Conditions for Global Optimality in Quantum Arimoto–Blahut Algorithms

Geng Liu and Masahito Hayashi

The Chinese University of Hong Kong, Shenzhen



香港中文大學（深圳）
The Chinese University of Hong Kong, Shenzhen

Outline

01 **Introduction**

02 **Main Results**

- AB and mirror descent
  The constraint geometry determines when their updates agree.
- Global optimality of fixed points
  Compatibility is required at the endpoint.
- Finite-candidate assessment
  Feasible finite differences bound the objective gap.

03 **Summary**

From AB to Quantum AB

Classical Arimoto–Blahut

- classical probability distributions
- alternating optimization method
- computes channel capacity and rate–distortion

Quantum Arimoto–Blahut

- replace distributions by density matrices
- uses closed-form updates on quantum states
- optional projection handles linear constraints
- used for quantum capacities, rate–distortion, QIB, etc.

Why Quantum Arimoto-Blahut is attractive:

QAB is computationally efficient and simple to implement: simple updates, no explicit gradients.

Optimization over a mixture family

Mixture family

$$\mathcal{M} = \{\rho \geq 0 : Tr\rho = 1,\ \mathrm{Tr}(\rho H_j) = c_j,\ j = 1, \ldots, k\}$$

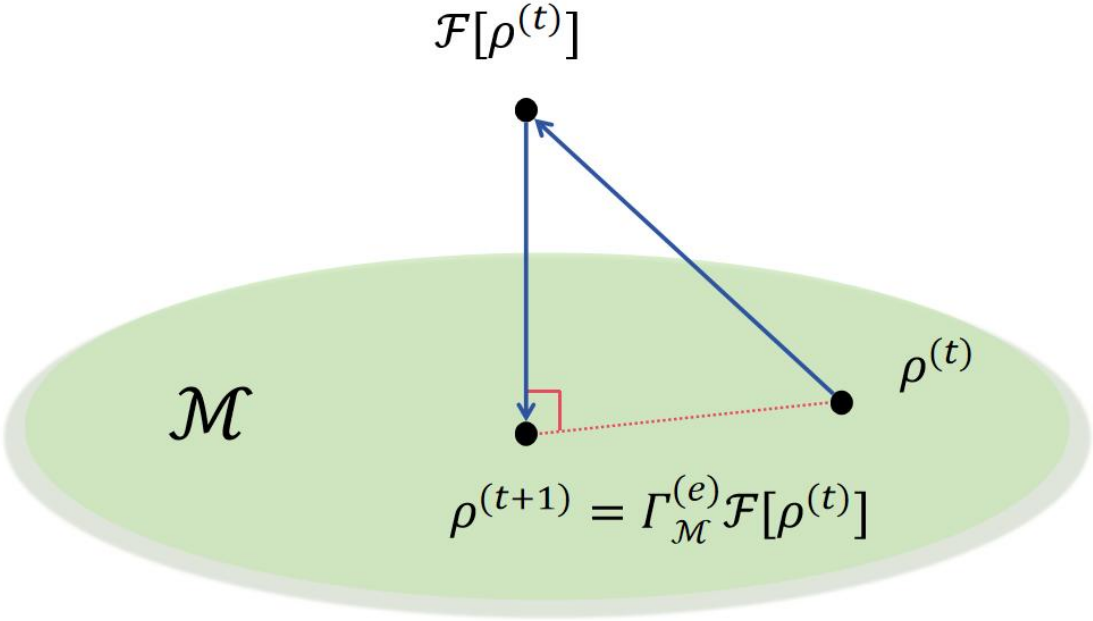
Objective and chosen Hermitian field

$$\min_{\rho \in \mathcal{M}} f(\rho), \qquad f(\rho) = \mathrm{Tr}[\rho \Omega(\rho)]$$

Standing assumptions: convex differentiable f; an interior feasible state

$$\mathcal{M}_{++} = \{\rho \in \mathcal{M} : \rho \succ 0\}$$

The generalized quantum AB update



1 Exponential step

$$\mathcal{F}[\rho] = \frac{\exp(\log\rho - \gamma^{-1}\Omega[\rho])}{\mathrm{Tr}\exp(\log\rho - \gamma^{-1}\Omega[\rho])}$$

