I reject the paper.

Correctness: 3: Some of the paper's claims have minor issues. A few statements are not well-supported, or require small changes to be made correct.

Technical Novelty And Significance: 3: The contributions are significant and somewhat new. Aspects of the contributions exist in prior work.

Empirical Novelty And Significance: 3: The contributions are significant and somewhat new. Aspects of the contributions exist in prior work.

Flag For Ethics Review: NO.


Recommendation: 6: marginally above the acceptance threshold

Confidence: 5: You are absolutely certain about your assessment. You are very familiar with the related work and checked the math/other details carefully.

[–] Response to reviewer aueW

ICLR 2023 Conference Paper5528 Authors Florian Tobias Schaefer (/profile?id=~Florian_Tobias_Schaefer1) (privately revealed to you)

15 Nov 2022 (modified: 16 Nov 2022) ICLR 2023 Conference Paper5528 Official

Comment Readers:  Everyone Show Revisions (/revisions?id=SFGH4uyiuH)

Comment:

Thank you for taking the time to read our paper. Below are our responses to your four main critiques.

1. The initial submission of this manuscript already referenced work similar to that of McClenny et al. in Sections 1 and 4. There, we discussed the closely related works by Wang et al. and van der Meer et al. As already described in these sections, these previous works learn positive weights to multiply with a positive loss function. Thus, the optimal strategy of the max player is always to set all weights to infinity, independent of the actions of the minimizing player. In contrast, our method using signed bets multiplied with the residual requires the maximizer to learn the mistakes of the minimizer. Thus, even if one were to use a neural network that allowed for spatially varying weights (McClenny et al. do not do this), the resulting formulation would be different. We understand that this might be a subtle point. In order to avoid future misunderstanding, we added the above comments to section 2.2, right after introducing CPINNs.
2. As the reviewer points out, many existing works attempt to a theoretical analysis of PINNs. However, the rigorous understanding of the PINN methodology is still limited despite their being studied since 2017 (with earlier work dating back to 1998). Given this precedence, we consider a rigorous understanding of CPINNs to be beyond the scope of this paper. Instead, we focus on a model problem that motivates our approach and, we believe, cuts to the key problem associated with training PINNs. Although limited in this particular way, our theoretical contribution is in line with most other physics-informed learning papers in top-tier conferences that introduce novel methodologies and study them empirically, including those of McClenny et al. and Zang et al. mentioned by the reviewer.
3. The plots of the error vs "forward passes" account for the cost due to every single gradient computation or Hessian vector product. We thus account fairly for the additional cost incurred by competitive gradient descent, which is necessitated by the min-max formulation. Besides this cost, the general claim that "More neural networks with a min-max optimization will increase the difficulties in training" would need to be substantiated further. This has not been our experience over the course of this project and we do not see a fundamental reason why it should be the case. In ongoing preliminary experiments on high-dimensional problems, we have so far not been able to obtain meaningful improvements with the CPINN method. Doing so remains an interesting direction of future research, but our current work focuses on attaining high accuracy on low-dimensional PDEs, which comprise a large proportion of PDE applications. We point out that the only comparison of weak adversarial networks to PINNs on a high-dimensional problem in Zang et al. in Figure 2 (b), which shows almost no performance difference. We were forced to stop the training process due to the required computational resources. Note that the number of iterations and forward passes are plotted on a log scale. As shown, the experiments take a few days to train. Adding a single additional major axis tick would require training for months or longer.
4. We hope that the additional explanation contrasting bets with weights in section 2.2 will be helpful in highlighting our reasons for introducing this terminology. As described there, their difference is crucial to understanding the relationship of our method to those based on reweighting. We, therefore, consider this terminology to be useful.

[–] Concerns

ICLR 2023 Conference Paper5528 Reviewer aueW

24 Nov 2022 (modified: 24 Nov 2022) ICLR 2023 Conference Paper5528

Official Comment Readers:  Everyone Show Revisions (/revisions?id=RV4nNoqmp4)

Comment:


Thanks for your reply. Here are some concerns, and I am happy if you could have further discussion. The main concern for me is the reason for achieving high accuracy.

