OPEN ACCESS

Computational Mathematics PERSPECTIVE

The Innovation Informatics

Optimization of neural network wave functions in variational Monte Carlo

Yuyang Wang[1,2,3] **and Xin Liu**[1,2,3,*]

[1]State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

[2]School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

[3]These authors contributed equally

*Correspondence: liuxin@lsec.cc.ac.cn

Received: September 24, 2025; Accepted: November 24, 2025; Published Online: November 27, 2025; https://doi.org/10.59717/j.xinn-inform.2026.100025

Citation: Wang Y. and Liu X. (2026). Optimization of neural network wave functions in variational Monte Carlo. *The Innovation Informatics* **2**:100025.

The central challenge in materials science and quantum chemistry is solving the electronic Schrodinger equation, complicated by the curse of dimensionality as system size grows. Neural network-based variational Monte Carlo (NN-VMC) offers a promising path forward, achieving unprecedented accuracy at far lower cost than traditional high-level methods. However, the flexibility of neural network wavefunctions introduces a bottleneck: their optimization is high-dimensional, stochastic, and nonconvex. This Perspective reviews the evolution of optimization methods in NN-VMC, from stochastic reconfiguration to approximate second-order algorithms and geometric insights. We highlight key challenges currently limiting scalability and efficiency, and outline future opportunities to advance the field. With continued progress in optimization, neural network techniques, and computer architectures, NN-VMC can tackle larger and more complex quantum systems and move from trailing experiments to guiding them.

INTRODUCTION

Quantum many-body problems are central to materials science and quantum chemistry, and a fundamental challenge is solving the stationary electronic Schrodinger equation, that is, finding the ground-state energy and eigenfunction of the many-electron Hamiltonian,

$$\mathcal{H} = -\sum_{i=1}^{N}\frac{1}{2}\Delta_{r_i} - \sum_{i=1}^{N}\sum_{I=1}^{M}\frac{Z_I}{\|r_i - R_I\|} + \sum_{1\le i<j\le N}\frac{1}{\|r_i - r_j\|} + \sum_{1\le I<I'\le M}\frac{Z_I Z_{I'}}{\|R_I - R_{I'}\|},$$

where r_i and R_I denote electron and nuclear coordinates (with charges Z_I), respectively. Achieving "chemical accuracy" is often required for reliable applications (e.g. Figure 1A). Conventional electronic structure methods face an accuracy-cost trade-off: high-level wavefunction methods are accurate but scale poorly, whereas lower-cost approaches like density functional theory sacrifice accuracy.

Variational Monte Carlo (VMC) offers a compromise between accuracy and efficiency. It turns the eigenvalue problem into a Rayleigh-quotient minimization, where a parametric trial wavefunction ψ_θ is optimized to minimize the Hamiltonian expectation:

$$\min_{\theta\in\mathbb{R}^{N_p}} L_{\mathrm{VMC}}(\theta) := \mathbb{E}_{r\sim\pi_\theta}\left[E_{\mathrm{loc}}(r;\theta)\right],$$

with local energy $E_{\mathrm{loc}}(r;\theta) := \psi_\theta^{-1}(r)\mathcal{H}\psi_\theta(r)$ and sampling distribution $\pi_\theta \propto |\psi_\theta|^2$. In practice, all expectations are estimated via Monte Carlo sampling.

VMC involves three core tasks: constructing the wavefunction, sampling from π_θ, and optimizing parameters (e.g. Figure 1B). Deep learning has greatly enhanced the expressiveness of trial wavefunctions. Meanwhile, improved variance-reduction Monte Carlo schemes have increased sampling efficiency and stabilized training (see Webber and Lindse and the references therein).[2] As a result, neural-network-based VMC (NN-VMC) has achieved chemical accuracy on atoms and small molecules, sometimes surpassing traditional high-accuracy methods.[1] These successes, often achieved at lower cost, highlight NN-VMC's potential to bridge the efficiency-accuracy gap (e.g. Figure 1C).

