

Machine learning kernels for real-time complex Langevin

Enno Carstensen^{a,*} and Dénes Sexty^a

^a*Institute of Physics, NAWI Graz, University of Graz,
Universitätsplatz 5, Graz, Austria*

E-mail: enno.carstensen@uni-graz.at, denes.sexty@uni-graz.at

Real time evolution in QFT poses a severe sign problem, which may be alleviated via a complex Langevin approach. However, simulation results fail to converge with a large real-time extent. A kernel in a complex Langevin equation is known to influence the appearance of the boundary terms and integration cycles, and thus kernel choice can improve the range of real-time extents with correct results. For multi-dimensional models the optimal kernel is searched for using machine learning methods. We test this approach by simulating the simplest possible case, a 0+1-dimensional scalar field theory in Minkowski space. The performance of band-diagonal kernels as well as the existence of integration cycles in the theory is also discussed.

*The 42nd International Symposium on Lattice Field Theory (LATTICE2025)
3-8 November 2025
Mumbai, India*

*Speaker

1. Introduction

The infamous sign problem denies us the ability to apply the standard lattice HMC method to complex actions, which appear e.g. in high-density QCD and also real-time evolution [1]. Common techniques to deal with the sign problem, like extrapolation and reweighting are limited to theories with weak sign problems, i.e. QCD with low μ/T (see e.g. the reviews [2, 3]). New methods based on complex analytic functions are applicable to severe sign problems [4], these include Lefschetz thimbles [5] and complex Langevin [6, 7], both using a complexified field space. Complex Langevin has been successfully applied to a variety of problems, including toy models [8], QCD at finite density [9] and real-time field theories [10]. However naive complex Langevin (CL) can produce wrong results under certain circumstances, such as QCD at low temperature and finite chemical potential or real-time evolution at large real-time extent. There are several ways a CL simulation can fail, but by introducing an appropriate kernel convergence can be restored. The main problem for finding a suitable kernel is the huge parameter space, as every holomorphic function is in principle usable. By utilizing machine learning in the form of optimization by simple gradient descent, such a kernel can be learned. This has been explored in the context of real-time scalar theories in [11–13] using dense constant kernels. In this work we try to build on this approach by trying to improve space and performance scaling of the learning process, investigate if non-trivial cycles play a role and look into extending the reachable real-time extend by widening the kernel space to first include linear field-dependent kernels.

2. Real-time ϕ^4 -theory on the lattice

We study an anharmonic quantum oscillator described by the action quartic action

$$S = \int_C dt \partial^\mu \phi \partial_\mu \phi + m^2 \phi^2 + \frac{\lambda}{4!} \phi^4 \quad (1)$$

on a $0 + 1$ -dimensional real-time contour C . In a continuum theory, one can use the Keldysh formalism to extract real-time equilibrium observables [14]. A naively discretized Schwinger-Keldysh contour, which connects the origin, the apex at t_{\max} and the return at $-i\beta$ through horizontal and vertical pieces, still needs to be regulated by a slight tilt $k > 0$ in the real legs [10]. If we are also not interested in reverse time-ordered correlators, we can ignore the back leg and connect apex and return directly, saving some compute. The final contour starts at the origin, extends up to the apex at $(t_{\max}, -k\beta)$, and returns in a straight line to the return at $(0, -\beta)$, with periodic boundary conditions.

The usual observables are the unequal-time correlator $\langle \phi(0)\phi(t) \rangle$ and the equal-time correlator $\langle \phi(t)\phi(t) \rangle$, which is constant over the whole lattice. This system is solvable by diagonalizing its Hamiltonian, which is useful since we can compare our method to the exact results. The standard benchmark for this setup is $m = 1$, $\lambda = 24$, $\beta = 1$ and we try to achieve an as high as possible t_{\max} whilst reproducing the correct result.