1. "Thus, even if one were to use a neural network that allowed for spatially varying weights (McClenny et al. do not do this), the resulting formulation would be different." Can you check the performance of the scheme? I mean, this would prove that the signed bets multiplied by the residual make something different. You also said the "theoretical" analysis is not rigorous and I indeed call for more proof maybe more experiments are needed.
2. Okay, if it is just a not rigorous explanation, that makes sense.
3. "We were forced to stop the training process due to the required computational resources. As shown, the experiments take a few days to train." As the computational cost is very high, an illustration of the complex problem that the traditional numerical methods can't afford would be much more reasonable, for example, complex boundary, high dimension (maybe the complex boundary problem is easier to take). Further, I am still concerned about the stopping time when the curve is still oscillating. You mentioned that because of the computational cost, it is enough to stop. Can I say that you just need good accuracy, so you can just stop at a good one? If it is, that makes sense to me.
4. The terminologies are not essential for the explanation, but if you insist on contrasting bets with weights, I accept it.
5. You may miss the question at the end of the third point: "Also, it is hard to determine which one enhanced the performance, the optimizer or the structure; see Table 1." I restated the question more clearly here. I mean, you compare CPINN + XCGD (X means different schemes of CGD) with PINN + Adam and SGD; is it fair? Further, Table 1 shows if you choose CPINN + XAdam the algorithm fails, and CPINN + XSGD didn't outperform much. That makes me confused whether the accuracy achieved is because of the XCGD.

[–] **Thank you for engaging in the discussion!**

ICLR 2023 Conference Paper5528 Authors Florian Tobias Schaefer (/profile?id=~Florian_Tobias_Schaefer1) (privately revealed to you)

29 Nov 2022 ICLR 2023 Conference Paper5528 Official

Comment Readers:  Everyone Show Revisions (/revisions?id=PaP2j4xm6Wq)

Comment:

Thank you for your response. We appreciate your effort to participate in the conversation and help us improve the paper. Please see our response below.

1. To perform the requested comparison, we implemented an adaptive weighting scheme where the product of the squares of the discriminator output and the PDE residual gives the loss. As shown here (<https://imgur.com/a/8Btkp0d>), we found it to perform similarly to weak adversarial networks (WAN) [Zang et al. 2019] and, thus, significantly worse than the CPINN. From a theoretical perspective, this is not so surprising because adaptive weighting is similar to WAN, though we agree with the reviewer that the experiments provided above are helpful in ascertaining this claim. The main difference between adaptive weighting and WANs is that WAN uses a logarithmic transformation of the loss, that WAN does not use adaptive weights for the boundary loss, and that the WAN loss is normalized by the magnitude of the discriminator output. This shows through both experiments and theoretical considerations that WAN and adaptive weighting are fundamentally different from the CPINN formulation, despite their superficial similarity.
2. Thank you for your understanding and for helping us to clarify the description.
3. *(Regarding comparison to traditional methods:)* We agree that finding cases where physics-informed networks can outperform classical methods is an important research direction, and that PDEs on complicated or high-dimensional domains are promising candidates. However, in this work we focus on a pervasive flaw of existing physics-informed networks that already appears in the relatively simple problems studied by a strikingly large amount of research works: Their inability to produce high-accuracy solutions even on these seemingly benign problems. We believe that resolving this limitation is a necessary and important stepping stone towards solving more challenging PDEs using neural networks and doing so performantly. *(Regarding stopping criteria:)*

You are correct that, in practice, one can stop training at any point once a good solution is reached. Of course, the L^2 error compared to the exact solution will generally be unavailable. However, we found that the norm of the PDE residual in the training points strongly correlates with the L^2 error (for both PINN and CPINN). Therefore, a practical approach is to choose the iteration that achieves the lowest PDE residual. This criterion can probably be improved by including the PDE residual at "holdout points" that were not used during training. We plot the results for "as long as we could train" instead of imposing a superficial stopping criterion to provide the maximum amount of information about the training behavior of the methods under consideration.

