Supplementary Information for: Halving the Cost of Quantum Algorithms with Randomization

John M. Martyn[1,2] and Patrick Rall[2]

[1] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[2] *IBM Quantum, MIT-IBM Watson AI Lab, Cambridge, Massachusetts 02142, USA*

I. POLYNOMIAL APPROXIMATION TO SMOOTH FUNCTIONS

Here we recall theorems on polynomial approximations to smooth functions, specifically the decay of Fourier coefficients and Chebyshev coefficients of smooth functions. For more details on these results, see Refs. [1, 2].

A. Decay of Fourier Coefficients

Fourier analysis can be used to determine the rate of decay of the Fourier coefficients of smooth functions. The following is a classic result [1], for which we provide a brief proof:

Theorem 1. *Let $G(\theta)$ be a function with period 2π and Fourier series*

$$G(\theta) = \frac{a_0}{2} + \sum_{n=-\infty}^{\infty} a_n e^{in\theta}. \tag{1}$$

If $G(\theta)$ is a C^k function on $\theta \in [0, 2\pi)$ (i.e., continuous and differentiable through order k), then the Fourier coefficients decay polynomially as

$$a_n = o\Big(\frac{1}{n^k}\Big). \tag{2}$$

Similarly, if $G(\theta)$ is a C^∞ function on $\theta \in [0, 2\pi)$, then the Fourier coefficients decay super-polynomially as

$$a_n = O\big(e^{-qn^r}\big) \tag{3}$$

for some $q, r > 0$.

Proof. The Fourier coefficients are

$$a_n = \frac{1}{2\pi}\int_0^{2\pi} G(\theta) e^{-in\theta} d\theta. \tag{4}$$

Using integration by parts repeatedly, we can write this as

$$\begin{aligned} a_n &= \frac{1}{2\pi}\int_0^{2\pi} G(\theta) e^{-in\theta} d\theta \\ &= \frac{i}{n}\frac{1}{2\pi}\int_0^{2\pi} G'(\theta) e^{-in\theta} d\theta \\ &= \quad ... \\ &= \frac{i^k}{n^k}\frac{1}{2\pi}\int_0^{2\pi} G^{(k)}(\theta) e^{-in\theta} d\theta. \end{aligned} \tag{5}$$

This implies that the Fourier coefficients of the kth derivative $G^{(k)}(\theta)$ are $(-in)^k a_n$.

According to the Riemann-Lebesgue lemma [1], the Fourier coefficients of a smooth function go to 0 as $n \to \infty$. Therefore, if $G(\theta)$ is a C^k function, then $G^{(k)}(\theta)$ is continuous, and thus its Fourier coefficients decay as $\lim_{n\to\infty}(-in)^k a_n = 0$. This limit implies that $a_n = o(1/n^k)$.

On the other hand, if $G(\theta)$ is a C^∞ function, then by an analogous argument, $\lim_{n\to\infty} n^k a_n = 0$ for all integers $k \geq 1$. This implies that a_n decays super-polynomially, i.e. as $a_n = O\big(e^{-qn^r}\big)$ for some $q, r > 0$. □

B. Decay of Chebyshev Coefficients

An analogous result can be derived for the decay of Chebyshev coefficients. Recall that the nth Chebyshev polynomial $T_n(x)$ is a degree n polynomial defined on $x \in [-1, 1]$ as

$$T_n(x) = \cos(n \arccos(x)). \tag{6}$$

It can be shown that $T_n(x)$ is a degree n polynomial of definite parity (either even or odd, depending on n) and bounded magnitude $|T_n(x)|_{[-1,1]} = 1$ [2]. The Chebyshev polynomials provide a convenient basis for expanding functions on $x \in [-1, 1]$. A function $F(x)$ can be expanded as

$$F(x) = \frac{c_0}{2} + \sum_{n=1}^{\infty} c_n T_n(x), \tag{7}$$

where

$$c_n = \frac{2}{\pi}\int_{-1}^{1} \frac{F(x)T_n(x)}{\sqrt{1-x^2}} dx \tag{8}$$

are the Chebyshev coefficients for all $n \geq 0$.