However, fully realizing NN-VMC's potential has exposed a major challenge in optimization. The flexibility of neural network wavefunctions and the complexity of the Hamiltonian expectation make the optimization landscape high-dimensional, nonconvex and ill-conditioned. Combined with stochastic noise and limited structure, this renders standard preconditioners, which scale updates only by energy gradient, ineffective at capturing subtle wavefunction variations. First-order optimizers like Adam often perform poorly. Moreover, since modern architectures were designed for well-behaved losses suited to Adam, the FermiNet's physics-informed design and the Hamiltonian expectation objective further limit its effectiveness.[1] Instead, practitioners rely on physics-inspired methods, like stochastic reconfiguration (SR)-based methods,[1,3-5] and variants of the linear method.[2,6] These exploit wavefunction geometry and have been critical for high accuracy. However, they remain sensitive to hyperparameters, require careful regularization, and scale poorly with system size. As shown in Figure 1E, NN-VMC applications remain limited to small systems, making improved optimization essential for future progress.

METHODOLOGICAL EVOLUTION OF VMC OPTIMIZATION

In recent decades, various strategies have been developed to tackle VMC optimization. Broadly, these methods fall into three categories, presented below along the timeline of their development.

Stochastic reconfiguration-based methods

SR remains the most widely used VMC optimizer.[3] Essentially, SR projects an infinitesimal imaginary-time evolution onto the variational parameter space under the Fubini-Study metric (e.g. Figure 1D), yielding a linear system each iteration:

$$S(\theta)\Delta\theta = -g(\theta),$$

where $g(\theta)$ is the energy gradient and $S(\theta)$ is the covariance matrix of $\nabla_\theta log|\psi_\theta|$ known as the SR matrix.[1] Solving this $N_p \times N_p$ system costs $\mathcal{O}(N_p^3)$, which is prohibitive for neural networks with millions of parameters. Moreover, since all ingredients are estimated from Monte Carlo samples $N_s \ll N_p$, $S(\theta)$ is often singular, necessitating careful regularization. In recent years, several developments have sought to address these issues and make SR more scalable and robust.

One important development was the adaptation of Kronecker-factored approximate curvature (K-FAC) method for VMC. Originally proposed to approximate natural gradient in deep learning,[7] it was first applied to NN-VMC in FermiNet to approximate $S(\theta)$ with a block-diagonal structure,[1] reducing per-iteration cost to roughly $\mathcal{O}(n_l^3 l)$, where n_l is the maximum hidden dimension and l is the number of layers. However, K-FAC must be tailored to each network and works best when $S(\theta)$ is well captured by layer-wise Kronecker factors; it may struggle for more complex ansatz.

Another set of methods tackles the issue that when $N_s \ll N_p$, the SR equations are underdetermined. MinSR selects the minimum-norm solution, using the Sherman-Morrison formula reduces the per-iteration cost to $\mathcal{O}(N_s^3)$.[4] SPRING views SR updates as Kaczmarz projections, iteratively projecting parameter updates along each sample's direction, similarly achieving $\mathcal{O}(N_s^3)$.[5] In practice, SPRING often outperforms MinSR and K-FAC, but it can diverge for certain hyperparameter choices, highlighting the need to deeper understand hyperparameters in SR-based optimizers.

Approximate second-order methods

Researchers have long explored second-order optimization methods for VMC. The linear method is a classic example: it optimizes in the subspace spanned by wavefunction's parameter derivatives, yielding a generalized eigenvalue problem of size N_p.[8] While very effective for few parameters, it becomes impractical for neural network wavefunctions due to its unreduced