3. CLE Basics

The Complex Langevin equation (CLE) is an alternative sampling method for evaluating path integrals based on stochastic quantization [15]. Unlike Monte Carlo methods, it does not rely on

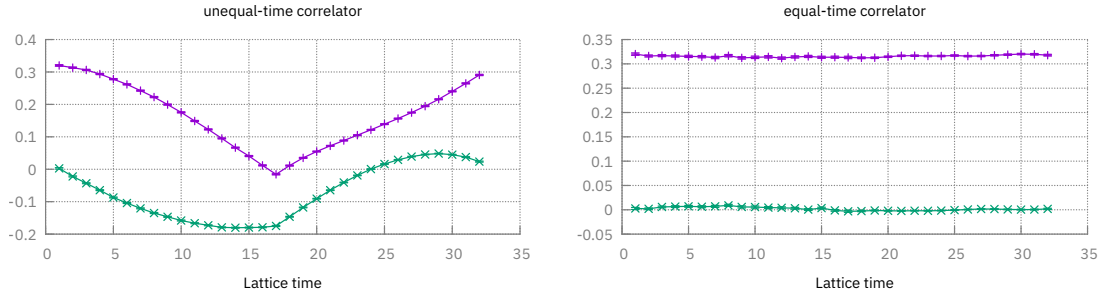


Figure 1: Correct results with trivial kernel for $t_{\max} = 0.8$, $\beta = 1$, $k = 0.01$, $m = 1$, $\lambda = 24$

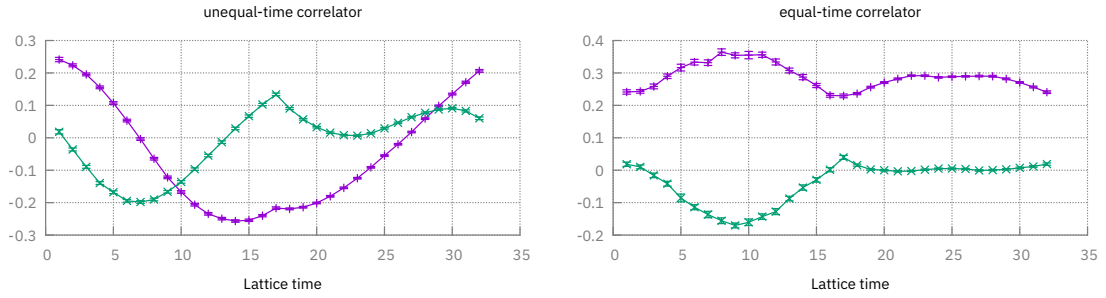


Figure 2: Wrong results with trivial kernel for $t_{\max} = 1.6$, $\beta = 1$, $k = 0.01$, $m = 1$, $\lambda = 24$

the fact that the action is real and positive and can therefore be applied to complex actions. After stochastic quantization, where one introduces an additional time dimension (in this context usually called Langevin time), the crucial step to sampling complex actions is the complexification of the degrees of freedom of the theory (in the case where one uses e.g. complex scalars, real and imaginary part of complex variables get complexified themselves), which allows us to turn the usual complex distribution of real fields into a real distribution of complexified fields, which we can sample from. This sampling is done using the Langevin equation, also called the complex Langevin equation when applied to complex actions.

$$d\phi = -\frac{\partial S}{\partial \phi} d\tau + \sqrt{2}dw \quad (2)$$

It is a first-order stochastic differential equation with a standard Wiener process w , which evolves the degrees of freedom of the system in Langevin time. The expectation values of observables can be obtained by taking the Langevin time average in the limit $\tau \rightarrow \infty$. Taking this limit is obviously impractical in practice, which is why in practice numerical discrete solvers are employed, which produce a distribution of field values, akin to a MCMC algorithm. The simplest choice for the discretization of the Langevin equation is the explicit, unimproved Euler-Maruyama

$$\Delta\phi = -\epsilon \frac{\partial S}{\partial \phi} + \sqrt{2\epsilon}\eta, \quad (3)$$

with step size ϵ and Gaussian random variable η with zero mean and unit variance. Gathering and analysis of observables follows the established methods of HMC lattice simulations.

