Optimization of Plasma-Surface Kinetics via Machine Learning Methods

José Afonso[1], Vasco Guerra[1], Rodrigo Ventura[2], Pedro Viegas[1]

[1] Instituto de Plasmas e Fusão Nuclear, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisboa, Portugal

[2] Instituto de Sistemas e Robótica, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisboa, Portugal

josefafonso@tecnico.ulisboa.pt

Fundação para a Ciência e a Tecnologia

Introduction

Optimizing plasma-surface kinetics is a critical challenge in material science, often hindered by computationally expensive simulations. Moreover, a key challenge is in the uncertainty of model parameters, such as energy barriers, for some chemical reactions, which are not well defined in the literature [1]. Therefore, this work employs a data-driven optimization approach to improve the physical models.

To overcome these bottlenecks, we introduce a dual-strategy approach tailored to different simulation regimes:

- For *fast* simulators (it takes less than ~10 secs to run): a hierarchical optimization algorithm that efficiently navigates complex and sloppy parameter landscapes [2];
- For *slow* simulators (it takes more than ~500 secs to run): a ML-accelerated framework using a Progressive Conditional Denoising Autoencoder (PCDAE) as an accurate surrogate model [3];

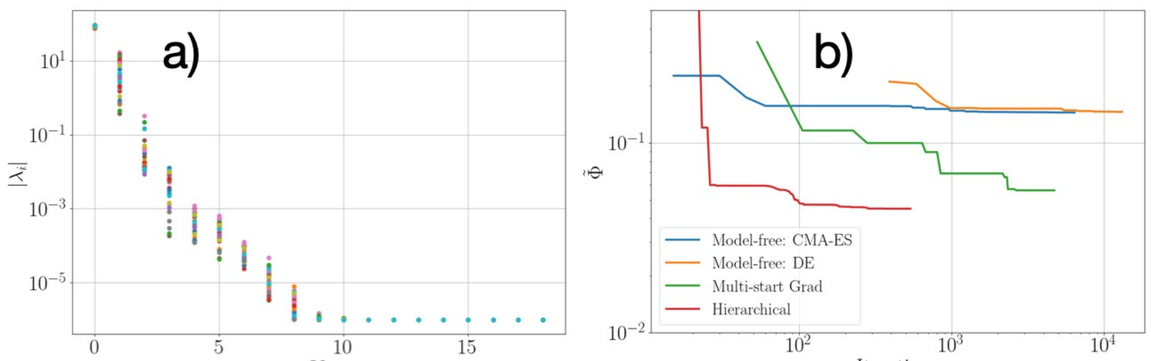
This work demonstrates that both strategies provide significant efficiency gains, accelerating the discovery of optimal plasma processing conditions.

Model-free Optimization for *Fast* Simulators

Optimizing parameters of interest in the O and CO kinetic scheme [1] is challenging due to the model's *sloppiness*.

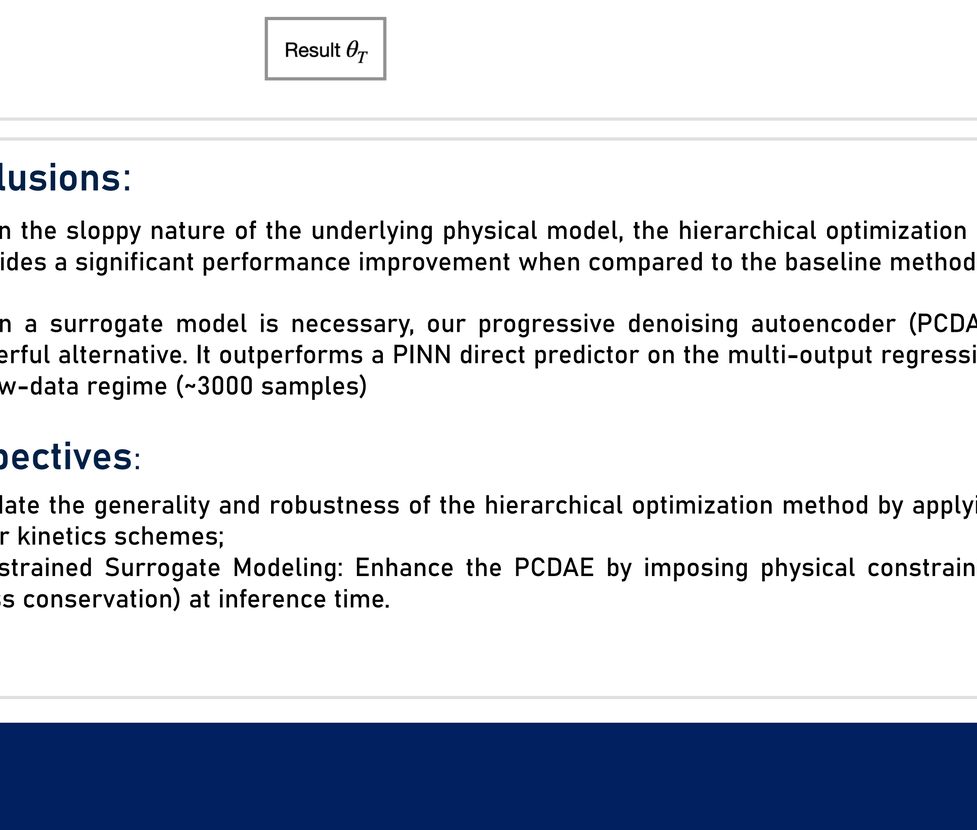
We define the loss function as $\widetilde{\Phi}(\theta) = \sum_i r_i^2$, where i iterates over the dataset and the residuals are given by $r_i = \frac{\gamma_i - \widehat{\gamma_i}}{\gamma_i}$. The γ_i represents the experimental measurement and $\widehat{\gamma_i}$ the predicted value. The vector θ concatenates all the parameters to be optimized.