By the relation between Chebyshev series and Fourier series, it can be shown that the Chebyshev coefficients decay in a manner analogous to the Fourier coefficients:

Theorem 2. *Let $F(x)$ be a function on $x \in [-1, 1]$ with the Chebyshev series*

$$F(x) = \frac{c_0}{2} + \sum_{n=1}^{\infty} c_n T_n(x), \tag{9}$$

If $F(x)$ is a C^k function on $x \in [-1, 1]$, then the Chebyshev coefficients decay polynomially as

$$c_n = o\Big(\frac{1}{n^k}\Big). \tag{10}$$

Similarly, if $F(x)$ is a C^∞ function on $x \in [-1,1]$, then the Chebyshev coefficients decay super-polynomially as

$$c_n = O\big(e^{-qn^r}\big) \tag{11}$$

for some $q, r > 0$.

Proof. Let $\theta = \arccos(x)$, or equivalently $x = \cos\theta$. Using this change of variables, we can re-express the Chebyshev coefficients as

$$\begin{aligned} c_n &= \frac{2}{\pi}\int_{-1}^{1} \frac{F(x)T_n(x)}{\sqrt{1-x^2}}dx = \frac{2}{\pi}\int_0^\pi F(\cos(\theta))\cos(n\theta)d\theta \\ &= \frac{1}{2\pi}\int_0^{2\pi} F(\cos(\theta))e^{-in\theta}d\theta \end{aligned} \tag{12}$$

where the last equality stems from $F(\cos(\theta))$ being an even function of θ. Comparing with Eq. (4), we see that the Chebyshev coefficients c_n are equal to the Fourier coefficients of the function $G(\theta) := F(\cos(\theta))$. Therefore, if we can show that $F(x)$ being a C^k function of x implies that $G(\theta) = F(\cos(\theta))$ is a C^k function of θ, then we can inherit the result of Theorem 1 to prove the purported decay pattern of the coefficients c_n.

This is easy to show. Observe that the first derivative of $G(\theta)$ is

$$G'(\theta) = F'(\cos\theta)(-\sin\theta) = -F'(x)\sin\theta. \tag{13}$$

If $F(x)$ is a C^1 function of $x = \cos\theta$, such that $F'(x)$ is a continuous function of x, then $G'(\theta) = -F'(\cos\theta)\sin(\theta)$ is a composition of continuous functions of θ. Therefore, $G'(\theta)$ is is also a continuous function of θ, which implies that $G(\theta)$ is a C^1 function of θ. A similar trend persists at higher orders: by using the chain rule and product rule, the kth derivative $G^{(k)}(\theta)$ can be expressed as a linear combination of products of derivatives $F^{(j)}(\cos\theta)$ for $j \le k$ and trigonometric polynomials $\sin^a(\theta)\cos^b(\theta)$ for some integers $a, b \ge 0$. Thus, if $F(x)$ is a C^k function of x, then this expression for $g^{(k)}(\theta)$ is also continuous, which implies that $G(\theta)$ is a C^k function.

Therefore, by appealing to Theorem 1, we see that if $F(x)$ is a C^k function, then the Chebyshev coefficients c_n decays polynomially as $c_n = o(1/n^k)$. Similarly, if $F(x)$ is a C^∞ function, then c_n decays super-polynomially as $c_n = O\big(e^{-qn^r}\big)$ for some $q, r > 0$. □

II. EXTENSIONS OF THE MIXING LEMMA

Here we present extensions of the mixing lemma to non-unitary evolution and block-encodings. The proofs of these variants parallel the proof of the mixing lemma presented in Ref. [3].

