

Reconciling Theory and Practice: Quadratic Acceleration of Variational Quantum Algorithms with Rotosolve*

Sayantana Pramanik^{1,2}, M Girish Chandra¹

¹TCS Research, TATA Consultancy Services, ²Indian Institute of Science

{sayantan.pramanik, m.gchandra}@tcs.com

Abstract

The **Rotosolve** algorithm [1, 2, 3, 4] leverages explicit knowledge about the structure of variational quantum circuits [5, 6] and objective functions to perform exact minimisation along one coordinate at a time. In a recent paper [7], we showed that the said method is not a heuristic. We established its convergence, and derived explicit rates under different assumptions on the objective function. The worst-case rates were found to match those of Randomised Coordinate Descent (RCD) [7, 8, 9, 10]. However, since **Rotosolve** uses additional information about the objective function, it is expected to outperform generic gradient-based optimisers like RCD. Empirical studies indeed demonstrate its superior performance [7, 11]. Further, the variance in the performance of **Rotosolve** across multiple experiments was inexplicably found to be higher than gradient-based methods [7, 11].

Questions: These observations suggest a disconnect between theory and practice, and naturally raise the following two questions: (1) Can sharper rates of convergence be derived for **Rotosolve** to bridge the gap between theory and practice? (2) Is it possible to theoretically justify the higher variance observed with **Rotosolve**?

Contributions: In this extended abstract, which acts as a followup to [7], our contributions towards answering the above questions are as follows:

- C.1** We start by noting that establishment of better convergence rates for **Rotosolve** requires the consideration of a stronger measure of performance than first-order stationarity. We introduce the notion of coordinate-wise second-order stationarity [Definition 1] which is quantified by $\|\rho(\theta)\|$. This measure is zero iff the first- and second-derivatives are zero and non-negative, respectively, along each of the coordinates. We refer to a point where $\|\rho(\theta)\| = 0$ as a *Coordinate-wise Second-order Stationary* (CSS) point.
- C.2** Next, we consider the convergence of **Rotosolve** and gradient-based methods in the noise-free case. While the latter can be shown to converge only to ε -first-order stationary points, **Rotosolve** is guaranteed to converge to ε -CSS points in the ℓ_1 norm. More specifically, we show that for a circuit with d optimisable parameters, the number of iterations needed by **Rotosolve** to converge to a point satisfying $\mathbb{E}[\|\rho(\theta)\|_1] \leq \varepsilon$ is $\mathcal{O}(d/\varepsilon)$ [Theorem 1]. On the other hand, if gradient-based methods *happen* to converge to an ε -CSS point in the ℓ_2 norm, they require $\mathcal{O}(d/\varepsilon^2)$ iterations [Remark 2]. Thus, in the worst case, **Rotosolve** provides a **quadratic acceleration** in terms of ε . We further argue and show that even in the absence of noise, **Rotosolve** has a higher variance than gradient-based methods [Remark 1].
- C.3** Finally, we extend our analysis to account for shot noise. Under an *additive noise model*, we establish conditions under which **Rotosolve** retains its iteration complexity of $\mathcal{O}(d/\varepsilon)$ [Theorem 2]. Since these assumptions are not guaranteed to hold in practice, we are motivated to consider a *relative-error noise model* instead, and modify **Rotosolve** accordingly. Owing to space constraints, discussion on the relative noise model, the modified algorithm, and its analysis are deferred to an extended version of this paper.

Due to the page limit, we state our results informally and omit their proofs. We complement our theoretical analysis with a concise set of experiments that corroborate our theoretical findings.

1 Introduction

We operate under the same setting as described in [7]. Briefly, the ansatz is composed of fixed, and parametrised single-qubit rotation gates. The components of the parameter-vector $\theta \in \mathbb{R}^d$ are uncorrelated with each other, and the final state is given by $|\psi(\theta)\rangle$. The spectrum of the Hermitian observable $\mathbf{H}$ is bounded in the interval $[-\lambda_{\min}, \lambda_{\max}]$, with $\lambda_{\min}, \lambda_{\max} \geq 0$. The resulting problem to be solved is:

*in the absence of noise.

