QUANTUM MOMENT MATCHING

RESEARCH NOTES IN THE ENEXA AND QROM PROJECTS

February 11, 2026

ABSTRACT

We present Quantum Moment Matching, a parameter-estimation oriented preparation algorithm of maximum-entropy distributions. More specifically we show that the amplitude amplification procedure corresponds with a classical increase of the canonical parameter in a Hybrid Logic Network.

Contents

1 Introduction **1**

2 Moment Amplification **1**

2.1 Effective reflection by single computation circuits 2

2.1.1 Decomposition of the q-sample 2

2.1.2 Decomposition into auxiliary statistic qubits 3

2.2 Rotation along a formula 3

2.2.1 Maximum Entropy 3

3 Moment Estimation **4**

3.1 Eigenvalues the formula-amplifying Grover operator 5

3.2 Quantum Phase Estimation 5

4 Moment Matching **5**

1 Introduction

Tensor network contractions have been identified as central inference mechanism of neuro-symbolic AI in **?**. However, their exact execution is NP-hard, and one either has to restrict itself to feasible situations, or apply approximations.

We in this work investigate, how specific contractions (especially forward inference on Computation-Activation Networks) can be performed more efficiently on a Quantum Computer.

2 Moment Amplification

We here apply computation circuits on anti-symmetric ancilla states (as for sign encoding)

- A single computation circuit effectively prepares a reflection across the models of the computed formula

- Building the reflections based on ancilla ground states (as for basis encoding), we would need an additional Pauli-Z and two computation circuits
- Further ancilla qubits used in circuit decompositions need to be uncomputed (i.e. by applying the same partial computation circuit again) for further usage

2.1 Effective reflection by single computation circuits

Note, that the computation circuit is not a reflection, but acts as one when preparing the ancilla in the real anti-symmetric state.

To start let

- $\psi^0\left[X_{[d]}\right]$ be a q-sample of a probability distribution $\mathbb{P}^0\left[X_{[d]}\right]$
- U be a unitary preparing the state
- $\mathcal{C}^f$ be a computation circuit to a formula f

2.1.1 Decomposition of the q-sample

The we split the initial state into vectors supported at the models of f and the complement as

$$\psi^0\left[X_{[d]}\right] = \psi^{\parallel}\left[X_{[d]}\right] + \psi^{\perp}\left[X_{[d]}\right]$$

where

$$\begin{aligned}
\psi^{\parallel}\left[X_{[d]}\right] &= \sum_{x_{[d]}\in\times_{k\in[d]}[m_k]\,:\,f\left[X_{[d]}=x_{[d]}\right]=1} \sqrt{\mathbb{P}\left[X_{[d]}=x_{[d]}\right]}\cdot\epsilon_{x_{[d]}}\left[X_{[d]}\right]\\
\psi^{\perp}\left[X_{[d]}\right] &= \sum_{x_{[d]}\in\times_{k\in[d]}[m_k]\,:\,f\left[X_{[d]}=x_{[d]}\right]=0} \sqrt{\mathbb{P}\left[X_{[d]}=x_{[d]}\right]}\cdot\epsilon_{x_{[d]}}\left[X_{[d]}\right]\,.
\end{aligned}$$

We further define for a boolean formula

$$\psi^{\mathbb{P}^0,f}\left[X_{[d]}\right] = \frac{\psi^{\parallel}\left[X_{[d]}\right]}{\left\|\psi^{\parallel}\left[X_{[d]}\right]\right\|_2}$$

and have consistently for its negation

$$\psi^{\mathbb{P}^0,\neg f}\left[X_{[d]}\right] = \frac{\psi^{\perp}\left[X_{[d]}\right]}{\left\|\psi^{\perp}\left[X_{[d]}\right]\right\|_2}$$

We define an angle α by

$$\sin\left(\frac{\alpha}{2}\right) := \left\|\psi^{\parallel}\left[X_{[d]}\right]\right\|_2$$

and have

$$\psi^0\left[X_{[d]}\right] = \sin\left(\frac{\alpha}{2}\right)\psi^{\mathbb{P}^0,f}\left[X_{[d]}\right] + \cos\left(\frac{\alpha}{2}\right)\psi^{\mathbb{P}^0,f}\left[X_{[d]}\right]\,.$$