4. Thank you for your understanding.


5. This is an important point, and we regret not addressing it more directly the first time. We provided our reasoning for why our improvements are due to the CPINN formulation and not just due to the choice of the optimizer in Section 3.6 of the revised paper. If appropriate, we are glad to expand it to the extent that space permits (or, alternatively, in an appendix). Our reasoning is as follows. The closest thing to applying BCGD to PINN would be to introduce an inconsequential "dummy agent" to obtain a problem of the form $\min_x \max_y f(x)$ and solve it with BCGD. The resulting algorithm is identical to GD. Similarly, applying ACGD to PINN is equivalent to using Adam. Indeed, the core tenet of [Schaefer and Anandkumar, 2019] is that CGD "is a natural generalization of gradient descent to the two-player setting." Thus, comparing CPINN + BCGD to PINN + GD and CPINN + ACGD to PINN + Adam strikes us as a fair comparison. As shown in Figure 6, CPINN significantly outperforms PINN in either of these cases. This justifies our claim that the improvement is due to the CPINN formulation and not just a particular optimizer. Additional evidence for this claim is provided by our experiments with WAN (in Figure 6) and adaptive weights (in our response 1. above). These experiments use ACGD, just as CPINN. Yet, CPINN achieves significantly higher accuracy.

We hope that we have now addressed all your outstanding concerns. Please let us know if you require further clarification.

[–] **Most of my concerns are addressed**

ICLR 2023 Conference Paper5528 Reviewer aueW

29 Nov 2022 ICLR 2023 Conference Paper5528 Official

Comment Readers:  Everyone Show Revisions (/revisions?id=i5z5fmhtb7d)

Comment:

Most of my concerns are addressed, thanks. "we are glad to expand it to the extent that space permits (or, alternatively, in an appendix)" is welcomed.

[–] **Official Review of Paper5528 by Reviewer 1E2V**

ICLR 2023 Conference Paper5528 Reviewer 1E2V

24 Oct 2022 ICLR 2023 Conference Paper5528 Official Review Readers: 

Everyone Show Revisions (/revisions?id=CVFx99xcQk)

Summary Of The Paper:

This paper introduces a new model called CPINN, which is an agent based approach to the machine learning solution of PDEs. CPINN is an adversarial approach where the discriminator and PINN, the existing approach to solve PDEs using neural networks, plays a game in order to improve the solution precision of PDEs. The authors motivate their work by the fact that the existing PINNs approaches fail to generate precise solution. The authors show that the proposed model, CPINN, is well-behaved compared to the original PINN model, from the optimization/training perspective. The authors support their claim using multiple PDE problems.

Strength And Weaknesses:

Strengths: 1- Novelty 2- Well-written 3- Experimental results

Weakness: Cannot think of any

Clarity, Quality, Novelty And Reproducibility:

The paper is very well-written and clear. The idea is novel, to the best of my knowledge. I cannot comment on the reproducibility as I am not sure if enough details are provided for that end.

Summary Of The Review:

I enjoyed reading this paper. Physics informed machine learning is a very important and hot problem and still in its infancy. The authors efforts in improving the performance of existing models for this problem needs to be highly acknowledged and rewarded. The authors were able to motivate their approach and support that using experimental study. Overall, I believe this paper has a high quality and ready for publication. I should only add that I am not fully familiar with the state of the arts in this topic and assumed that the authors reflected it well in the paper.

One minor issue: I am curious to see similar plots to that of Figure 1 for other datasets.

Correctness: 4: All of the claims and statements are well-supported and correct.

Technical Novelty And Significance: 4: The contributions are significant, and do not exist in prior works.

Empirical Novelty And Significance: 4: The contributions are significant, and do not exist in prior works.

Flag For Ethics Review: NO.


Recommendation: 8: accept, good paper

Confidence: 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.

[–] Response to reviewer 1E2V

ICLR 2023 Conference Paper5528 Authors Florian Tobias Schaefer (/profile?id=~Florian_Tobias_Schaefer1) (privately revealed to you)

15 Nov 2022 (modified: 16 Nov 2022) ICLR 2023 Conference Paper5528 Official

Comment Readers:  Everyone Show Revisions (/revisions?

id=YQpiSeKuch9)

Comment:

We thank you for your time and the positive review. We have added heatmaps for the remaining experiments to the appendix of the revised version.