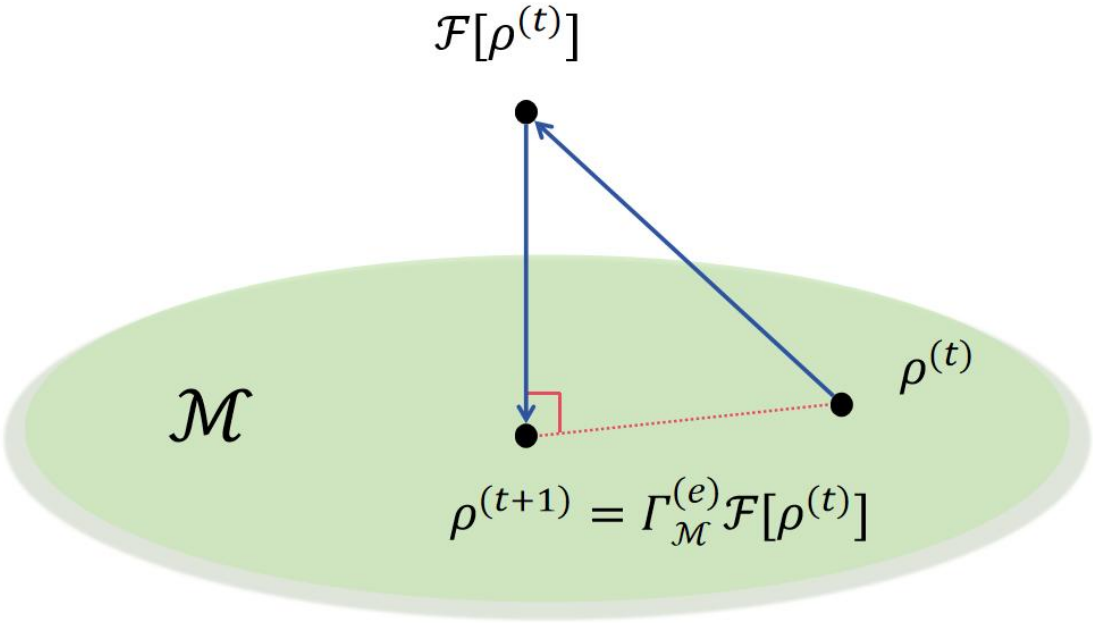
2 e-projection

$$\Gamma^{(e)}_{\mathcal{M}}[\rho] = \arg\min_{\sigma\in\mathcal{M}} D(\sigma\|\rho)$$

$$\rho^{(t+1)} = \arg\min_{\sigma\in\mathcal{M}} \{\mathrm{Tr}[\sigma\Omega(\rho^{(t)})] + \gamma D(\sigma\|\rho^{(t)})\}$$

e-projection preserves feasibility

The generalized quantum AB update



Algorithm 1. Minimization of $\mathcal{G}(\rho)$.

Choose the initial value $\rho^{(1)} \in \mathcal{M}$;
repeat
  Calculate $\rho^{(t+1)} := \Gamma^{(e)}_{\mathcal{M}_a}[\mathcal{F}_3[\rho^{(t)}]]$;
until convergence.

$$\rho^{(t+1)} = \arg\min_{\sigma\in\mathcal{M}} \{\mathrm{Tr}[\sigma\Omega(\rho^{(t)})] + \gamma D(\sigma\|\rho^{(t)})\}$$

e-projection preserves feasibility

Theoretical guarantees for AB algorithm

Condition 1 (Monotonicity) : For some $\gamma > 0$, and all pairs (ρ, σ), we have

$$D_\Omega(\rho\|\sigma) := Tr\rho(\Omega[\rho] - \Omega[\sigma]) \leq \gamma D(\rho\|\sigma)$$

Condition 2 (Global optimality) : For all pairs (ρ, σ), we have

$$D_\Omega(\rho\|\sigma) \geq 0$$

$$f(\rho^{(t+1)}) - f(\rho_*) \leq \frac{\gamma D(\rho_*\|\rho^{(1)})}{t}$$

Proving or verifying these conditions is very hard!
(problem dependent)

- Ramakrishnan, Navneeth, et al. "Computing quantum channel capacities." IEEE Transactions on Information Theory 67.2 (2020): 946-960.
- He, Kerry, James Saunderson, and Hamza Fawzi. "A Bregman proximal perspective on classical and quantum Blahut-Arimoto algorithms." IEEE Transactions on Information Theory 70.8 (2024): 5710-5730.