Further we have that

$$\begin{aligned}
\mu^0 &= \left\langle \mathbb{P}^0\left[X_{[d]}\right], f\left[X_{[d]}\right]\right\rangle_{[\varnothing]}\\
&= \sum_{x_{[d]}\in\times_{k\in[d]}[m_k]} \left|\psi^0\left[X_{[d]}=x_{[d]}\right] f\left[X_{[d]}=x_{[d]}\right]\right|^2\\
&= \left(\|\psi^{\parallel}\left[X_{[d]}\right]\|_2\right)^2\\
&= \left(\sin\left(\frac{\alpha}{2}\right)\right)^2\,.
\end{aligned}$$

We prepare an ancilla qubit in the anti-symmetric state and consider as initial state

$$\psi^0\left[X_{[d]}\right] \otimes \sqrt{\frac{1}{2}} \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix} [A] \,.$$

The computation circuit $\mathcal{C}^f$ acting on this state prepares a state

$$\left(\psi^{\perp}\left[X_{[d]}\right] - \psi^{\parallel}\left[X_{[d]}\right]\right) \otimes \sqrt{\frac{1}{2}} \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix} [A] \,.$$

This is the reflection on the subspace spanned by the one-hot encodings of the models of f.

2.1.2 Decomposition into auxiliary statistic qubits

We can exploit decomposition sparsity to find a sparse representation of a computation circuit with further auxiliary statistic qubits. Any auxiliary statistic qubits Y is prepared in the initial state $\epsilon_0[Y]$, only the head statistic qubit is prepared in the anti-symmetric state. To prepare a sign encoding such that the auxiliary statistic qubits are disentangled, one has to uncompute all auxiliary statistic qubits, by applying the respective computation circuits with auxiliary target qubits again.

To be more precise, let $Y_{[p]}$ be auxiliary statistic qubits representing the subformulas t in a syntactic decomposition of f, applying a decomposed computation circuit on the initial state then gives

$$\left(\sum_{x_{[d]} \in \times_{k\in[d]}[m_k]} (-1)^{f(x_{[d]})} \psi^0\left[X_{[d]} = x_{[d]}\right] \cdot \epsilon_{x_{[d]}}\left[X_{[d]}\right] \otimes \epsilon_{t(x_{[d]})}\left[Y_{[p]}\right]\right)$$

Notice, that in this case the auxiliary qubits are entangled with the distributed qubits. This can be resolved by applying the adjoint of the computation circuit except for those where the control is on the ancilla qubit. We then have

$$\begin{aligned}&\left(\sum_{x_{[d]} \in \times_{k\in[d]}[m_k]} (-1)^{f(x_{[d]})} \psi^0\left[X_{[d]} = x_{[d]}\right] \cdot \epsilon_{x_{[d]}}\left[X_{[d]}\right]\right) \otimes \epsilon_{t(x_{[d]})\oplus t(x_{[d]})}\left[Y_{[p]}\right] \\ &= \left(\sum_{x_{[d]} \in \times_{k\in[d]}[m_k]} (-1)^{f(x_{[d]})} \psi^0\left[X_{[d]} = x_{[d]}\right] \cdot \epsilon_{x_{[d]}}\left[X_{[d]}\right]\right) \otimes \epsilon_{0_{[p]}}\left[Y_{[p]}\right]\end{aligned}$$

Alternatively, one can include the auxiliary statistic qubits into the distributed qubits, and understand the parts of the computation circuit not affecting the ancilla qubit as part of the initial state preparing unitary U. Note that this perspective has been applied in section Sect. **??**, where we took U as the Computation Activation Circuit preparing the q-sample of the ancilla augmentation.