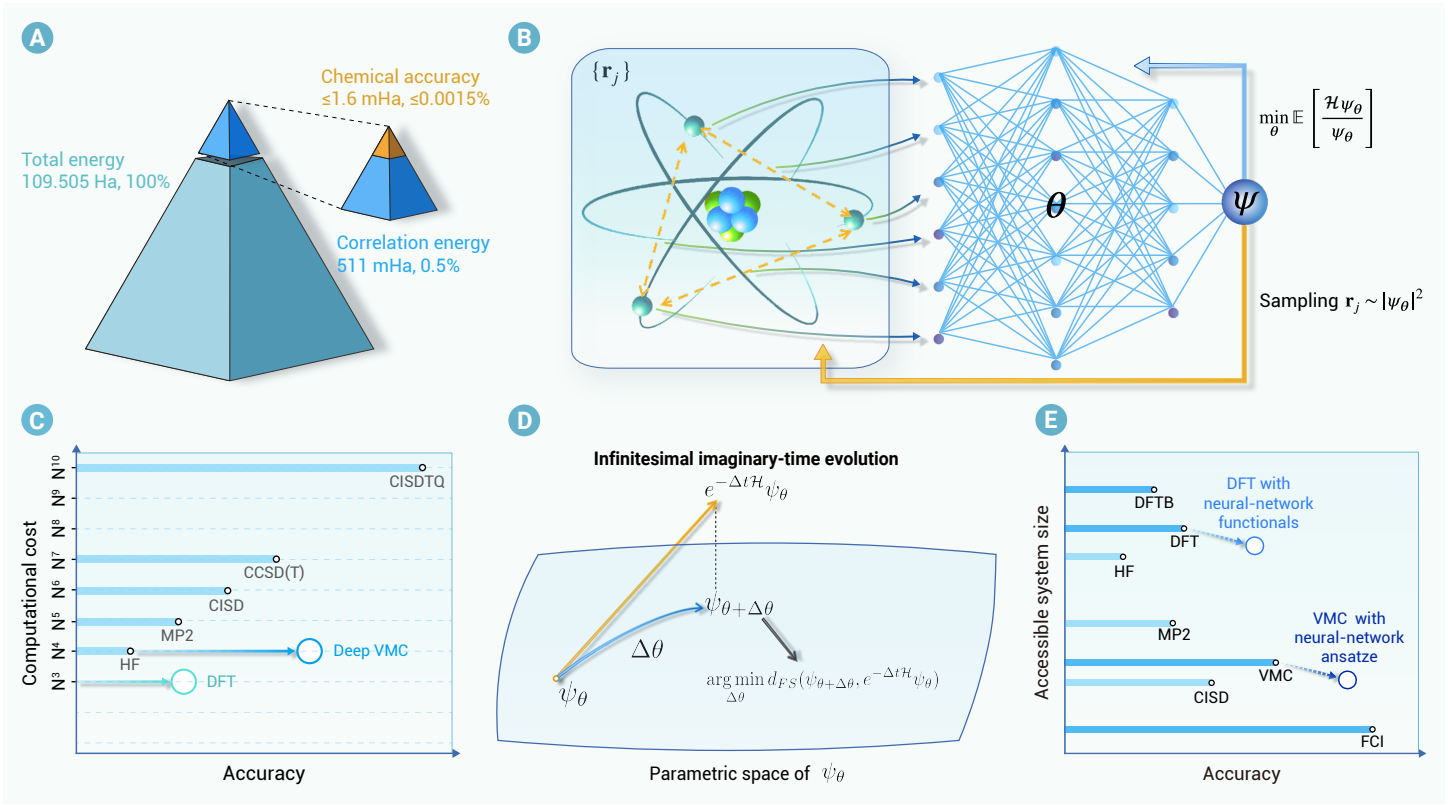

Figure 1. Workflow, comparison and SR update of NN-VMC (A) Ground-state energy of Nitrogen molecule. (B) NN-VMC workflow. (C) Cost-accuracy comparison of different methods. (D) Illustration of SR. (E) Scope of different methods by system size and accuracy.

$\mathcal{O}(N_p^3)$ cost.

The Rayleigh–Gauss–Newton (RGN) method is a compromise between SR and the linear method.[2] It derives from second-order expansions of the wavefunction and energy, retaining certain Hessian terms while neglecting others. On lattice models, RGN converges in fewer iterations than SR.[2] However, applying it to electronic systems remains costly: each iteration requires solving an N_p-dimensional linear system and computing parameter derivatives of the local energy, roughly doubling per-iteration cost. These costs currently limit RGN to proof-of-concept studies on small systems.

Manifold and geometric perspectives

Another emerging direction leverages the geometry insights. Fundamentally, ground-state computation is an infinite-dimensional eigenvalue problem. The objective can naturally be viewed as a functional on the manifold of normalized wavefunctions or probability densities rather than the variational parameters. Respecting this geometric structure may yield algorithms that converge more efficiently.