As shown in Fig. a), the parameter sensitivities span many orders of magnitude. This mixture of *stiff* (sensitive) and *sloppy* (insensitive) directions makes optimization with standard algorithms challenging [2].



- Fig. a) shows the eigenvalue spectrum ($|\lambda_i|$) of the model's Gauss-Newton Hessian, $H_{GN} \approx \nabla^2\tilde{\Phi}(\theta)$, as a function of the eigendirections V_i. The H_{GN} is expressed as $H_{GN} = V\Lambda V^T$, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.
- The eigenvalues span many orders of magnitude, confirming the model's sloppy nature. The large eigenvalues correspond to *stiff* directions, while the tiny eigenvalues correspond to sloppy directions that are challenging for standard optimization methods to resolve.
- Fig. b) compares the convergence of the loss function ($\widetilde{\Phi}$) versus iterations for our Hierarchical method (red) against three baselines, showing it converges faster and for a better solution when compared to the baselines.

To solve this, our hierarchical algorithm (see scheme) iteratively partitions parameters into *stiff* and *sloppy* subspaces. It then optimizes each subspace in sequence, efficiently navigating the complex landscape to find the final parameters θ_T.



Surrogate-Model for *Slow* Simulators

Our Progressive Conditional Denoising Autoencoder (PCDAE) replaces a slow simulator, overcoming regression model limits by implicitly learning data geometry and utilizing robust, conditional contrastive learning to reconstruct clean samples from noisy versions [3,4]

Training Stage: Learning to Denoise

The model g_ϕ is trained to reconstruct the true chemical densities y from a noisy version $\tilde{y}$, conditioned on experimental inputs x.
Noise is added via $\tilde{y} \sim q_\sigma(\cdot|\, y)$, where the noise level σ is randomly sampled from a distribution p_σ during training. The objective is to minimize the loss:

$$\mathcal{L}(\phi) = \mathbb{E}_{(x,y)\sim p_{\mathrm{data}}} \mathbb{E}_{\sigma\sim p_\sigma} \mathbb{E}_{\tilde{y}\sim q(\cdot|y)} \left[\| g_\phi(\tilde{y}, x; \sigma) - z \|_2^2 \right]$$

with $z^T = [y, x]$.
By being trained to recover clean samples from noisy inputs, the network g_ϕ is compelled to learn the geometry and correlations of the data manifold implicitly [3].

Inference Stage: Iterative Refinement

To generate a prediction, we start with pure noise and denoise it iteratively. Following a schedule of decreasing noise levels, $\{\sigma_k\}$, from $\sigma_{max} \gg 0$ to $\sigma_{min} \approx 0$, our model g_ϕ guides the prediction out of the noise. The refinement update is given by:

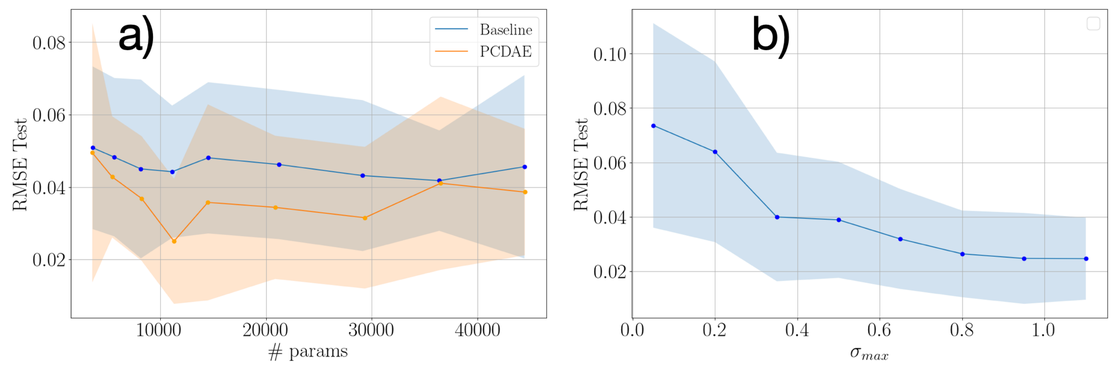
$$y \leftarrow y + \eta\, v(y, x; \sigma)$$