AB and entropic mirror descent

Motivation:

- **Previous analyses require sufficient, but not necessary conditions for global optimality.**
- **Equivalence between AB and Mirror Descent algorithm.**

MD updating map:

$$\mathcal{T}_{MD}(\rho) = \arg\min_{\sigma\in\mathcal{M}} \{\mathrm{Tr}[\sigma\nabla f(\rho)] + \gamma D(\sigma\|\rho)\}$$

AB updating map:

$$\mathcal{T}_{AB}(\rho) = \arg\min_{\sigma\in\mathcal{M}} \{\mathrm{Tr}[\sigma\Omega(\rho)] + \gamma D(\sigma\|\rho)\}$$

Condition1 & 2 → $\Omega(\rho) = \nabla f(\rho)$ → $\mathcal{T}_{AB}(\rho) = \mathcal{T}_{MD}(\rho)$

Feasible directions and the normal space

Tangent space

$$T_{\mathcal{M}} = \{X = X^{\dagger} : \mathrm{Tr}X = 0,\ \mathrm{Tr}(XH_j) = 0\ \forall j\}$$

Normal space

$$N_{\mathcal{M}} = T_{\mathcal{M}}^{\perp} = \mathrm{span}\{I, H_1, \ldots, H_k\}$$

$$N = a_0 I + \sum_j a_j H_j \quad \Longrightarrow \quad \mathrm{Tr}(\sigma N) = a_0 + \sum_j a_j c_j$$

Normal components add a constant to the feasible linear objective

One-step and pathwise equivalence

At a full-rank input state

$$\mathcal{T}_{AB}(\rho) = \mathcal{T}_{MD}(\rho) \quad \Leftrightarrow \quad \Omega(\rho) - \nabla f(\rho) \in N_{\mathcal{M}}$$

Along the common trajectory

$$\Omega(\rho^{(t)}) - \nabla f(\rho^{(t)}) \in N_{\mathcal{M}}, \qquad t = 0, \ldots, T-1$$

Pathwise equivalence requires compatibility at every common iterate

Accumulation points and AB fixed points

Accumulation set of the AB trajectory

$$\omega(\rho^{(0)}) := \{\, \overline{\rho} \in \mathcal{M} : \rho^{(t_k)} \to \overline{\rho} \ \text{ for some subsequence } \ t_k \to \infty \}$$

Trajectory assumption

$$D_\Omega(\rho^{(t+1)} \| \rho^{(t)}) \leq \gamma D(\rho^{(t+1)} \| \rho^{(t)}) \,, \qquad D_\Omega(\sigma \| \rho) := \mathrm{Tr}\sigma[\Omega(\sigma) - \Omega(\rho)]$$

Consequences

$$\|\rho^{(t+1)} - \rho^{(t)}\|_1 \longrightarrow 0 \qquad \mathrm{dist}(\rho^{(t)}, \omega(\rho^{(0)})) \ \to \ 0$$

$$\begin{cases} \rho^{(t_k)} \longrightarrow \overline{\rho} \in \mathcal{M}_{++} \\ \text{Continuity of the } \mathcal{T}_{AB} \end{cases} \quad \Longrightarrow \quad \mathcal{T}_{AB}(\overline{\rho}) = \overline{\rho}$$

Global optimality at a full-rank AB fixed point

Assume a convex differentiable objective and an exact full-rank AB fixed point $\rho^\star$.

AB stationarity

$$\Omega(\rho^\star) \in N_{\mathcal{M}}$$

Objective optimality

$$\nabla f(\rho^\star) \in N_{\mathcal{M}}$$

Necessary and sufficient condition

$$\rho^\star \in \operatorname{argmin}_{\rho \in \mathcal{M}} f(\rho) \quad \Leftrightarrow \quad \Omega(\rho^\star) - \nabla f(\rho^\star) \in N_{\mathcal{M}}$$

Compatibility at the endpoint suffices

Global optimality at a full-rank AB fixed point

Condition for global optimality $\Omega(\rho^\star) - \nabla f(\rho^\star) \in N_{\mathcal{M}}$

How to evaluate this condition?

Let P_T be the orthogonal projection onto $T_{\mathcal{M}}$, we consider the distance between an operator and the normal space $N_{\mathcal{M}}$.

$$\mathrm{dist}_2(Y, N_{\mathcal{M}}) = \|P_T Y\|_2 = \sup_{X \in T_{\mathcal{M}}, \|X\|_2 \leq 1} |\,\mathrm{Tr}(XY)|$$

$$\mathrm{dist}_2(\Omega(\rho) - \nabla f(\rho), N_{\mathcal{M}}) = \sup_{X \in T_{\mathcal{M}}, \|X\|_2 \leq 1} |\,\mathrm{Tr}(\rho D\Omega(\rho)[X])|$$

Let $E_1, \ldots, E_m$ be the orthonormal basis of $T_{\mathcal{M}}$, $m = d^2 - k - 1$, then

$$\mathrm{dist}_2^2(\Omega(\rho) - \nabla f(\rho), N_{\mathcal{M}}) = \sum_{a=1}^{m} [\,\mathrm{Tr}(\rho D\Omega(\rho)[E_a])]^2$$

Global optimality at a full-rank AB fixed point

Condition for global optimality $\Omega(\rho^\star) - \nabla f(\rho^\star) \in N_{\mathcal{M}}$

How to evaluate this condition?