The Wasserstein QMC (WQMC) endows the Wasserstein–Fisher–Rao (WFR) metric over probability manifold.[9] It evolves the probability density π_θ along the WFR gradient flow, then projects this evolution back to the parameter space under KL divergence, analogous to SR in Figure 1D. While effective for small molecules, WQMC remains costly: each iteration requires two extra backpropagations for local-energy derivatives and an N_p-dimensional linear solve. Nevertheless, exploring metrics beyond the standard Fisher-Rao geometry could lead to more efficient optimization trajectories.

Notably, both SR and RGN can be interpreted as approximate manifold optimization on the unit sphere of wavefunctions.[6] By treating parameter derivatives as a tangent space basis, SR arises as a Galerkin-projected Riemannian gradient descent, while RGN corresponds to an approximate Riemannian Newton step. This viewpoint inspired the projected inverse iteration (PII) method, which applies a damped second-order step on the wavefunction manifold, essentially performing an inverse power iteration in the function space. PII showed promising results on lattice models, suggesting further geometric insights could improve NN-VMC. However, it remains costly, at each iteration requires an N_p-dimensional linear solve and additional backpropagations to assemble the approximate Riemannian Hessian.

CURRENT CHALLENGES

Despite algorithmic progress, several fundamental challenges still impede the efficient optimization of neural network wavefunctions.

Underdetermined subproblem and choice of update

The SR linear system is underdetermined, admitting infinitely many possible updates. While practical variants like MinSR and SPRING provide specific choices, their influence on the optimization trajectory remains unclear. Establishing principled criteria for selecting an optimal update direction is an open problem.

Lack of theoretical understanding of the SR matrix

Although SR-based methods remain central to VMC, the theoretical properties of the SR matrix are poorly understood. How network architecture and system physics shape its spectrum, and how this affects convergence and efficiency, remains unknown. This lack of insight hampers further improvement of SR methods.

Lack of convergence theory

So far, in NN-VMC, only stochastic gradient descent has been analyzed to converge to stationary points, not the true ground-state. Correlated and biased MCMC samples, together with the evolving distribution, make convergence analysis challenging. Without theoretical guidance, algorithm design relies heavily on empirical tuning and trial-and-error. This hinders understanding of why certain methods work and prevents deeper insights that could drive fundamental advances.

Scaling law of NN-VMC

Larger neural networks are increasingly used to improve VMC accuracy, yet it is unclear whether scaling network size consistently yields proportional gains. Establishing scaling laws could guide the selection of appropriate architectures and optimizers for specific systems, preventing unnecessary computational overhead.

FUTURE OPPORTUNITIES

The challenges above point to several opportunities and research directions to improve optimization in NN-VMC.

Advanced manifold optimizers

Manifold optimization techniques, like Riemannian conjugate gradient and trust-region methods, could accelerate NN-VMC atop the SR framework and to better navigate the wavefunction landscape. However, adapting these methods to stochastic settings remains a major challenge.

Hessian approximations at lower cost

Existing second-order methods are often impractical due to additional backpropagations. A more efficient route is to approximate Hessian effects using only first-order quantities like SR matrix and gradients, possibly through secant equations. This strategy could capture much of the acceleration of Newton-like updates while keeping the computational cost comparable to first-order methods.

Understanding of hyperparameters

Adaptive tuning has improved NN-VMC optimizers like K-FAC, yet their influence on stability and performance remains poorly understood. For instance, SPRING's momentum parameter is highly sensitive: an improper value easily destabilizes training. Linking such parameters to measurable quantities like variance or convergence rate, could enable more interpretable and robust adaptation.

Exploring new metrics

The WFR-based formulation suggests the choice of metric fundamentally shapes optimization behavior. Exploring alternatives, like metrics from quantum information theory (e.g. the Bures metric) may reveal new trade-offs and yield more efficient optimization paths.

Advanced neural network techniques

Embedding more physics into network architectures shrinks the function class, potentially smoothing the variational landscape, and accelerating convergence. For example, the hierarchical interaction model encodes known interactions directly into the wavefunction, yielding high accuracy.[10] Designing wavefunction-tailored networks can yield more compact ansatz that achieve high accuracy with less training.