[–] Official Review of Paper5528 by Reviewer hCa9

ICLR 2023 Conference Paper5528 Reviewer hCa9

21 Oct 2022 ICLR 2023 Conference Paper5528 Official Review Readers: 

Everyone Show Revisions (/revisions?id=-sNAPTWPYh)

Summary Of The Paper:

This paper produces more accurate physics-informed neural networks (PINNs) by including adversarial training through a discriminator which is rewarded for predicting PINN errors.

Strength And Weaknesses:

Strengths: • Claims are supported by reported results. • Concept is sound and sensible. • Making PINNs more accurate has good impact for the field. • Paper is well-written and easy to read and navigate. • Mathematics is well explained. • Citations are thorough. • Tradeoffs between this method and conventional PINNs are discussed.

Weaknesses: • Figures describing the method are not included, and they would make the method easier to understand in this case.

Clarity, Quality, Novelty And Reproducibility:

The work is well-written and easy to follow. Bolded subheadings are used to create small sections, allowing for quick navigation of the paper.

The experimental quality is quite high, testing on a variety of problems with good results.

The novelty of the work is middling but non-trivial, with the addition of a discriminator network to reduce error in the PINN.

Reproducibility is possible with hyperparameters and implementational details available in the appendix.

Summary Of The Review:

This paper seems to be a solid step forward in the field of physics-based neural networks – a field that has high impact on a number of domains. The paper is generally well-written. The novelty of the work is middling (maybe too straightforward because of the recent popularity of GANs) but non-trivial.

Correctness: 4: All of the claims and statements are well-supported and correct.

Technical Novelty And Significance: 2: The contributions are only marginally significant or novel.

Empirical Novelty And Significance: 3: The contributions are significant and somewhat new. Aspects of the contributions exist in prior work.

Flag For Ethics Review: NO.


Recommendation: 6: marginally above the acceptance threshold

Confidence: 3: You are fairly confident in your assessment. It is possible that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work. Math/other details were not carefully checked.

[–] **Response to reviewer hCa9**

ICLR 2023 Conference Paper5528 Authors Florian Tobias Schaefer (/profile?id=~Florian_Tobias_Schaefer1) (privately revealed to you)

15 Nov 2022 (modified: 16 Nov 2022) ICLR 2023 Conference Paper5528 Official

Comment Readers:  Everyone Show Revisions (/revisions?id=8uu2yWEz9G)

Comment:

We thank you for taking the time to review our work and for your positive feedback. We have included a figure describing our method in the appendix, which we will move to the main part of the camera-ready version if space permits. We did not include it in the main part for now in order to keep the figure numbering constant during the discussion.

[–] **Official Review of Paper5528 by Reviewer QL5E**

ICLR 2023 Conference Paper5528 Reviewer QL5E

14 Oct 2022 (modified: 16 Nov 2022) ICLR 2023 Conference Paper5528 Official

Review Readers:  Everyone Show Revisions (/revisions?id=QRbjzpZvHr)

Summary Of The Paper:

This paper introduce competitive physics-informed networks where two neural networks solve a partial differential equation by playing a zero-sum game. The formulation is built using a weak form. The authors claim that this weak form have smaller condition number.

Strength And Weaknesses:

Strength: The experiments look strong, and the writing is super clear to understand.

Weakness: Several claims are hand waving, and the reviewer is not convinced. I think if the paper is accepted at this stage, it will make confusion in the community.

Regards Experiment The reviewer's first major concern is the relationship of this paper with [1] which also uses the weak formulation to train a PDE solution. I think an experiment comparing these works is essential (using the same optimization algorithm).

The paper only includes results comparison with respect to the number of iterations. Comparison with respect to the computation time is also needed. For CPINN introduces a new NN and CGD have a matrix inverse. Further computation cost have been included.

Regards Mni-max min-max optimization is hard. The complexity lower bound proposed of minimax optimization has a term $K_x K_y$ [2]. I guess using min-max has the same computational complexity as the original formulation.

To reduce the condition number of laplace squares, we can still use PINN to do this by considering $\min_{\{u_1, u\}} \|u_1 - \nabla u\|^2 + \|f - \nabla \cdot u_1\|^2$ to solve $\Delta u = f$. I don't think the minimax formulation is essential.