4. Kernelled CLE

A common issue of this method is wrong convergence where some (typically all) observables of the theory deviate from their correct value, see 2. This is most often due to the appearance of boundary terms [16, 17]. A possible remedy is the Kernel. The kernel K can be any arbitrary holomorphic function and in the case of real actions, does not change the result.

$$\frac{\partial \phi}{\partial \tau} = -K \frac{\partial S}{\partial \phi} + \frac{\partial K}{\partial \phi} + \sqrt{2K} \eta \quad (4)$$

On the computer it is more practical to work with H , since taking the square root of a matrix is quite expensive.

$$\Delta \phi = -H^T H \epsilon D(\phi) + \sqrt{2\epsilon} H \eta \quad (5)$$

In a complex setting however, the kernel can influence the results and restore correct convergence. (It is also possible to induce wrong convergence.) Since the kernel can be any holomorphic function, picking the right kernel for the right theory is quite a hard problem. Hand-tuned kernels have restored correct convergence in 1-dof toy models [18], real-time gauge theories [19] and real-time scalar theories.

5. Machine learning Kernels

Another method of achieving a convergence-restoring kernel is machine learning [11–13], optimizing the kernel based on heuristics which indicate correct convergence. Heuristics can include boundary terms, symmetries of the system, euclidean observables or the unitary norm [11]. Different heuristics can of course be combined to arrive at a comprehensive loss function, but the simplicity of the unitarity norm

$$U(\phi) = \sum \text{Im}(\phi_i)^2 \quad (6)$$

(where the some is meant for all the degrees of freedom for the action) specifically makes it an appealing choice as the single heuristic for a loss function. The general idea is to keep the sampling close to the real axis, where one would also integrate along when solving the continuum path integral. To make it a loss function it has to be the unitarity norm of the next step, as this field configuration actually depends on the kernel, such that the kernel update

$$\Delta H_{ij} = -r \frac{\partial L(\Phi)}{\partial K_{ij}} \quad (7)$$

can be computed with r , the learning rate.

The general method is as follows: First the kernel gets optimized via the aforementioned gradient descent via the loss function gradient. This quantity is sampled multiple times over a range of Langevin steps, averaged and applied to the kernel. This process is a single epoch. The whole optimization process consists of multiple epochs, ranging from 100 to 10^7 , depending on the system and the other hyperparameters such as learning rate $r \in (10^{-6}, 1)$ and sampling interval $\tau_l \in (0.1, 100)$. This optimized kernel is then used in the usual manner to gather observables using the CLE, starting from a new hot or cold start.

This setup was already used in [12], yielding correct results for $t_{\max} = 1.6$, with wrong convergence appearing above $t_{\max} = 2$, using the standard benchmark of $\beta = 1$, $m = 1$, $\lambda = 24$.

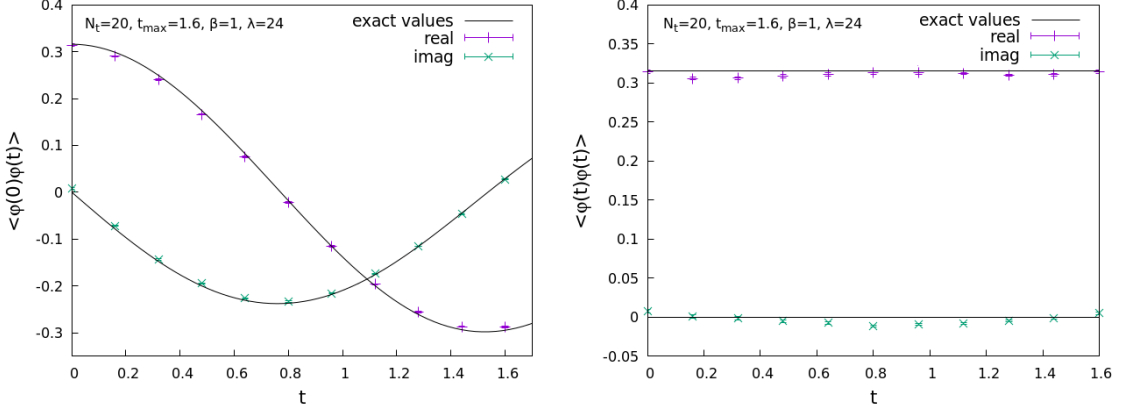


Figure 3: Observables at $t_{\max} = 1.6$ with optimized Kernel [12]

