



Benchmarking optimization methods in quantum information theory



IML Scholars: David Huerta, Zhiyuan Ma, Tingyu Wu

Graduate Mentor: Sujeet Bhalerao

Faculty Advisor: Felix Leditzky

1. Background

Definition (Density operator). A density operator or quantum state is a linear operator $\rho \in \mathcal{L}(\mathcal{H})$ on a Hilbert space $\mathcal{H}$ satisfying $\rho \succeq 0$ and $\text{tr}(\rho) = 1$.

Definition (Partial trace). Let $\{|i\rangle_B\}_{i=0}^{|B|-1}$ be an orthonormal basis of $\mathcal{H}_B$ (where $|B| = \dim \mathcal{H}_B$). The partial trace of an operator $X_{AB} \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is defined as

$$X_A := \text{tr}_B(X_{AB}) = \sum_{i=0}^{|B|-1} (\mathbb{1}_A \otimes \langle i|_B) X_{AB} (\mathbb{1}_A \otimes |i\rangle_B)$$

Definition (Quantum channel). A quantum channel is a linear, completely positive, trace-preserving map acting on density operators.

A quantum channel is a model for noise in an open quantum system S with environment E , given by unitary evolution of the larger system SE and tracing out the environment E .

Kraus representation of $\mathcal{N}$: Kraus operators $\{K_i\}$ satisfying $\sum_i K_i^\dagger K_i = \mathbb{1}$ and

$$\mathcal{N}(\rho) = \sum_i K_i \rho K_i^\dagger.$$

Channel models considered in this project:

- Pauli channels: For the Pauli matrices X, Y, Z and a probability vector (p_0, p_1, p_2, p_3) ,

$$\rho \mapsto p_0 \rho + p_1 X \rho X + p_2 Y \rho Y + p_3 Z \rho Z \quad (1)$$

- Depolarizing channel for $p \in [0, 1]$: $p_0 = 1 - p, p_1 = p_2 = p_3 = p/3$ in (1).
- Two-Pauli channel for $p \in [0, 1]$: $p_0 = 1 - p, p_1 = p_3 = p/2, p_2 = 0$ in (1).
- Low-noise Pauli channel for probability vector (p_x, p_y, p_z) and $t \in [0, 1]$: $p_0 = 1 - t, (p_1, p_2, p_3) = t(p_x, p_y, p_z)$ in (1).
- Qutrit-platypus channel [3]: For $s \in [0, 1]$,

$$\rho \mapsto K_0 \rho K_0^\dagger + K_1 \rho K_1^\dagger \quad \text{where} \quad \begin{aligned} K_0 &= \sqrt{s}|0\rangle\langle 0| + |2\rangle\langle 1| \\ K_1 &= \sqrt{1-s}|0\rangle\langle 1| + |2\rangle\langle 2|. \end{aligned}$$

- Damping-dephasing channel: For $p, g \in [0, 1]$, the channel has Kraus operators

$$K_0 = \sqrt{1-p}(|0\rangle\langle 0| + \sqrt{1-g}|1\rangle\langle 1|) \quad K_1 = \sqrt{g}|0\rangle\langle 1| \quad K_2 = \sqrt{p}(|0\rangle\langle 0| + \sqrt{1-g}|1\rangle\langle 1|).$$

2. Optimization in quantum information theory

2.1 Channel Coherent Information

Given a channel $\mathcal{N}$: $A \rightarrow B$, the *channel coherent information* is defined as

$$I_c(\mathcal{N}) = \max \{S(\mathcal{N}(\psi_A)) - S((\text{id}_R \otimes \mathcal{N})(\psi_{RA})) : |\psi\rangle_{RA} \text{ pure}\}$$

The channel coherent information is an achievable rate for quantum information transmission through $\mathcal{N}$ [1]. The *quantum capacity* of the channel $\mathcal{N}$ is equal to

$$Q(\mathcal{N}) = \lim_{n \rightarrow \infty} \frac{1}{n} I_c(\mathcal{N}^{\otimes n}).$$

3. Performance and Data profiles

3.1 Performance profiles

Define a problem set $\mathcal{P}$ and a set of solvers $\mathcal{S}$.