A. The Generalized Mixing Lemma

Consider a mixing lemma for arbitrary operators, including non-unitary evolution. In this scenario, we we wish to implement a channel $\mathcal{S}(\rho) = S\rho S^\dagger$, where S is not necessarily unitary, yet we only have access to operators R_j that approximate S. Then we can prove the following:

Lemma 3 (Generalized Mixing Lemma)**.** *Let S be a target operator, possibly non-unitary, and $\mathcal{S}(\rho) = S\rho S^\dagger$ the corresponding channel. Suppose there exist operators R_j and a probability distribution p_j that approximate S as*

$$\begin{aligned} &\|R_j - S\| \le a \text{ for all } j, \\ &\Big\|\sum_j p_j R_j - S\Big\| \le b, \end{aligned} \tag{14}$$

for some $a, b > 0$. Then, the corresponding channel $\Lambda(\rho) = \sum_j p_j R_j \rho R_j^\dagger$ approximates the channel $\mathcal{S}$ as

$$\|\Lambda - \mathcal{S}\|_\diamond \le a^2 + 2b\|S\|. \tag{15}$$

Proof. Paralleling the proof of the mixing lemma in Ref. [3], let $\delta_j = R_j - S$. This obeys $\|\delta_j\| \le a$ and $\big\|\sum_j p_j\delta_j\big\| \le b$ by the assumed conditions of the theorem. The action of the channel $\mathcal{S}$ can then be expanded as

$$\begin{aligned} \Lambda(\rho) &= \sum_j p_j R_j \rho R_j^\dagger \\ &= \sum_j p_j (S+\delta_j)\rho(S+\delta_j)^\dagger \\ &= S\rho S^\dagger + \Big(\sum_j p_j\delta_j\Big)\rho S^\dagger \\ &\quad + S\rho\Big(\sum_j p_j\delta_j^\dagger\Big) + \sum_j \delta_j\rho\delta_j^\dagger. \end{aligned} \tag{16}$$

We can then bound the 1-norm $\|\Lambda(\rho) - \mathcal{S}(\rho)\|_1$ via the triangle inequality as

$$\begin{aligned} \|\Lambda(\rho) - \mathcal{S}(\rho)\|_1 \le &\Big\|\Big(\sum_j p_j\delta_j\Big)\rho S^\dagger\Big\|_1 \\ &+ \Big\|S\rho\Big(\sum_j p_j\delta_j^\dagger\Big)\Big\|_1 + \Big\|\sum_j \delta_j\rho\delta_j^\dagger\Big\|_1. \end{aligned} \tag{17}$$

The first term on the right hand side of Eq. (17) can be bounded by appealing to Holder's inequality (specifically, $\|AB\|_1 \le \|A\|\|B\|_1$ and $\|AB\|_1 \le \|A\|_1\|B\|$ in this context):

$$\begin{aligned} \Big\|\Big(\sum_j p_j\delta_j\Big)\rho S^\dagger\Big\|_1 &\le \Big\|\sum_j p_j\delta_j\rho\Big\|_1\|S^\dagger\| \\ &\le \Big\|\sum_j p_j\delta_j\Big\|\|\rho\|_1\|S\| \\ &\le \|S\|b, \end{aligned} \tag{18}$$

where we have used that $\|\rho\|_1 = 1$. Analogously, the second term in Eq. (17) can be upper bounded by $\|S\|b$.

Again invoking Holder's inequality, the last term in Eq. (17) can be bounded as

$$\begin{aligned}\Big\|\sum_j \delta_j \rho \delta_j^\dagger\Big\|_1 &\le \sum_j p_j \|\delta_j \rho \delta_j^\dagger\|_1 \\ &\le \sum_j p_j \|\delta_j \rho\|_1 \|\delta_j^\dagger\| \\ &\le \sum_j p_j \|\delta_j\| \|\rho\|_1 \|\delta_j^\dagger\| \\ &\le \sum_j p_j \|\delta_j\| \|\delta_j\| \\ &\le a^2.\end{aligned} \tag{19}$$

Therefore we have

$$\big\|\Lambda(\rho) - \mathcal{S}(\rho)\big\|_1 \le a^2 + 2b\|S\|, \tag{20}$$

for all density matrices ρ. This translates to a diamond norm bound:

$$\begin{aligned}\big\|\Lambda - \mathcal{S}\big\|_\diamond &= \sup_{\|\sigma\|_1 \le 1} \|(\Lambda \otimes \mathcal{I})(\sigma) - (\mathcal{S} \otimes \mathcal{I})(\sigma)\|_1 \\ &\le a^2 + 2b\|S\|,\end{aligned} \tag{21}$$

because appending the identity channel $\mathcal{I}$ does not alter the proof of 1-norm bound in Eq. (20), which holds true for all density matrices. □

Notably, this result is analogous to the original mixing lemma, but modified by the spectral norm of S. For unitary evolution where $\|S\| = 1$, the generalized mixing lemma reduces to the usual mixing lemma.