6. Band-diagonal Kernels

Besides the still limited range that optimized constant kernels provide, there are two other issues that need to be addressed to achieve practicability.

The first issue has to do with the size of the kernel. In the setup as used so far, the kernel size scales with the square of the system size (lattice sites). In a 0+1-dimensional or 1+1 dimensional simulations done so far, this is manageable since $\|H\| = N_t^2$, but in a 4-dimensional setup, which is needed for realistic models, $\|H\| = N_s^6 N_t^2$, which is unfortunate. Luckily for us, it seems like most of the entries in the kernel do not seem to matter and the only relevant entries are on or close to the main diagonal. Since we are dealing with a local theory, this makes sense. We can therefore try to reduce the effective kernel size by restricting the kernel to only the seemingly relevant elements. This seems to be the main and first sub-diagonals in our setup with maybe the second sub-diagonals also playing a minor role. Restricting the kernel to only a few diagonals reduces the effective size to $\|H\| \sim N_t$ in the one-dimensional case and $\|H\| \sim N_s^3 N_t$ in the 4-dimensional one.

We tested this setup, restricting the kernel to a few different numbers of diagonals, starting from 3, forcing all other entries to 0, and going through the usual optimization process.

This did clearly not work out as we intended. Maybe the hyperparameters need to be drastically adjusted or the far-off-diagonal elements do play a significant role and are somehow correlated. If we are lucky, the former case is true and we can with relative ease reduce kernel size and learning cost. If the latter is the case, then the kernel size will be quadratic in the system size and we need to learn every single entry. However even determining if this is the case will be already quite hard. Either way, starting with a full kernel and then taking out diagonals one-by-one is a more prudent approach, which we implement currently.

7. Integration cycles in field theories

The Salcedo-Seiler Theorem [20] states that in complex Langevin for one degree of freedom in the absence of boundary terms the observables are actually given by a combination of observables in different integration cycles. Good evidence for this holding for multiple degrees of freedom exists [18], but a non-trivial integration cycle has not yet been observed in a field theory. Of the