At the same time, the paper uses ADAM, which is an adaptive method. The adaptive methods can cancel the condition number. I'm confused by this mismatch of the experiment with the theory claimed.

Regards the Squared residuals. Changing the loss into the Sobolev norm also introduces a further matrix A. But [3,4] showed this would only accelerate the training both in experiments and theory for the machine learning algorithms. This paper have considered the theory as [4] includes a machine learning model into the learning objective. Can the authors comment on this?

Related work [5] The reviewer thinks the paper is still valuable. [5] proposed a similar trick from the optimization community and figured out the improvement of this method. I encourage the authors to think about combining the theory in [4] and [5] to figure out the true reason this method has an acceleration.

[1] Yaohua Zang, Gang Bao, Xiaojing Ye, and Haomin Zhou. Weak adversarial networks for high-dimensional partial differential equations. Journal of Computational Physics, 411:109409, 2020

[2] Zhang J, Hong M, Zhang S. On lower iteration complexity bounds for the convex-concave saddle point problems. Mathematical Programming, 2022, 194(1): 901-935.

[3] Son H, Jang J W, Han W J, et al. Sobolev training for the neural network solutions of pdes[J]. arXiv preprint arXiv:2101.08932, 2021.

[4] Lu Y, Blanchet J, Ying L. Sobolev Acceleration and Statistical Optimality for Learning Elliptic Equations via Gradient Descent[J]. arXiv preprint arXiv:2205.07331, 2022.

[5] <https://proceedings.neurips.cc/paper/2018/file/08048a9c5630ccb67789a198f35d30ec-Paper.pdf>
(<https://proceedings.neurips.cc/paper/2018/file/08048a9c5630ccb67789a198f35d30ec-Paper.pdf>)

===

The author's response convinced me this is a worth pursuing direction, I've raised my score.

Clarity, Quality, Novelty And Reproducibility:

The writing of this paper is super great. I enjoyed reading this paper.

Summary Of The Review:

The experiments look strong, and the writing is super clear to understand. Several claims are hand waving, and the reviewer is not convinced. I think if the paper is accepted at this stage, it will make confusion in the community.

I still consider this method have made a huge contribution to the PINN family. I suggest authors to investigate paper [5] in detail to see the benefit of this method.

Correctness: 3: Some of the paper's claims have minor issues. A few statements are not well-supported, or require small changes to be made correct.

Technical Novelty And Significance: 3: The contributions are significant and somewhat new. Aspects of the contributions exist in prior work.

Empirical Novelty And Significance: 4: The contributions are significant, and do not exist in prior works.

Flag For Ethics Review: NO.


Recommendation: 8: accept, good paper

Confidence: 5: You are absolutely certain about your assessment. You are very familiar with the related work and checked the math/other details carefully.

[–] Response to reviewer QL5E, Part 1/2

ICLR 2023 Conference Paper5528 Authors Florian Tobias Schaefer (/profile?id=~Florian_Tobias_Schaefer1) (privately revealed to you)

15 Nov 2022 (modified: 16 Nov 2022) ICLR 2023 Conference Paper5528 Official

Comment Readers:  Everyone Show Revisions (/revisions?)

id=Uy2ixdck_pD)

Comment:

Thank you for your careful review of our work. You made a number of insightful comments that helped us improve the paper. We respond to them below, in the order in which they were raised.

Regarding experiments:

Comparison to WAN:

We have results for weak adversarial networks (WAN), trained with Adam or ACGD, to Figure 6. We observe that ACGD improves the stability of WAN training compared to the combination of Adam and Adagrad used by Zang et al. However, we were unable to achieve accuracy compared to CPINN with ACGD despite trying different sets of activation functions, loss functions (dropping the logarithm as suggested in section 3.5 of Zang et al.), and training modes (minibatching vs fixed batch). In summary, we believe that Figure 6 now makes a compelling case that the improvements of CPINN are not just a function of the optimizer used for its training.

Computational Cost:

We believe that we have already accounted for the additional costs associated with the CPINN methodology. This was done by tracking the number of "forward passes," which accounts for the gradient or Jacobian-vector product computation with respect to either network, including those used by the iterative solution of the matrix-inverse in the update of rule of (A)CGD. These are the only additional components that make CPINN potentially more (or less) costly than a standard PINN. We use this measure of computational cost since it is independent of hardware and implementation.