Problem 1.

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^d} (f(\boldsymbol{\theta}) := \langle \psi(\boldsymbol{\theta}) | \mathbf{H} | \psi(\boldsymbol{\theta}) \rangle).$$

While generic gradient-based optimisers, such as Stochastic Gradient Descent (SGD) [12] and RCD, take a small step along the negative of the gradient, **Rotosolve** performs exact minimisation along one coordinate at a time. This is possible because, in the given setting, if all-but-one parameters are fixed, the resulting univariate objective function assumes a sinusoidal form, whose coefficients can be ascertained with three evaluations of the objective function along the chosen direction. As in [7], for ease of analysis, we consider a version of the algorithm where at each iteration, the index of the parameter to be updated is chosen randomly. Let the index j be selected at the t^{th} iteration of the algorithm, and let $\mathbf{e}_j$ denote the corresponding unit vector. Further, the first- and second-derivatives along $\mathbf{e}_j$ are given by $\nabla_j f(\boldsymbol{\theta}^{(t)})$ and $\nabla_j^2 f(\boldsymbol{\theta}^{(t)})$, respectively. Then, the parameters are updated as [7]:

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \left(\pi/2 + \arctan2(-2\nabla_j^2 f(\boldsymbol{\theta}^{(t)}), 2\nabla_j f(\boldsymbol{\theta}^{(t)})) \right) \mathbf{e}_j. \quad (1)$$

Another noteworthy fact about the **Rotosolve** algorithm is that, contrary to popular opinion, it is a second-order method that implicitly uses first- and second-derivatives to update the parameters [7]. In fact, in the vicinity of stationary points, the algorithm reduces to Newton’s method in one dimension:

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \left((\nabla_j^2 f(\boldsymbol{\theta}^{(t)}))^{-1} \nabla_j f(\boldsymbol{\theta}^{(t)}) + c \right) \mathbf{e}_j, \quad (2)$$

where, $c = 0$ if $\nabla_j^2 f(\boldsymbol{\theta}^{(t)})$ is positive, and $c = \pi$ otherwise. It is also remarkable that unlike Newton’s method, which is prone to divergence unless the algorithm is initiated within a “Newton’s region”, **Rotosolve** is guaranteed to converge from any initial point [7].

2 A New Measure of Stationarity

The objective landscape induced by variational quantum algorithms is highly non-convex [10], and contains numerous local minima [13] and *trainable pockets* [14]. As global optimality is generally unattainable for non-convex objective functions, optimization algorithms are instead analyzed in terms of their convergence to stationary points. For smooth, non-convex problems, first-order gradient-based methods are typically guaranteed to find an ε -first-order stationary point, i.e., a point satisfying $\|\nabla f(\boldsymbol{\theta})\|_p \leq \varepsilon$, within a finite number of iterations [12]. This notion of stationarity, however, depends solely on the gradient and fails to distinguish (local) minima from saddle points or (local) maxima, all of which satisfy the first-order necessary condition.

When analysed under the conventional notion of first-order stationarity, **Rotosolve** is found to have the same rate of convergence as first-order methods [7]. However, as discussed above, **Rotosolve** exploits additional information about the trigonometric form of the landscape, along with first- and second-derivatives, to perform exact minimisation along a single coordinate at each iteration. We therefore argue that its convergence should be characterized using a stronger notion of stationarity that reflects this additional structure. Since **Rotosolve** updates only one coordinate at a time, without considering any *interaction* between the coordinates, the desired stationarity measure should also be coordinate-wise separable. Motivated by this observation, and inspired by [15, 16], we introduce the notion of *Coordinate-wise Second-order Stationarity*, which incorporates both first- and second-order derivative information along individual coordinates. We subsequently define a corresponding measure of stationarity, denoted by $\|\boldsymbol{\rho}(\boldsymbol{\theta})\|_p$, to quantify the extent to which a point satisfies this condition. These notions are formally defined below.