2.2 Rotation along a formula

Rotating j times across the initial state then gives the tensor product of the antisymmetric ancilla state with

$$\psi^j\left[X_{[d]}\right] = \sin\left(\left(\frac{1}{2}+j\right)\alpha\right)\psi^{\mathbb{P}^0,f}\left[X_{[d]}\right] + \cos\left(\left(\frac{1}{2}+j\right)\alpha\right)\psi^{\mathbb{P}^0,\neg f}\left[X_{[d]}\right] \,.$$

This reflection can be performed by the operator on the distributed qubits

$$U^t\left(\delta\left[X^{\text{in}}_{[d]}, X^{\text{out}}_{[d]}\right] - 2\epsilon_0\left[X^{\text{in}}_{[d]}\right] \otimes \epsilon_0\left[X^{\text{out}}_{[d]}\right]\right)U \,.$$

2.2.1 Maximum Entropy

When $\left(\frac{1}{2}+j\right)\alpha \leq \frac{\pi}{2}$ then all amplitudes remain positive, and $\psi^j\left[X_{[d]}\right]$ is a q-sample. What is more, we can show that the property of maximum entropy with respect to a statistic containing the amplified formula.

Theorem 1. *Let ψ^0 be the q-sample of a maximum entropy distribution with respect to the uniform base measure and the statistic $\mathcal{F}$ containing the amplified formula, and let $j \in \mathbb{N}$ be a rotation number such that $\left(\frac{1}{2}+j\right)\alpha \leq \frac{\pi}{2}$. Then also ψ^j is the q-sample of a maximum entropy distribution.*

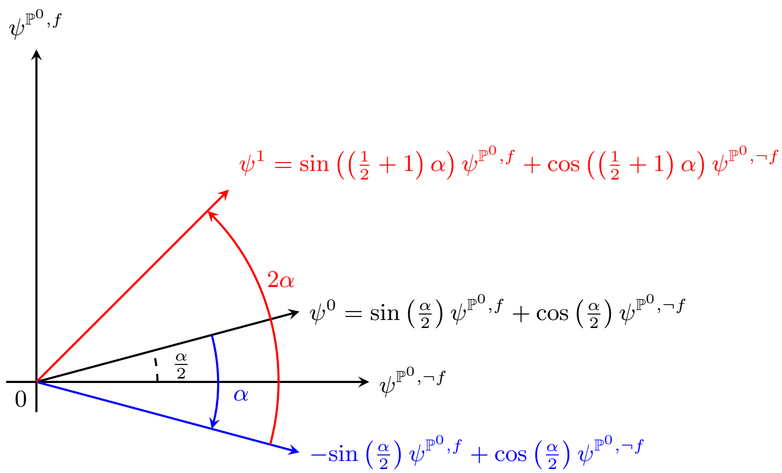

Figure 1: Geometric visualization of the states ψ^1 resulting from a Grover rotation on ψ^0, in the subspace spanned by $\psi^{\mathbb{P}^0,f}$ and $\psi^{\mathbb{P}^0,\neg f}$. The effective reflection by the computation circuit results in the gray state $\psi^\perp - \psi^\parallel$. Reflecting again on the initial state ψ^0 results in the amplified state ψ^1. Each such Grover iteration is therefore a rotation by the angle α in the.

We show Thm. 1 later based on a more technical description of the rotated state, which we show as the next lemma.

Lemma 1. *Let ψ^0 be the q-sample of the elementary Computation-Activation Network with canonical parameters (A, y_A, θ), and let us amplify the formula f_ℓ. Then ψ^j is the q-sample of the elementary Computation-Activation Network with canonical parameters*

$$A^j = \begin{cases} A \cup \{\ell\}, y_A^\ell = (y_A, 1) & \textit{if} \quad \sin\left((1+2\cdot j)\sin^{-1}\left(\sqrt{\mu_0\left[L=\ell\right]}\right)\right) = 1 \\ A, y_A & \textit{else} \end{cases}$$

and

$$\theta_j\left[L=\ell\right] = \begin{cases} 0 & \textit{if} \quad \sin\left((1+2\cdot j)\sin^{-1}\left(\sqrt{\mu_0\left[L=\ell\right]}\right)\right) = 1 \\ \theta\left[L=\ell\right] + \frac{\cos^2\left(\frac{\alpha}{2}\right)\left(1-\cos^2\left(\left(\frac{1}{2}+j\right)\alpha\right)\right)}{\sin^2\left(\frac{\alpha}{2}\right)\cdot\cos^2\left(\left(\frac{1}{2}+j\right)\alpha\right)} & \textit{else} \end{cases}.$$