Regarding min-max:

Complexity lower bounds:

We believe that you are correct that the asymptotic complexity for a generic matrix \mathbf{A} , with difficulty measured by the condition number of \mathbf{A} , can not be improved by simply switching to min-max formulation. In particular, directly applying a Krylov subspace method to the symmetric dilation $\begin{pmatrix} 0 & \mathbf{A}^\top \\ \mathbf{A} & 0 \end{pmatrix}$ is not going to be any more efficient than applying a Krylov method to $\mathbf{A}^\top \mathbf{A}$. However, there exist many matrices, including those that are symmetric positive-definite or those that are symmetric with a different condition number on the positive and negative real line, where matrix-vector products with \mathbf{A} allow for a significantly (both in practice and asymptotically as a function of the condition number) faster solution than what would be possible by matrix-vector products with $\mathbf{A}^\top \mathbf{A}$, see Chapter 3 in (Greenbaum, 1997) (<https://epubs.siam.org/doi/book/10.1137/1.9781611970937>). We have modified Section 2.3 to make these caveats more clear.

Alternative way of reducing the condition number:

The approach that you suggest could indeed help to reduce the condition number arising from the partial differential operator, since the resulting gradient can implement multiplication with the PDE operator instead of its square. Similar approaches are commonly used in the solution of fourth-order equations, transforming them into a system of second-order equations. The caveat of this approach is that it requires the a priori knowledge of a factorization of the ill-conditioned operator (which is available for differential operators). It is, therefore, unable to address the ill-conditioning that arises in the network Jacobian. Beyond this caveat, the practical improvements of

this approach would need to be verified empirically. We do not claim that CPINN is the only possible way of improving the performance of PINNs though we believe that it cuts a key reason that PINNs do not train well and provides a robust and general method that achieves greatly improved accuracy on many problems.

Adaptive Methods and Ill-conditioning

We agree that methods like Adam attempt to improve the conditioning by applying a diagonal rescaling. However, such a rescaling can only remove ill-conditioning that arises from singular vectors aligned with the coordinate axes. Many ill-conditioned systems have large and small singular values for which the respective singular vectors are not aligned with the coordinate axes. For instance, finite-difference discretizations of the Laplace operator are ill-conditioned, and yet their condition number can not be meaningfully reduced by a diagonal rescaling. Absent special structure, it seems unlikely that a generic ill-conditioned matrix can be well-conditioned by diagonal scaling. The fact that both PINN and CPINN require a large number of iterations also indicates that adaptive methods are not sufficient to overcome the ill-conditioning in practice.

[+] **Official Comment by Paper5528 Authors • Response to reviewer QL5E, Part 2/2**

[–] **Response to reviewer QL5E, Part 1/2**

[Deleted] Florian Tobias Schaefer (/profile?id=~Florian_Tobias_Schaefer1) (privately revealed to you)

ICLR 2023 Conference Paper5528 Official Comment Readers: Program Chairs,
Paper5528 Senior Area Chairs, Paper5528 Area Chairs, Paper5528 Reviewers,
Paper5528 Reviewers Submitted, Paper5528 Authors Show Revisions (/revisions?
id=IV_iGtEMWk)

[About OpenReview \(/about\)](/about)

[Hosting a Venue \(/group?
id=OpenReview.net/Support\)](/group?id=OpenReview.net/Support)

[All Venues \(/venues\)](/venues)

[Sponsors \(/sponsors\)](/sponsors)

[Frequently Asked Questions](#)

[\(https://docs.openreview.net/getting-
started/frequently-asked-questions\)](https://docs.openreview.net/getting-started/frequently-asked-questions)

[Contact \(/contact\)](/contact)

[Feedback](#)

[Terms of Use \(/legal/terms\)](/legal/terms)

[Privacy Policy \(/legal/privacy\)](/legal/privacy)

[OpenReview \(/about\)](/about) is a long-term project to advance science through improved peer review, with legal nonprofit status through [Code for Science & Society \(https://codeforscience.org/\)](https://codeforscience.org/). We gratefully acknowledge the support of the [OpenReview Sponsors \(/sponsors\)](#).