Definition 1 (Coordinate-wise Second-order Stationarity). *A point $\boldsymbol{\theta} \in \mathbb{R}^d$ is said to be coordinate-wise second-order stationary if $\nabla_j f(\boldsymbol{\theta}) = 0$ and $\nabla_j^2 f(\boldsymbol{\theta}) \geq 0$ for all $j \in [d]$. We quantify the violation of this property through the vector-valued mapping $\boldsymbol{\rho}_j(\boldsymbol{\theta}) := \sqrt{(\nabla_j^2 f(\boldsymbol{\theta}))^2 + (\nabla_j f(\boldsymbol{\theta}))^2} - \nabla_j^2 f(\boldsymbol{\theta})$; and define the corresponding stationarity measure as $\|\boldsymbol{\rho}(\boldsymbol{\theta})\|_p$. If $\|\boldsymbol{\rho}(\boldsymbol{\theta})\|_p = 0$, then $\boldsymbol{\theta}$ is referred to as a CSS point.*

The connection between **Rotosolve** and $\boldsymbol{\rho}(\boldsymbol{\theta})$ will become apparent in [Lemma 1](#). It is trivial to verify that for all $j \in [d]$ and $\boldsymbol{\theta} \in \mathbb{R}^d$, $\boldsymbol{\rho}_j(\boldsymbol{\theta}) \geq 0$, $\boldsymbol{\rho}_j(\boldsymbol{\theta}) = 0$ iff $\nabla_j f(\boldsymbol{\theta}) = 0$ and $\nabla_j^2 f(\boldsymbol{\theta}) \geq 0$, and $\|\boldsymbol{\rho}(\boldsymbol{\theta})\|_p = 0$ implies that $\boldsymbol{\theta}$ is a first-order stationary point. Additionally, a CSS point is a second-order local minimum if the Hessian matrix is *diagonally dominant*.

3 Convergence Rates in the Absence of Noise

We now analyse the convergence of **Rotosolve** in the absence of noise, and start by establishing a relation between the expected descent at each iteration and $\|\rho(\theta)\|$ in the following lemma:

Lemma 1 (Rotosolve Descent Lemma, Informal). *With **Rotosolve**, the expected change in the objective value between consecutive iterations is:*

$$\mathbb{E}_j \left[f(\theta^{(t+1)}) - f(\theta^{(t)}) | \theta^{(t)} \right] = -\frac{\|\rho(\theta^{(t)})\|_1}{d}. \quad (3)$$

Remark 1. *It is remarkable that equality holds in Eq. (3), which allows us to find the variance of the descent at each iteration. While the variance can be modulated by the step-size in gradient-based methods, it is upper bounded by $\|\rho(\theta^{(t)})\|_2^2/d$ with **Rotosolve**. This behaviour is clearly evident in Fig. 1a.*

Equipped with the result in Lemma 1, we can now bound the iteration complexity of **Rotosolve**.

Theorem 1 (Convergence rate of **Rotosolve**, Informal). *In the setting of Sec. 1, **Rotosolve** finds an ε -CSS point of Prob. 1, satisfying $\mathbb{E}[\|\rho(\theta)\|_1] \leq \varepsilon$, in $\mathcal{O}(d/\varepsilon)$ iterations.*

Remark 2. *In contrast to **Rotosolve**, gradient-based methods such as SGD and RCD guarantee convergence only to an ε -first-order stationary point. As discussed in Sec. 2, first-order stationarity is a weaker condition than coordinate-wise second-order stationarity. Consequently, **Rotosolve** is guaranteed to converge to a stronger notion of stationarity. If, in addition, the point obtained by SGD or RCD happens to be ε -CSS, then the iteration complexity required to obtain a point such that $\mathbb{E}[\|\rho(\theta)\|_2^2] \leq \varepsilon^2$ remains $\mathcal{O}(d/\varepsilon^2)$ [7]. It should be noted that in this case, the ε -CSS condition is satisfied in the ℓ_2 norm.*