The following conditions are equivalent

- $\rho^\star$ is a global minimizer of f over $\mathcal{M}$
- $\Omega(\rho^\star) - \nabla f(\rho^\star) \in N_{\mathcal{M}}$
- $Df(\rho^\star)[X] = 0 \quad \forall X \in T_{\mathcal{M}}$
- $\mathrm{Tr}(\rho^\star D\Omega(\rho^\star)[X]) = 0 \quad \forall X \in T_{\mathcal{M}}$
- $\mathrm{Tr}(\rho^\star D\Omega(\rho^\star)[E_a]) = 0$ for $a = 1, \ldots, m$

An objective-gap bound from the tangent residual

Let f be convex and differentiable, let $f_* := \min_{\sigma \in \mathcal{M}} f(\sigma)$

Tangent-gradient residual

$$\varepsilon(\rho) = \|P_T \nabla f(\rho)\|_2 = \mathrm{dist}_2(\nabla f(\rho), N_{\mathcal{M}})$$

$$R_{\mathcal{M}}(\rho) = \sup_{\sigma \in \mathcal{M}} \|\rho - \sigma\|_2 \leq \sqrt{2}$$

Global objective-gap guarantee

$$0 \leq f(\rho) - f_\star \leq R_{\mathcal{M}}(\rho)\varepsilon(\rho) \leq \sqrt{2}\varepsilon(\rho)$$

A posteriori optimality certificates

Choose $h_j > 0$ so that $\rho \pm h_j E_j \in \mathcal{M}$

$$\ell_j = \frac{f(\rho) - f(\rho - h_j E_j)}{h_j} \qquad u_j = \frac{f(\rho + h_j E_j) - f(\rho)}{h_j}$$

$$\ell_j \le Df(\rho)[E_j] \le u_j, \qquad \beta_j = max\{|\ell_j|, |u_j|\}$$

$$\Rightarrow \quad \epsilon(\rho) \le \left(\sum_{j=1}^{m} \beta_j^2\right)^{\frac{1}{2}} \qquad 0 \le f(\rho) - f^\star \le \sqrt{2}\left(\sum_{j=1}^{m} \beta_j^2\right)^{1/2}$$

If one derivative interval excludes zero, the interior candidate is nonoptimal

Application: quantum relative entropy of channels

QRE of channels: quantify distinguishability between quantum channels $\mathcal{N}_1$ and $\mathcal{N}_2$.

$$D(\mathcal{N}_1\|\mathcal{N}_2) = \sup_{\rho_{AR}} D((\mathcal{N}_1 \otimes id_R)(\rho_{AR})\|(\mathcal{N}_2 \otimes id_R)(\rho_{AR}))$$

$$= \sup_{\rho_A} D(\sqrt{\rho_A}\Gamma^{\mathcal{N}_1}\sqrt{\rho_A}\|\sqrt{\rho_A}\Gamma^{\mathcal{N}_2}\sqrt{\rho_A}) \qquad \text{(Concave in } \rho_A\text{)}$$

Convex minimization over the input marginal

$$f(\rho_A) = -D(\sqrt{\rho_A}\Gamma^{\mathcal{N}_1}\sqrt{\rho_A}\|\sqrt{\rho_A}\Gamma^{\mathcal{N}_2}\sqrt{\rho_A})$$

$$f(\rho_A) = \mathrm{Tr}[\rho_A\Omega(\rho_A)]$$

Application: quantum relative entropy of channels

Define $\Omega[\rho]$ so that the convex objective has the form $f(\rho) = \mathrm{Tr}\,\rho\Omega[\rho]$.