With $v(y, x; \sigma) = g_\phi(y, x; \sigma) - y$

Dataset

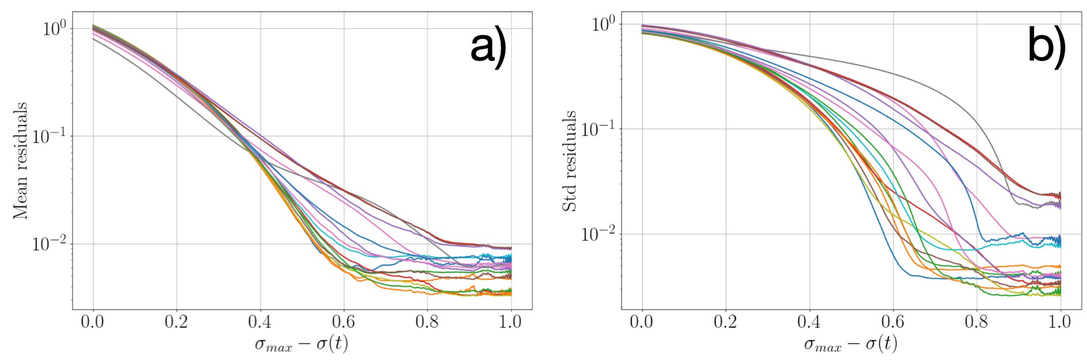
The dataset was generated using LoKI-B+C [5], obtained from [6], and corresponds to a deterministic system that predicts densities of chemical species, y (17 vars), based on the input experimental conditions for the plasma system, x (3 $vars$). It is composed by 3000 samples (x, y).

Results



- Fig. a) Superior Accuracy: our PCDAE model outperforms the PINN with Projection [6] baseline, achieving a lower test RMSE across all model sizes. The $\#params$ corresponds to the number of learnable weights in each neural net.
- Fig. b) Importance of Noise: high performance requires training with a large maximum noise level ($\sigma_{max} \geq 0.9$), reinforcing the contrastive nature of our method.

The following figures show the evolution of absolute value residuals, with each line representing a different y component, starting from a set of random initial guesses y.



- Fig. a) Stable Mean: the mean of each y component converges to a stable solution near the ground truth;
- Fig. b) Variance Collapse: the standard deviations of each y component collapse to (almost) zero.

This proves the model reliably finds a unique solution regardless of the random starting point, an attractive features as we are dealing with a deterministic dataset.

Conclusions:

- Given the sloppy nature of the underlying physical model, the hierarchical optimization method provides a significant performance improvement when compared to the baseline methods.
- When a surrogate model is necessary, our progressive denoising autoencoder (PCDAE) is a powerful alternative. It outperforms a PINN direct predictor on the multi-output regression task in low-data regime (~3000 samples)

Perspectives:

1. Validate the generality and robustness of the hierarchical optimization method by applying it to other kinetics schemes;
2. Constrained Surrogate Modeling: Enhance the PCDAE by imposing physical constraints (e.g., mass conservation) at inference time.

References:

[1] Pedro Viegas et al 2024 Plasma Sources Sci. Technol. 33 055003 https://doi.org/10.1088/1361-6595/ad4055,
[2] M. K. Transtrum, B. B. Machta, and J. P. Sethna. (2010). Why are nonlinear fits so challenging? Physical Review Letters, 104(060201), 060201,
[3] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol. (2008). "Extracting and composing robust features with denoising autoencoders." In Proceedings of the 25th International Conference on Machine Learning ,
[4] Song, Y., & Ermon, S. (2019). Generative Modeling by Estimating Gradients of the Data Distribution. In Advances in Neural Information Processing Systems 32 (NeurIPS 2019),
[5] https://nprime.tecnico.ulisboa.pt/loki
[6] Valente, M., Dias, T. C., Guerra, V., & Ventura, R. (2025). Physics-consistent machine learning: output projection onto physical manifolds. arXiv preprint arXiv:2502.15755.

Acknowledgements:

This work was supported by the European Union under Horizon Europe project CANMILK (DOI:https://doi.org/10.3030/101069491.