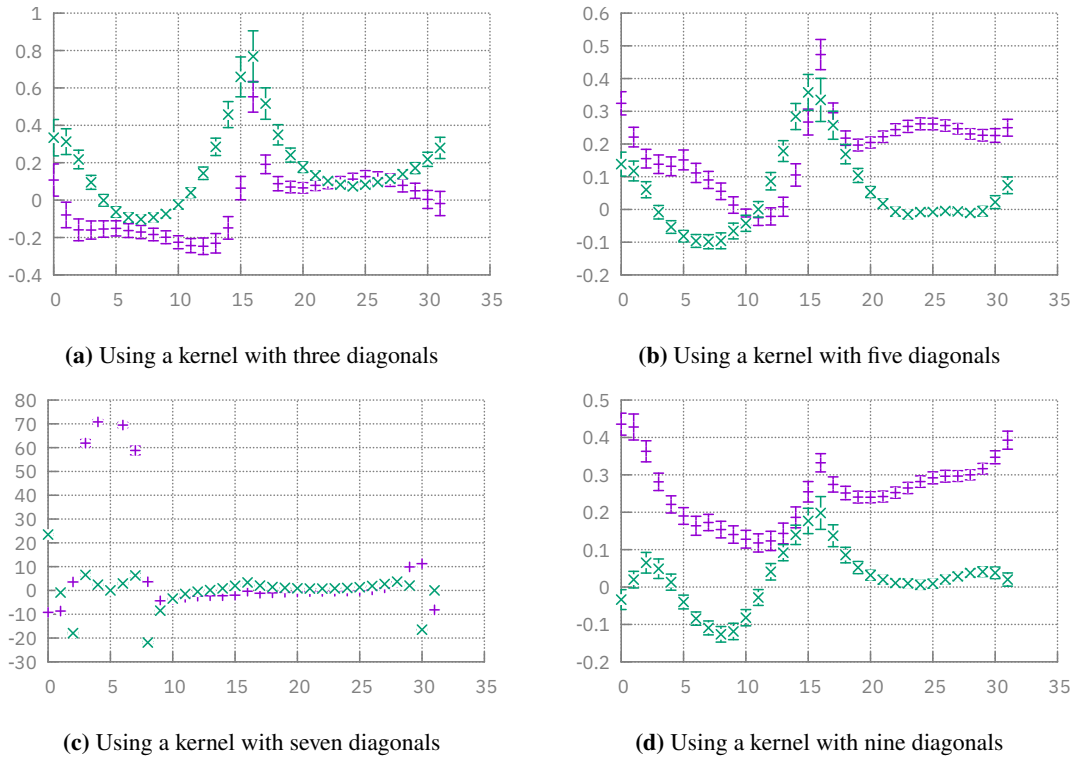


Figure 4: Equal-time correlators using optimized band-diagonal kernels with three, five, seven and nine diagonals respectively. None of them are able to restore correct convergence.

several failure modes of CLE, non-trivial integration cycles is perhaps the hardest to diagnose. Recently, a criterion that can reliably determine whether the results are from a certain cycle has been invented [21], but the criterion needs further improvement to allow application to field theories.

Since our model is arguably the simplest non-trivial field theory, we want to investigate if cycles can show up and how we can distinguish if we are in a non-trivial cycle (meaning any cycle but the real one). For one degree of freedom with a quartic action there should be 3 different integration cycles, giving a basis for connecting the zeroes of the action at $\infty, i\infty, -\infty, -i\infty$. In a field theory, we get the trivial cycle where we integrate each field from $-\infty$ to ∞ and the "fully imaginary cycle" where the integration goes from $-i\infty$ to $i\infty$ for each field. More combinations exist, as at each lattice site we can choose a combination of 3 independent integration contours, leading to a total of 3^N cycles (with N giving the number of lattice points), which are probably not all linearly independent [18]. Determining if the fully imaginary cycle and the trivial cycle are independent is an important first step in understanding integration cycles in field theories.

We investigated the above mentioned two cycles by integrating numerically using trapezoidal integration with a small amount of lattice sites. The cost of this numerical integral increases exponentially with N , thus it is manageable for small N only. The smallest lattice that makes sense in the setup detailed above is $N = 4$. Preliminary studies indicate that the trivial and fully imaginary cycles indeed produce different observables.

Since it seems like that there are indeed different non-degenerate integration cycles in this

theory, we can try to reproduce their results using kernelled CLE. We can alter the starting value of the kernel to see if we end up with different observables and kernels than usual. We can also change the loss function to e.g. minimize the real part of the fields, this should correspond to sampling the fully imaginary cycle. Both approaches are currently being pursued.

8. Conclusions

Using complex Langevin with machine-learned kernels still seems to be the best method for real-time scalar theories. However to reach wide applicability in physical systems, a lot of work needs to be done. More in-depth investigation is needed with regards to band-diagonal kernels to potentially improve space and performance scaling. The possibility of integration cycles also needs to be addressed such that results can be 100% trustable in real-world scenarios. The way forward to increase the accessible real-time extent is to introduce learnable field-dependencies into the kernel.

Acknowledgments

We thank Michael Mandl for valuable discussions. E.C. acknowledges financial support from the FWF under Project P 36875. The numerical simulations were performed on the Graz Scientific Cluster and VSC5 of the ASC.