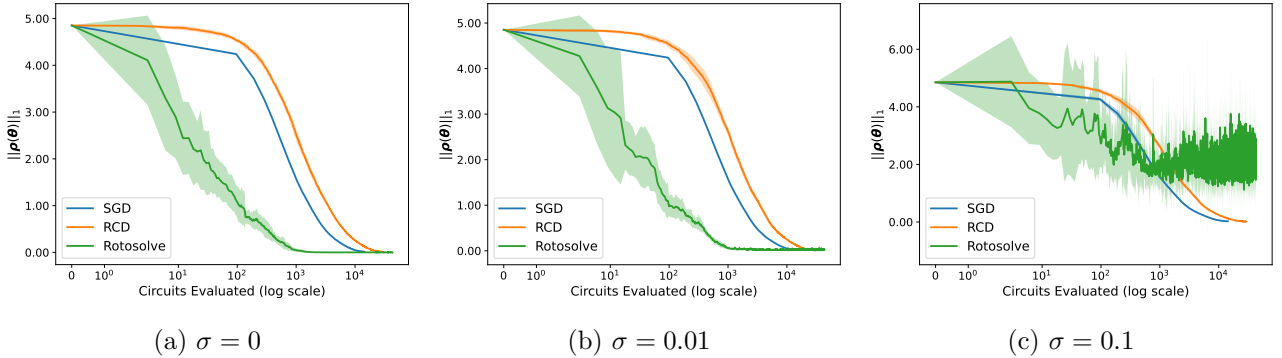


Figure 1: Experimental results (replicated and adapted from [17]) for estimating the ground-state energy of the water molecule. From (a) to (c), the standard deviation σ of noise increases.

4 Convergence Rates under Finite Measurements

Finally, we incorporate the effect of shot-noise into the analysis. We assume that we have access to stochastic estimates $\hat{f}(\theta) = f(\theta) + \xi$, where ξ is a random variable with zero-mean and variance $\leq \sigma^2$ [7]. Additionally, our analysis requires that the noise does not cause the arctan2 function in Eq. (1) to change its quadrant, and that for all $\theta \in \mathbb{R}^d$, $|\xi| \ll |f(\theta)|$.

Theorem 2 (Convergence rate of **Rotosolve** with shot-noise, Informal). *Let the above assumptions be satisfied, and let $\sigma^2 \leq \varepsilon\delta/d$, where $\delta := \inf_{\theta \in \mathbb{R}^d, j \in [d]} \sqrt{(\nabla_j^2 f(\theta))^2 + (\nabla_j f(\theta))^2}$. Under these conditions, **Rotosolve** finds an ε -CSS point of Prob. 1, satisfying $\mathbb{E}[\|\rho(\theta)\|_1] \leq \varepsilon$, in $\mathcal{O}(d/\varepsilon)$ iterations, and with a total of $\mathcal{O}(d^2/\delta\varepsilon^2)$ measurements.*

In contrast, SGD and RCD require $\mathcal{O}(d^2/\varepsilon^4)$ measurements. The performance of these algorithms has been compared in Figs. 1b and 1c. The observed degradation in **Rotosolve**'s performance, as the noise variance increases and exceeds the theoretical threshold of $\varepsilon\delta/d$, is consistent with Theorem 2, whereas SGD and RCD appear considerably more robust to noise. However, it should be noted that while the *additive* noise-model is a standard assumption [10, 11], the conditions required for our analysis are generally not satisfied in practice. This motivates us to adopt a *relative* shot-noise model and develop a modified version of **Rotosolve**. Owing to space constraints, the corresponding assumptions, algorithmic modifications, and theoretical analysis are deferred to the full version of this paper.