We define: $\Omega_1[\rho_A] = -\mathrm{Tr}_B(\Gamma^{\mathcal{N}_1}\rho_A^{0.5}(log\,\rho_A^{0.5}\Gamma^{\mathcal{N}_1}\rho_A^{0.5} - log\,\rho_A^{0.5}\Gamma^{\mathcal{N}_2}\,\rho_A^{0.5})\,\rho_A^{-0.5})$

$$\Omega[\rho] = \frac{\Omega_1[\rho_A] + \Omega_1[\rho_A]^\dagger}{2}$$

$$\Rightarrow \quad -D(\mathcal{N}_1||\mathcal{N}_2) = \inf_{\rho_A} -D(\sqrt{\rho_A}\Gamma^{\mathcal{N}_1}\sqrt{\rho_A}||\sqrt{\rho_A}\Gamma^{\mathcal{N}_2}\sqrt{\rho_A}) = \inf_{\rho_A} \mathrm{Tr}\rho_A\Omega[\rho_A]$$

No linear constraint: update is a deterministic fixed-point iteration on ρ_A.

$$\rho^{(t+1)} = \frac{\exp(\log\rho^{(t)} - \gamma^{-1}\Omega[\rho^{(t)}])}{\mathrm{Tr}\exp(\log\rho^{(t)} - \gamma^{-1}\Omega[\rho^{(t)}])}$$

With linear constraints: add the e-projection step.

Dephasing and depolarizing channels: exact solutions

Qubit channels: dephasing (q = 0.4) and depolarizing (p = 0.05)

$$\mathcal{D}_{deph}(\rho) = q\rho + (1-q)\sigma_z\rho\sigma_z, \qquad \mathcal{D}_p(\rho) = (1-p)\rho + pI/2$$

No extra constraint

$$\rho^\star = I/2$$

One Pauli expectation

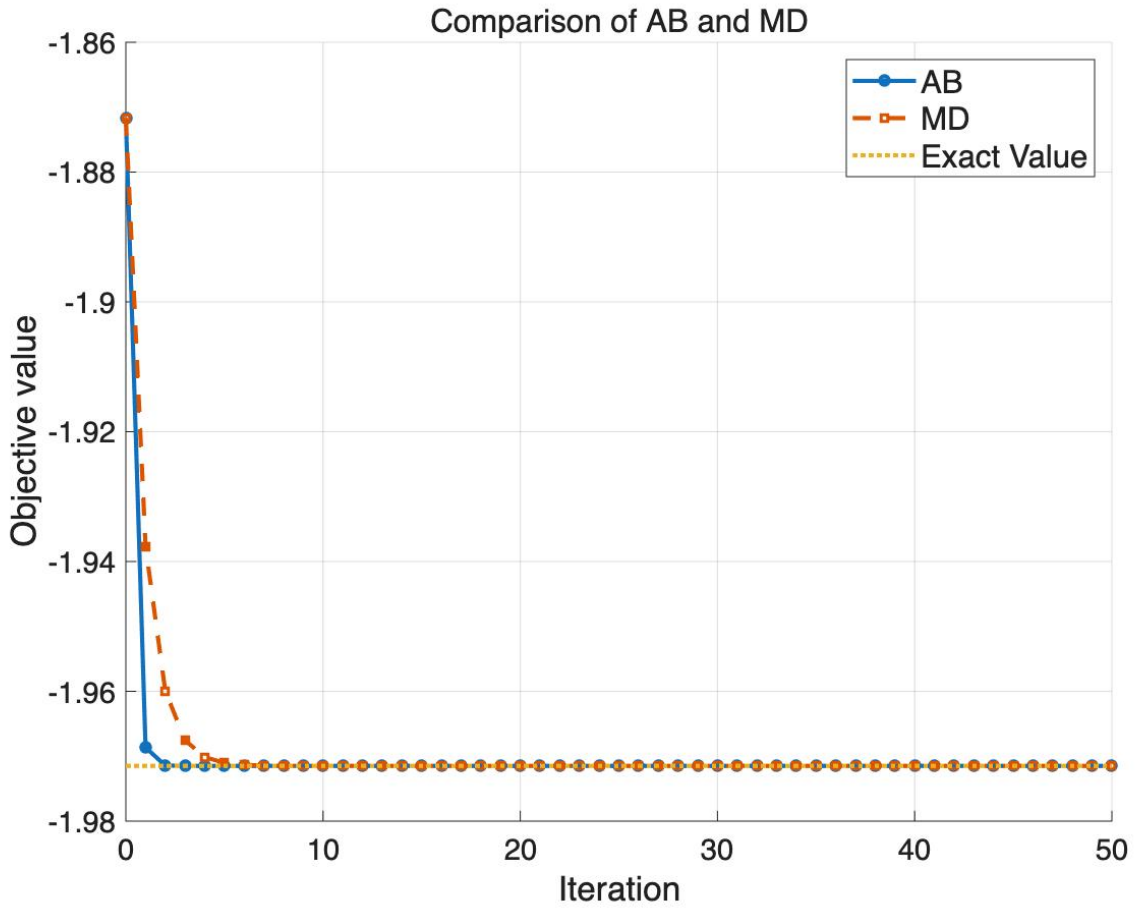
$$\mathrm{Tr}(\rho H) = E, \quad \rho^\star = (I + EH)/2$$

$$H \in \{\sigma_X, \sigma_Y, \sigma_Z\}, \ |E| < 1$$

In both cases, the optimizer is the unique full-rank AB fixed point.