- More and Wild Convergence Test** [4]: For a minimization problem $\min\{f(x) : x \in \mathbb{R}^m\}$ and a predefined accuracy τ , an optimization run is said to converge if

$$\frac{f(x_0) - f(x)}{f(x_0) - f_L} \geq 1 - \tau. \quad (2)$$

(x_0 : starting point of optimization run; f_L : best value found by any solver $s \in \mathcal{S}$ given prescribed budget μ_p of function evaluations.)

- Performance measure** $t_{p,s} > 0$ for given $p \in \mathcal{P}$, $s \in \mathcal{S}$ and the convergence test (2).

In this project: $t_{p,s} = \#$ function evaluations.

- Performance Ratio**

$$r_{p,s} = \frac{t_{p,s}}{\min\{t_{p,s} : s \in \mathcal{S}\}}.$$

If solver s does not converge for problem p w.r.t. (2), we set $r_{p,s} = \infty$.

- Performance Profile** for $\alpha \geq 1$:

$$\rho_s(\alpha) = \frac{1}{|\mathcal{P}|} \# \{p \in \mathcal{P} : r_{p,s} \leq \alpha\}.$$

Performance profile measures the fraction of problems for which the performance of solver s is the best-performing solver up to a factor α .

3.2 Data profiles

Let n_p be the number of variables in a benchmark problem. With $t_{p,s}$ as defined above and n_p the number of variables in the problem, the *data profile* of a solver $s \in \mathcal{S}$ is defined for $\alpha \geq 1$ as

$$d_s(\alpha) = \frac{1}{|\mathcal{P}|} \# \left\{ p \in \mathcal{P} : \frac{t_{p,s}}{n_p + 1} \leq \alpha \right\}.$$

$t_{p,s} \equiv \infty$ if solver s does not converge for problem p according to the criterion (2).

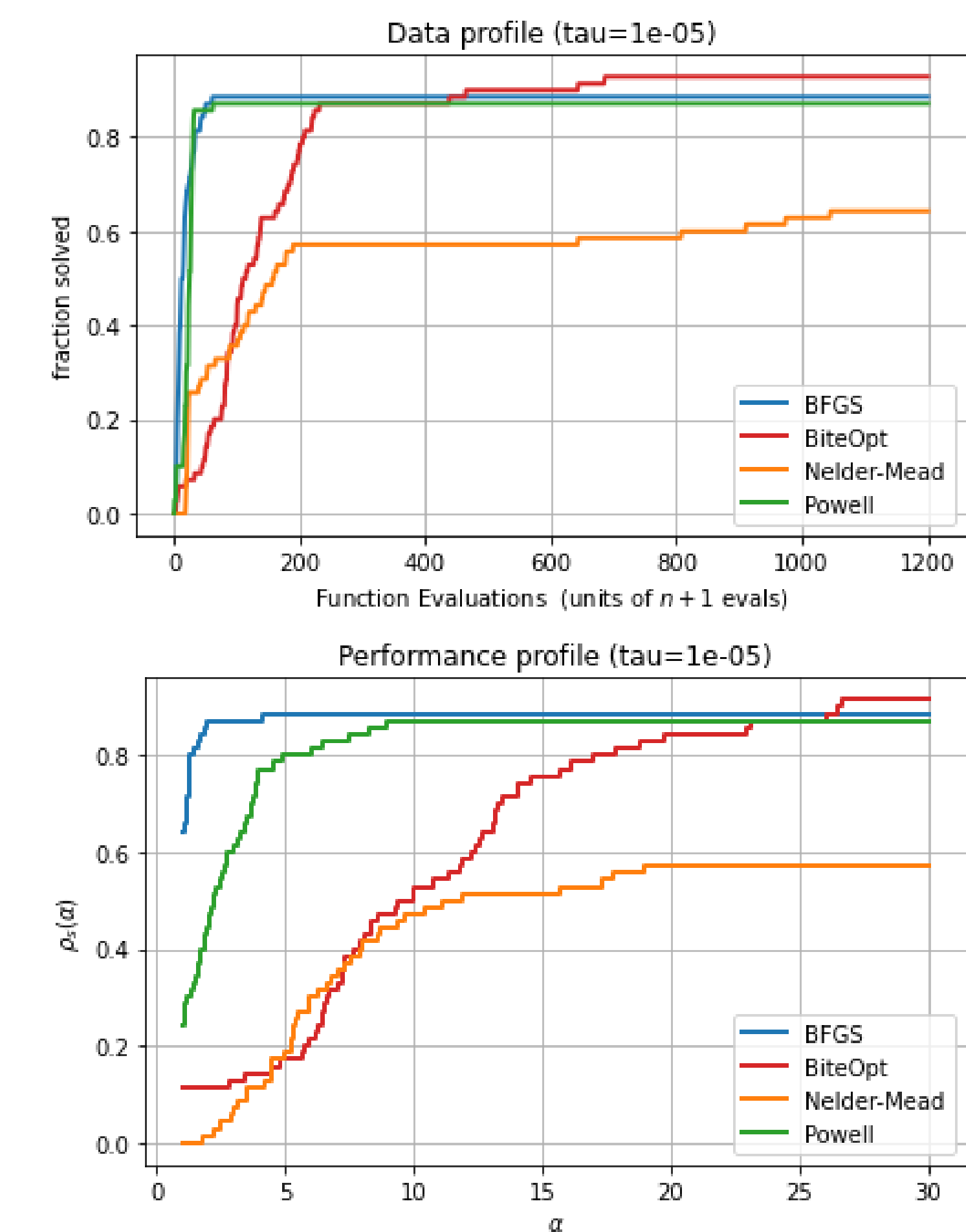
Data profiles are useful for deciding on a solver given a certain budget of function evaluations, measured in “gradient descent” units.