References

- [1] J. G. Vidal and D. O. Theis, “Calculus on parameterized quantum circuits,” 2018. [Online]. Available: <https://arxiv.org/abs/1812.06323>
- [2] K. M. Nakanishi, K. Fujii, and S. Todo, “Sequential minimal optimization for quantum-classical hybrid algorithms,” *Physical Review Research*, vol. 2, no. 4, Oct. 2020. [Online]. Available: <http://dx.doi.org/10.1103/PhysRevResearch.2.043158>
- [3] M. Ostaszewski, E. Grant, and M. Benedetti, “Structure optimization for parameterized quantum circuits,” *Quantum*, vol. 5, p. 391, Jan. 2021. [Online]. Available: <https://doi.org/10.22331/q-2021-01-28-391>
- [4] R. M. Parrish, J. T. Iosue, A. Ozaeta, and P. L. McMahon, “A jacobi diagonalization and anderson acceleration algorithm for variational quantum algorithm parameter optimization,” 2019. [Online]. Available: <https://arxiv.org/abs/1904.03206>
- [5] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien, “A variational eigenvalue solver on a photonic quantum processor,” *Nature Communications*, vol. 5, no. 1, jul 2014. [Online]. Available: <https://doi.org/10.1038%2Fncomms5213>
- [6] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, “Variational quantum algorithms,” *Nature Reviews Physics*, vol. 3, no. 9, p. 625–644, Aug. 2021. [Online]. Available: <http://dx.doi.org/10.1038/s42254-021-00348-9>
- [7] S. Pramanik and M. G. Chandra, “One coordinate at a time: Convergence guarantees for rotosolve in variational quantum algorithms,” 2026. [Online]. Available: <https://arxiv.org/abs/2604.25613>
- [8] Y. Nesterov, “Efficiency of coordinate descent methods on huge-scale optimization problems,” *SIAM J. Optim.*, vol. 22, no. 2, pp. 341–362, Jan. 2012.
- [9] S. J. Wright, “Coordinate descent algorithms,” 2015. [Online]. Available: <https://arxiv.org/abs/1502.04759>
- [10] Z. Ding, T. Ko, J. Yao, L. Lin, and X. Li, “Random coordinate descent: A simple alternative for optimizing parameterized quantum circuits,” *Phys. Rev. Res.*, vol. 6, p. 033029, Jul 2024. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevResearch.6.033029>
- [11] Z. Lai, J. Hu, T. Ko, J. Wu, and D. An, “Interpolation-based coordinate descent method for parameterized quantum circuits,” *Communications Physics*, vol. 9, no. 1, Jan. 2026. [Online]. Available: <http://dx.doi.org/10.1038/s42005-025-02473-8>
- [12] G. Garrigos and R. M. Gower, “Handbook of convergence theorems for (stochastic) gradient methods,” 2024. [Online]. Available: <https://arxiv.org/abs/2301.11235>
- [13] E. R. Anschuetz and B. T. Kiani, “Quantum variational algorithms are swamped with traps,” *Nat. Commun.*, vol. 13, no. 1, p. 7760, Dec. 2022.
- [14] A. Kulshrestha, R. Puig, D. García-Martín, L. Cincio, I. Safro, Z. Holmes, and M. Cerezo, “Exponentially many initializations to avoid barren plateaus,” 2026. [Online]. Available: <https://arxiv.org/abs/2606.18515>
- [15] Y. Nesterov and B. T. Polyak, “Cubic regularization of newton method and its global performance,” *Math. Program.*, vol. 108, no. 1, pp. 177–205, Aug. 2006.
- [16] A. Beck and Y. C. Eldar, “Sparsity constrained nonlinear optimization: Optimality conditions and algorithms,” 2012. [Online]. Available: <https://arxiv.org/abs/1203.4580>
- [17] L. Ion and A. Smith, “Variational quantum eigensolver for chemical molecules,” 2026. [Online]. Available: <https://arxiv.org/abs/2512.22572>