AB and MD can have different paths while sharing the optimal endpoint

Different paths, the same optimum



Unconstrained qubit input

Starting state

$$\rho^{(0)} = \mathrm{diag}(3/8, 5/8)$$

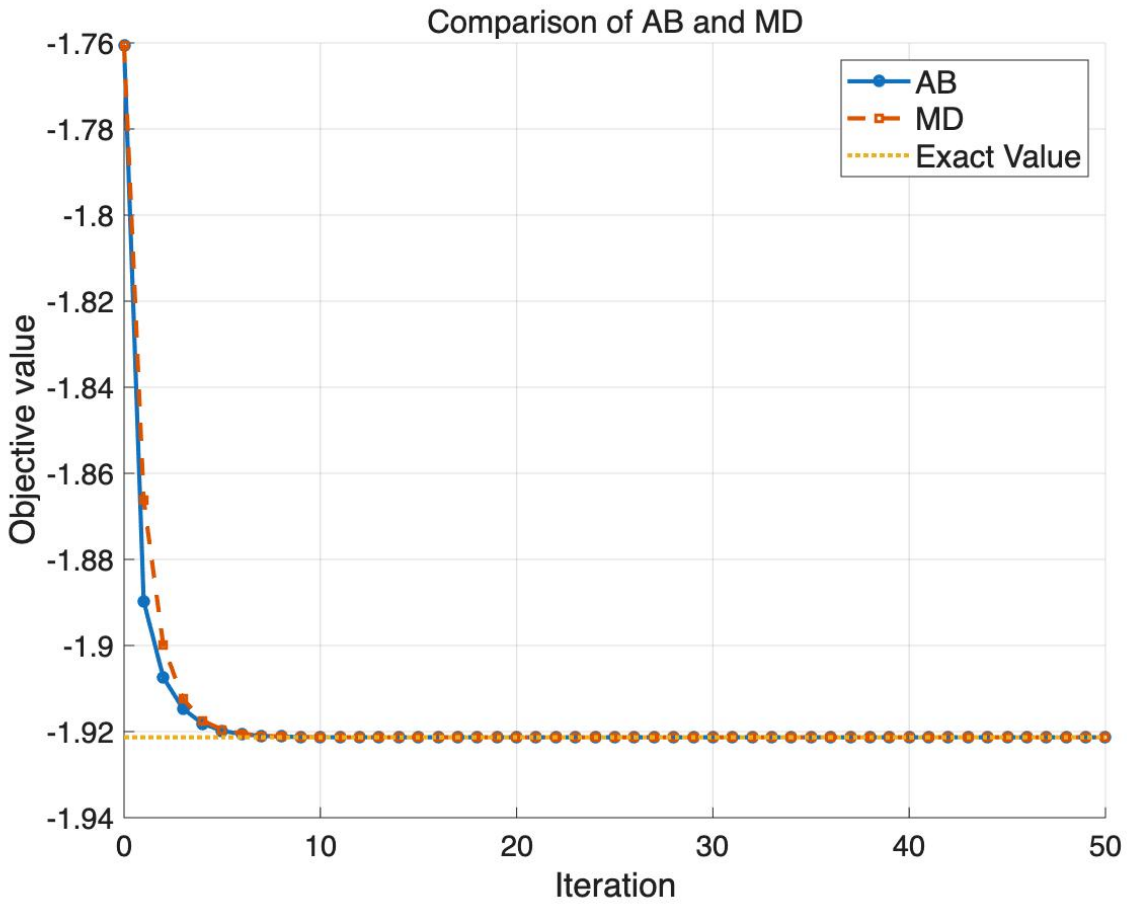
Step parameter and optimizer

$$\gamma = 1, \qquad \rho^{\star} = I/2$$

Initial tangent mismatch

$$\|P_T(\Omega - \nabla f)\|_2 \simeq 0.2674$$

The same pattern under a Pauli constraint