Proof. Notice, that

$$\left\langle \mathbb{P}^0\left[X_{[d]}\right], \beta^{f_\ell}\left[Y_\ell, X_{[d]}\right]\right\rangle_{[Y_\ell]} = \begin{bmatrix} \left(\cos\left(\frac{\alpha}{2}\right)\right)^2 \\ \left(\sin\left(\frac{\alpha}{2}\right)\right)^2 \end{bmatrix}\left[Y_\ell\right] = \begin{bmatrix} 1-\mu_0\left[L=\ell\right] \\ \mu_0\left[L=\ell\right] \end{bmatrix}\left[Y_\ell\right]$$

and

$$\left\langle \mathbb{P}^j\left[X_{[d]}\right], \beta^{f_\ell}\left[Y_\ell, X_{[d]}\right]\right\rangle_{[Y_\ell]} = \begin{bmatrix} \left(\cos\left(\left(\frac{1}{2}+j\right)\alpha\right)\right)^2 \\ \left(\sin\left(\left(\frac{1}{2}+j\right)\alpha\right)\right)^2 \end{bmatrix}\left[Y_\ell\right]\begin{bmatrix} 1-\mu_j\left[L=\ell\right] \\ \mu_j\left[L=\ell\right] \end{bmatrix}\left[Y_\ell\right]$$

If $\mu_j\left[L=\ell\right] = 1$ (that is $\sin\left(\left(\frac{1}{2}+j\right)\alpha\right) = 1$), the amplitude amplification amounts to adding f_ℓ as a hard constraint. This is equal to $A^j = A \cup \{\ell\}$ and $y_A^j = (y_A, 1)$.

If $\mu_j\left[L=\ell\right] \neq 1$ we choose $\theta_\Delta\left[L=\ell\right] \geq 0$ as in the claim and have after some algebra

$$\frac{\left(\cos\left(\frac{\alpha}{2}\right)\right)^2}{\left(\cos\left(\frac{\alpha}{2}\right)\right)^2 + \theta_\Delta\left[L=\ell\right]\cdot\left(\sin\left(\frac{\alpha}{2}\right)\right)^2} = \left(\cos\left(\left(\frac{1}{2}+j\right)\alpha\right)\right)^2$$

and therefore

$$\left\langle \mathbb{P}^j\left[X_{[d]}\right], \beta^{f_\ell}\left[Y_\ell, X_{[d]}\right]\right\rangle_{[Y_\ell]} = \left\langle \exp\left[\theta_\Delta\left[L=\ell\right]\cdot f_\ell\left[\theta\right]\right]\mathbb{P}^0\left[X_{[d]}\right], \beta^{f_\ell}\left[Y_\ell, X_{[d]}\right]\right\rangle_{[Y_\ell|\varnothing]}.$$

Thus, the rotation of the q-sample corresponds with an increase of the canonical parameter $\theta_0\left[L=\ell\right]$ by $\theta_\Delta\left[L=\ell\right]$. $\square$

Proof of Thm. 1. By Lem. 1 the amplitude amplified state is an elementary Computation-Activation Network and thus a maximum entropy distribution. ☐

Notice, that we can only increase the canonical parameter by amplitude amplification. Decreasing could be done by amplitude amplification on $\neg f_\ell$ instead.

3 Moment Estimation

Given a distribution $\mathbb{P}\left[X_{[d]}\right]$ and a propositional formula $f\left[X_{[d]}\right]$, we now investigate how to use Quantum Phase Estimation to estimate

$$\mu = \left\langle \mathbb{P}\left[X_{[d]}\right], f\left[X_{[d]}\right]\right\rangle_{[\varnothing]} .$$

We assume that

- a circuit U preparing a q-sample of $\mathbb{P}\left[X_{[d]}\right]$
- a computation circuit for $f\left[X_{[d]}\right]$

3.1 Eigenvalues the formula-amplifying Grover operator

The Grover operator $G^{\mathbb{P},f}\left[X_{[d]}^{\mathrm{in}}, X_{[d]}^{\mathrm{out}}, A^{\mathrm{in}}, A^{\mathrm{out}}\right]$ amplifying f

$$(U \circ \mathcal{S}^0 \circ U) \otimes \mathcal{C}^f$$

acting on the initial state

$$\epsilon_0\left[X_{[d]}\right] \otimes \sqrt{\frac{1}{2}} \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix}[A] .$$