References

- [1] M. Troyer and U.-J. Wiese, *Computational complexity and fundamental limitations to fermionic quantum Monte Carlo simulations*, *Physical Review Letters* **94** (2005) 170201 [[cond-mat/0408370](#)].
- [2] P. de Forcrand, *Simulating QCD at finite density*, *PoS LAT2009* (2009) 010 [[1005.0539](#)].
- [3] J.N. Guenther, *An overview of the QCD phase diagram at finite T and μ* , *PoS LATTICE2021* (2022) 013 [[2201.02072](#)].
- [4] D. Sexty, *New algorithms for finite density QCD*, *PoS LATTICE2014* (2014) 016 [[1410.8813](#)].
- [5] A. Collaboration, M. Cristoforetti, F.D. Renzo and L. Scorzato, *High density QCD on a Lefschetz thimble?*, *Physical Review D* **86** (2012) 074506 [[1205.3996](#)].
- [6] G. Parisi, *ON COMPLEX PROBABILITIES*, *Phys. Lett. B* **131** (1983) 393.
- [7] J.R. Klauder, *A Langevin approach to fermion and quantum spin correlation functions*, *Journal of Physics A: Mathematical and General* **16** (1983) L317.
- [8] H. Okamoto, K. Okano, L. Schulke and S. Tanaka, *The Role of a Kernel in Complex Langevin Systems*, *Nucl. Phys. B* **324** (1989) 684.
- [9] D. Sexty, *Simulating full QCD at nonzero density using the complex Langevin equation*, *Phys. Lett. B* **729** (2014) 108.

- [10] J. Berges, S. Borsanyi, D. Sexty and I.O. Stamatescu, *Lattice simulations of real-time quantum fields*, *Phys. Rev. D* **75** (2007) 045007 [hep-lat/0609058].
- [11] D. Alvestad, R. Larsen and A. Rothkopf, *Towards learning optimized kernels for complex Langevin*, *Journal of High Energy Physics* **2023** (2023) 57 [2211.15625].
- [12] N.M. Lampl and D. Sexty, *Real time evolution of scalar fields with kernelled Complex Langevin equation*, Sept., 2023. 10.48550/arXiv.2309.06103.
- [13] D. Alvestad, A. Rothkopf and D. Sexty, *Lattice real-time simulations with learned optimal kernels*, *Phys. Rev. D* **109** (2024) L031502.
- [14] L.V. Keldysh, *Diagram Technique for Nonequilibrium Processes*, *Sov. Phys. JETP* **20** (1965) 1018.
- [15] G. Parisi and Y.-s. Wu, *Perturbation Theory Without Gauge Fixing*, *Sci. Sin.* **24** (1981) 483.
- [16] M. Scherzer, E. Seiler, D. Sexty and I.-O. Stamatescu, *Complex Langevin and boundary terms*, *Phys. Rev. D* **99** (2019) 014512.
- [17] M. Scherzer, E. Seiler, D. Sexty and I.-O. Stamatescu, *Controlling Complex Langevin simulations of lattice models by boundary term analysis*, *Physical Review D* **101** (2020) 014501 [1910.09427].
- [18] M.W. Hansen, M. Mandl, E. Seiler and D. Sexty, *The Role of Integration Cycles in Complex Langevin Simulations*, *Physical Review D* **111** (2025) 074502 [2412.17137].
- [19] K. Boguslavski, P. Hotzy and D.I. Müller, *Real-time correlators in 3+1D thermal lattice gauge theory*, *Physical Review D* **109** (2024) 094518 [2312.03063].
- [20] L.L. Salcedo and E. Seiler, *Schwinger-Dyson equations and line integrals*, *Journal of Physics A: Mathematical and Theoretical* **52** (2019) 035201 [1809.06888].
- [21] M. Mandl, E. Seiler and D. Sexty, *Necessary and sufficient conditions for correctness of complex Langevin*, *Journal of Physics A: Mathematical and Theoretical* **58** (2025) 495202 [2508.14512].