Feasible inputs

$$\mathrm{Tr}(\rho\sigma_x) = 0.4$$

Exact optimizer

$$\rho^\star = (I + 0.4\sigma_x)/2$$

Initial tangent mismatch

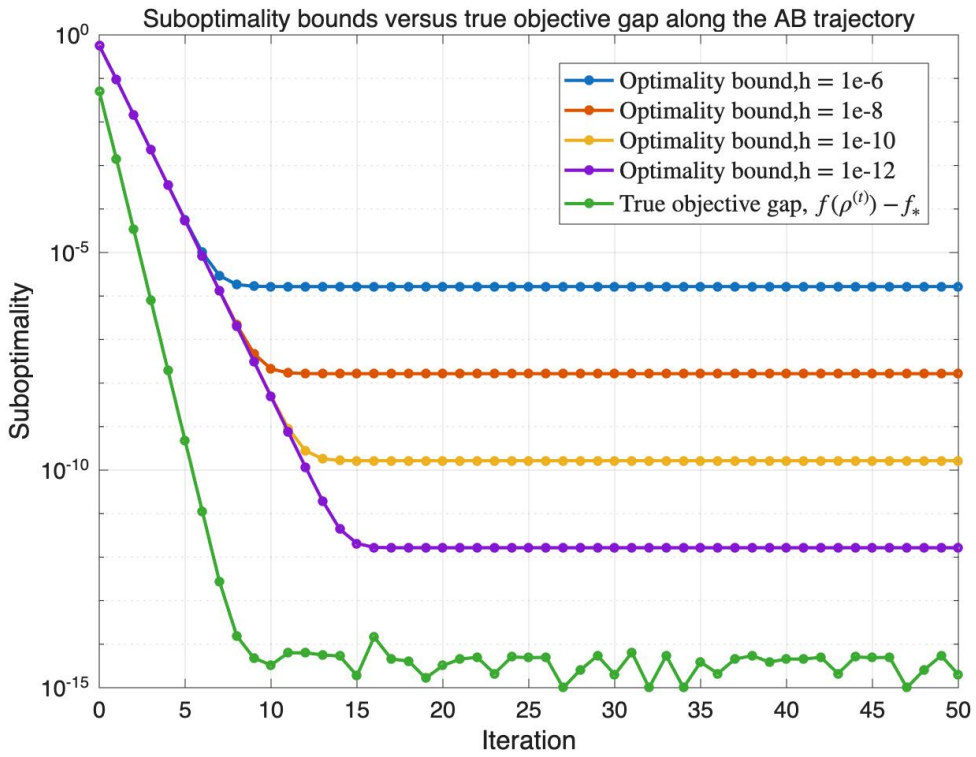
$$\|P_T(\Omega - \nabla f)\|_2 \simeq 0.2714$$

The constraint changes the optimizer. The two updates still differ.

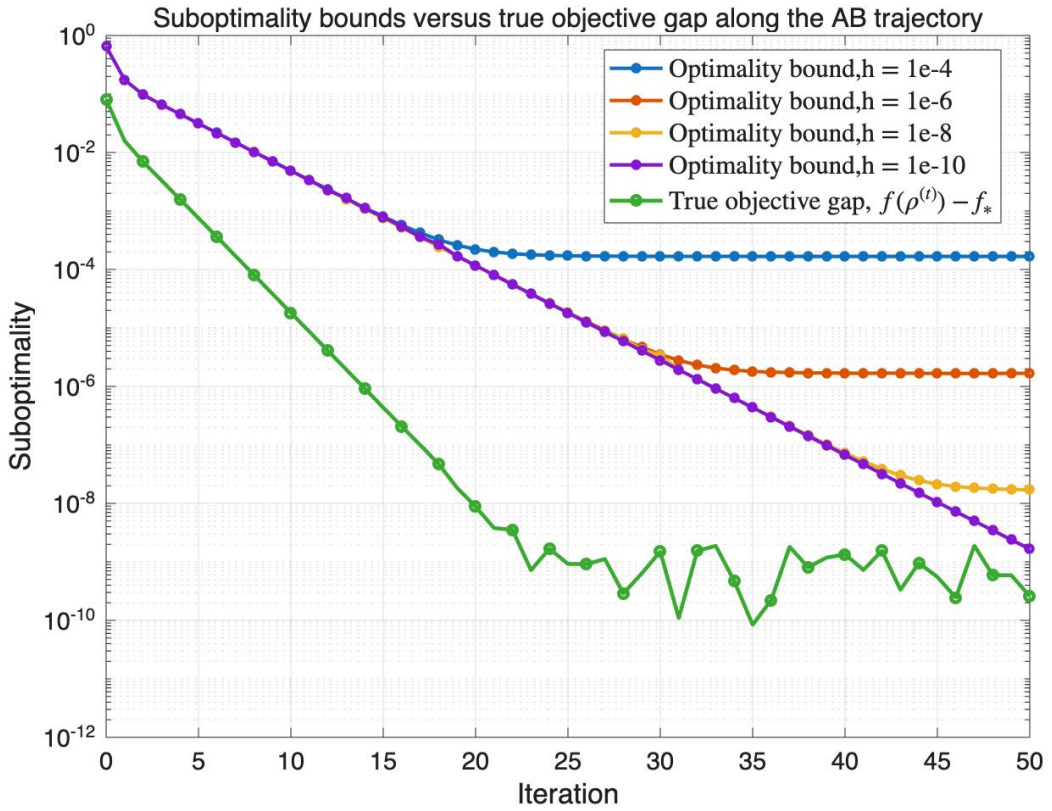
Finite-difference bounds along the AB trajectories