We describe the Grover operator by its action on the projection onto the plane

$$\mathrm{span}\left(\psi^{\mathbb{P},f}, \psi^{\mathbb{P},\neg f}\right)$$

as

$$\begin{bmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{bmatrix}[O_{\mathrm{in}}, O_{\mathrm{out}}] .$$

Here we introduce a boolean variables O to select the two axis on the plane, and copy it to $O_{\mathrm{in}}, O_{\mathrm{out}}$. The rotation angle is related to μ as

$$\alpha = 2 \cdot \sin^{-1}\left(\sqrt{\mu}\right)$$

Its eigenvalues are thus

$$\exp\left[\pm i\alpha\right]$$

to the eigenstates

$$\frac{1}{2}\begin{bmatrix} 1 \\ \pm i \end{bmatrix}[O] .$$

3.2 Quantum Phase Estimation

We prepare auxiliary qubits $A_{[p]}$ and perform for each $\ell \in [p]$ 2^ℓ repetitions of $G^{\mathbb{P},f}$ controlled on A_ℓ. The inverse Fourier transform on the auxiliary qubits is supported on the (2-adic representation of the) phases on the eigenvalues, i.e. $\pm\alpha$, and can be detected by a computational basis measure of the ancilla variables.

4 Moment Matching

We now investigate which Computation-Activation Networks can be prepared directly with amplitude amplification. While we already have studied sparse representations of the computation network parts, we here apply amplitude amplification by Grover rotations as an activation mechanism.

We orient on the iterative proportional fitting algorithm, which is used to estimate the canonical parameters in exponential families **?** and more generally of Computation-Activation Networks **?**.

The algorithm loops over the moment estimation and moment amplification subroutines, which have been introduced in the sections above. The idea is to first compute the current moments of a distribution and second amplify one.

The mean parameter of f with respect to the measurement distribution is

$$\mu^j = \left\langle \mathbb{P}^j \left[X_{[d]}\right], f\left[X_{[d]}\right]\right\rangle_{[\varnothing]} = \sin\left(\left(\frac{1}{2}+j\right)\alpha\right)^2 = \sin\left((1+2j)\sin^{-1}\left(\sqrt{\mu^0}\right)\right)^2$$

Here we used that $\alpha = 2\cdot\sin^{-1}\left(\sqrt{\mu^0}\right)$.

We understand this rotation as a moment matching operation to the formula f, see Algorithm 1.

Algorithm 1 Quantum Circuit Moment Matcher

Require: Formulas $f_p\left[X_{[d]}\right]$ for $\ell \in [p]$, vector $\mu_D[L]$ of mean parameters
Ensure: Circuit U preparing a state (when applied on the ground state) which measurement distribution matched $\mu_D[L]$

$U \leftarrow (\sigma_1 \circ H)[A_{\text{in}}, A_{\text{out}}] \otimes \left(\bigotimes_{k\in[d]} H\left[X_{k,\text{in}}, X_{k,\text{out}}\right]\right)$
while Convergence criterion not met **do**
Estimate $\mu[L]$,

- either by particle bases inference
- or by quantum counting (measuring the eigenvalue of the Grover operator), see Sect. 3

Select an $\ell \in [p]$, which moment has to be updated (e.g. by comparison of $\mu[L]$ with $\mu_D[L]$)
Choose whether to amplify f_ℓ or $\neg f_\ell$ based on whether the estimated mean is smaller than the target mean
Compute optimal j to match $\mu_D[L]$
Extend Circuit U by j alternations of

- $\mathcal{C}^{f_\ell}$ (effectively an reflection across the subspace spanned by one-hot encoded models)
- reflections across the current state (by $U(I-2e_0e_0)U$)

end while

We can use Algorithm 1 as a preparation algorithm of a q-sample before doing ancilla augmentation as in the main section. Note that is important to keep track of the canonical parameters in the preparation. When activating the statistic qubits, only the difference of the already prepared canonical parameters and the target canonical parameters has to be used (importance sampling interpretation).