3.3 Solvers used in this project

- BFGS** A local, gradient-based optimizer that uses first-derivative information to build an approximate inverse Hessian (curvature) to move towards a (local) minimum.
- Nelder-Mead** A local, gradient-free optimizer that searches by repeatedly reshaping a simplex (a set of points) to move toward better objective values.
- Powell** A local, gradient-free optimizer that minimizes along a sequence of search directions (line searches) and updates those directions over time.
- BiteOPT** A global, stochastic, derivative-free optimizer for bound-constrained problems that evolves a population of candidate solutions using objective-function evaluations and a variety of solution generators.

4. Results

We compute performance and data profiles for the channel coherent information $I_c(\mathcal{N}^{\otimes n})$, where $\mathcal{N}$ is one of the channels listed in Section 1. We only consider channels for which the coherent information is *additive*, i.e., $I_c(\mathcal{N}^{\otimes n}) = nI_c(\mathcal{N})$, and in each case compute $I_c(\mathcal{N}^{\otimes n})$ for $n = 1, 2, 3$ for 5 random instances of each channel class, using a tolerance $\tau = 1 \times 10^{-5}$ in (2).



5. Conclusion

Our results show that the gradient based method performs the best on a majority of the problems, converging the fastest on about 65 percent of problems and converging on about 90 percent of total problems. Another conclusion that can be made is that, although BiteOPT is able to solve slightly more problems than the local optimizers, it does so at the cost of many more function evaluations.

6. Future Work

In future work we plan to conduct a similar analysis for two other entropic optimization problems, the minimum output entropy of a channel and the entanglement of formation of a mixed bipartite quantum state. We will also study channels for which the entropic functionals may be super-additive, e.g., $I_c(\mathcal{N}^{\otimes n}) > nI_c(\mathcal{N})$ for the channel coherent information.

References

- [1] Devetak, The private classical capacity and quantum capacity of a quantum channel. IEEE Transactions on Information Theory 51.1 (2005): 44-55. [2] Wilde, Quantum Information Theory (2nd ed.). Cambridge: Cambridge University Press, 2017. [3] Leditzky et al., The platypus of the quantum channel zoo. IEEE Transactions on Information Theory 69.6 (2023): 3825-3849. [4] Moré, Jorge J., and Stefan M. Wild. "Benchmarking derivative-free optimization algorithms." SIAM Journal on Optimization 20.1 (2009): 172-191.

Acknowledgements: Support for this project was provided by the Illinois Geometry Lab, the Department of Mathematics at the University of Illinois at Urbana-Champaign.