The known optimum supplies a benchmark for these experiments.

Unconstrained



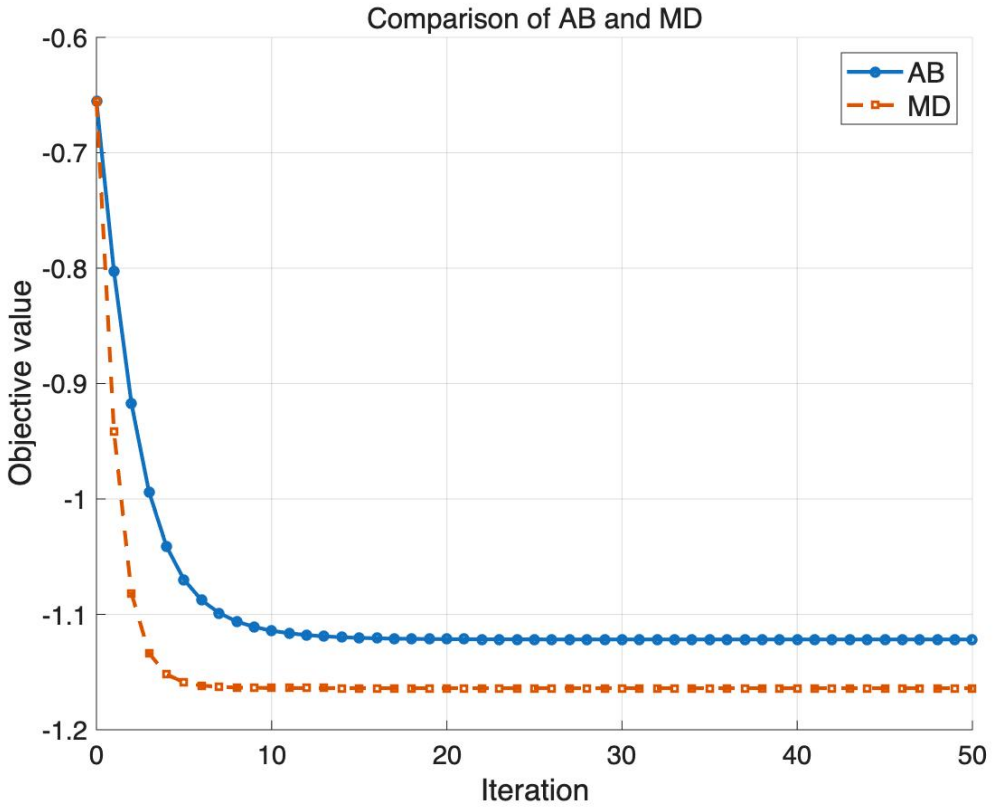
Fixed X expectation: 0.4



The bound uses only nearby objective values. Green shows the true gap.

A stable AB endpoint can be nonoptimal

Amplitude damping η = 0.5; depolarizing p = 0.05



After 50 AB updates

$$\|\mathcal{T}_{AB}(\rho^{(50)}) - \rho^{(50)}\|_1 \simeq 7.8 \times 10^{-9}$$

Feasible derivative interval

$$\ell_z \leq Df(\rho^{(50)})[E_z] \leq u_z$$

$$0.6127 < \ell_z \leq u_z < 0.6129$$

The interval excludes zero

Summary

Constraint geometry

AB–MD agreement depends on their feasible tangent components.

Endpoint optimality

A full-rank AB fixed point is optimal exactly when compatibility holds there.

Computable accuracy

Feasible finite differences bound the gap and detect nonstationary candidates.