Theoretical foundations of neural network wavefunctions

A deeper theoretical understanding of neural network wavefunctions will be invaluable for guiding future developments. Insights from optimization and statistical learning theory, especially understanding the loss landscape, could inform better initialization and optimizing schemes.

Computer architecture

Hybrid quantum-classical algorithms present an intriguing possibility, using quantum co-processors to assist parts of the VMC loop (e.g., sampling). While still in early stages, such hybrid approaches could eventually tackle larger electron counts or more complex interactions by offloading tasks to quantum hardware.

CONCLUSION

NN-VMC has rapidly advanced the accuracy, bringing us closer to solving many-electron problems with chemical accuracy at reasonable cost. How to efficiently train these powerful but complex wavefunctions is the key challenge. This Perspective outlined the progress in tackling this optimization problem and the challenges that remain. The field stands at an inflection point where insights from physics, numerical analysis, and machine learning are converging. By continuing to improve optimization algorithms and theoretical understanding of them, we can unlock the full potential of NN-VMC. Success will mean pushing simulations to larger molecules and materials, enabling new discoveries in chemistry and condensed matter physics. The coming years hold exciting opportunities to make "AI-powered" VMC not just accurate and efficient, but also robust and widely applicable.

REFERENCES

1. Pfau D., Spencer J., Matthews A., et al. (2020). Ab initio solution of the many-electron Schrodinger equation with deep neural networks. *Phys. Rev. Res.* **2**:033429. DOI:10.1103/PhysRevResearch.2.033429
2. Webber R. and Lindsey M. (2022). Rayleigh-Gauss-Newton optimization with enhanced sampling for variational Monte Carlo. *Phys. Rev. Res.* **4**:033099. DOI:10.1103/PhysRevResearch.4.033099
3. Sorella S. (1998). Green function Monte Carlo with stochastic reconfiguration. *Phys. Rev. Lett.* **80**:4558. DOI:10.1103/PhysRevLett.80.4558
4. Chen A. and Heyl M. (2024). Empowering deep neural quantum states through efficient optimization. *Nat. Phys.* **20**:1476–1481. DOI:10.1038/s41567-024-02566-1
5. Goldshlager G., Abrahamsen N. and Lin L. (2024). A Kaczmarz-inspired approach to accelerate the optimization of neural network wavefunctions. *J. Comput. Phys.* **516**:113351. DOI:10.1016/j.jcp.2024.113351
6. Armegioiu V., Carrasquilla J., Mishra S., et al. (2025). Functional neural wavefunction optimization. *arXiv* preprint. DOI:10.48550/arXiv.2507.10835
7. Martens J. and Grosse R. (2015). Optimizing neural networks with Kronecker-factored approximate curvature. *In Proceedings of the 32nd ICML of ICML'15. PMLR.* **37**:2408–2417. DOI:10.5555/3045118.3045374
8. Umrigar C. and Filippi C. (2005). Energy and variance optimization of many-body wave functions. *Phys. Rev. Lett.* **94**:150201. DOI:10.1103/PhysRevLett.94.150201
9. Neklyudov K., Nys J., Thiede L., et al. (2023). Wasserstein quantum Monte Carlo: A novel approach for solving the quantum many-body Schrodinger equation. *In Adv. Neural Inf. Process.* **36**:63461–63482. DOI:10.5555/3666122.3668892
10. Zhou D., Chen H., Ho C.H., et al. (2024). A multilevel method for many-electron Schrodinger equations based on the atomic cluster expansion. *SIAM J. Sci. Comput.* **46**:A105–A129. DOI:10.1137/23M1565887

FUNDING AND ACKNOWLEDGMENTS

The work of Xin Liu was supported in part by the National Natural Science Foundation of China (12125108, 12021001, 12288201), and RGC grant JLFS/P-501/24 for the CAS AMSS–PolyU Joint Laboratory in Applied Mathematics. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Yuyang Wang and Xin Liu contributed equally to the manuscript and approved the final version. All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

Xin Liu is an Editorial Board member of The Innovation Informatics and was blinded from reviewing or making final decisions on the manuscript. Peer review was handled independently of this member and their research group. The other authors declare no conflicts of interest.