B. The Mixing Lemma for Block-Encodings

We can consider an analog of the mixing lemma for block-encodings. In this scenario, we wish to evolve under an operator S (which may be non-unitary) encoded in a unitary V, yet we only have access to operators R_j block-encoded in unitaries U_j. We first study the case in which the operators are block-encoded in the $|0\rangle\langle 0|$ block of their respective unitaries, which is the simplest form of a block-encoding and most often considered in the literature. Afterwards, we proceed to the general case, in which the operators are encoded through arbitrary projectors Π and Π'.

First consider encodings in the $|0\rangle\langle 0|$ block:

Lemma (Mixing Lemma for Block-Encodings: $|0\rangle\langle 0|$ Block)**.** *Let V be a unitary that block-encodes a (possibly non-unitary) target operator S as $S = (\langle 0| \otimes I)V(|0\rangle \otimes I)$. Suppose there exist unitaries U_j that block-encode operators R_j as $R_j = (\langle 0| \otimes I)U_j(|0\rangle \otimes I)$. Also suppose there exists a probability distribution p_j such that*

$$\begin{aligned}&\|R_j - S\| \le a \text{ for all } j, \\ &\Big\|\sum_j p_j R_j - S\Big\| \le b,\end{aligned} \tag{22}$$

for some $a, b \ge 0$. Then, the corresponding unitary channel $\Lambda(\rho) = \sum_j p_j U_j \rho U_j^\dagger$ approximates the action of the channel $\mathcal{V}$ as

$$\big\|\bar{\Lambda} - \bar{\mathcal{V}}\big\|_\diamond \le a^2 + 2b, \tag{23}$$

where $\bar{\Lambda}$ and $\bar{\mathcal{V}}$ are channels that access the block-encodings of Λ and $\mathcal{V}$ by appending an ancilla qubit and post-selecting:

$$\begin{aligned}\bar{\Lambda}(\rho) &= (\langle 0| \otimes I) \cdot \Lambda\big(|0\rangle\langle 0| \otimes \rho\big) \cdot (|0\rangle \otimes I) \\ \bar{\mathcal{V}}(\rho) &= (\langle 0| \otimes I) \cdot \mathcal{V}\big(|0\rangle\langle 0| \otimes \rho\big) \cdot (|0\rangle \otimes I).\end{aligned} \tag{24}$$

Proof. First, observe that by linearity

$$\begin{aligned}\bar{\Lambda}(\rho) &= (\langle 0| \otimes I)\Lambda\big(|0\rangle\langle 0| \otimes \rho\big)(|0\rangle \otimes I) \\ &= \sum_j p_j R_j \rho R_j^\dagger,\end{aligned} \tag{25}$$

and

$$\begin{aligned}\bar{\mathcal{V}}(\rho) &= (\langle 0| \otimes I)\mathcal{V}\big(|0\rangle\langle 0| \otimes \rho\big)(|0\rangle \otimes I) \\ &= S\rho S^\dagger.\end{aligned} \tag{26}$$

Therefore, $\bar{\Lambda}$ and $\bar{\mathcal{V}}$ are analogous to the channels Λ and $\mathcal{S}$ considered in Theorem 3. Likewise, S and R_j obey the requisite conditions of Theorem 3 as per the assumptions of this theorem. Accordingly, Theorem 3 implies that

$$\|\bar{\Lambda} - \bar{\mathcal{V}}\|_\diamond \le a^2 + 2b\|S\| \le a^2 + 2b, \tag{27}$$

where we have used that $\|S\| \le 1$ because S is block-encoded in a unitary. □

Next, consider the general case in which the operators are block-encoded by arbitrary projectors Π and Π'. Specifically, in this case, an operator is block-encoded as $S = \Pi V \Pi'$ for projection operators Π, Π'. In practice, this means that S is accessed by first applying Π' to project the quantum state into the block of interest, then applying V, and finally applying Π' to extract the desired block. With this intuition, we can prove the following:

Lemma 4 (Mixing Lemma for Block-Encodings: General Encoding)**.** *Let V be a unitary, that block-encodes a (possibly non-unitary) target operator S as $S = \Pi V \Pi'$, where Π, Π' are orthogonal projectors. Suppose there exist unitaries U_j that block-encode operators R_j as $R_j = \Pi U_j \Pi'$. Also suppose there exists a probability distribution p_j such that*

$$\begin{aligned}&\|R_j - S\| \le a \text{ for all } j, \\ &\Big\|\sum_j p_j R_j - S\Big\| \le b.\end{aligned} \tag{28}$$

Then, the corresponding unitary channel $\Lambda(\rho) = \sum_j p_j U_j \rho U_j^\dagger$ approximates the action of the channel $\mathcal{V}$ as

$$\big\|\bar{\Lambda} - \bar{\mathcal{V}}\big\|_\diamond \le a^2 + 2b, \tag{29}$$

where $\bar{\Lambda}$ and $\bar{\mathcal{V}}$ are channels that access the block-encodings of Λ and $\mathcal{V}$ by applying the projectors Π and Π':

$$\begin{aligned}\bar{\Lambda}(\rho) &= \Pi \cdot \Lambda\big(\Pi'\rho\Pi'\big) \cdot \Pi \\ \bar{\mathcal{V}}(\rho) &= \Pi \cdot \mathcal{V}\big(\Pi'\rho\Pi'\big) \cdot \Pi.\end{aligned} \tag{30}$$

Proof. Analogous to the previous proof, observe that by linearity

$$\begin{aligned}\bar{\Lambda}(\rho) &= \Pi \cdot \Lambda\big(\Pi'\rho\Pi'\big) \cdot \Pi \\ &= \sum_j p_j \Pi U_j \Pi' \cdot \rho \cdot \Pi' U_j^\dagger \Pi \\ &= \sum_j p_j R_j \rho U_j^\dagger,\end{aligned} \tag{31}$$

and

$$\begin{aligned}\bar{\mathcal{V}}(\rho) &= \Pi \cdot \mathcal{V}\big(\Pi'\rho\Pi'\big) \cdot \Pi \\ &= \Pi V \Pi' \cdot \rho \cdot \Pi' V^\dagger \Pi \\ &= S\rho S^\dagger,\end{aligned} \tag{32}$$

where we have use the fact the the orthogonal projectors are Hermitian: $\Pi^\dagger = \Pi$, $\Pi'^\dagger = \Pi'$. Therefore, we again see that $\bar{\Lambda}$ and $\bar{\mathcal{V}}$ are analogous to the channels Λ and $\mathcal{S}$ considered in Theorem 3, which implies that

$$\|\bar{\Lambda} - \bar{\mathcal{V}}\|_\diamond \leq a^2 + 2b\|S\| \leq a^2 + 2b, \tag{33}$$

where we have again used that $\|S\| \leq 1$ because S is block-encoded in a unitary. □

C. The Mixing Lemma for Controlled Operations

The mixing lemma naturally also extends to controlled operations. That is, if the unitaries $\{U_j\}$ and probability distribution p_j satisfy the conditions of the mixing lemma in approximating a target unitary V, then one can approximate a controlled-V operation to the same level of precision by a probabilistic mixture of controlled-U_j's. More precisely, letting c-U_j denote a controlled U_j operation, we have the following:

Lemma 5 (Mixing Lemma for Controlled Operations)**.** *Let c-$V = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes V$ be a controlled unitary, and c-$\mathcal{V}(\rho) =$ (c-$V)\rho($c-$V)^\dagger$ the corresponding channel. Suppose there exist m unitaries $\{U_j\}_{j=1}^m$ and an associated probability distribution p_j that approximate V as*

$$\begin{aligned}&\|U_j - V\| \leq a \text{ for all } j, \\ &\Big\|\sum_{j=1}^m p_j U_j - V\Big\| \leq b,\end{aligned} \tag{34}$$

for some $a, b > 0$. Then, the channel comprising a mixture of controlled unitaries c-$\Lambda(\rho) = \sum_{j=1}^m p_j$(c-$U_j)\rho($c-$U_j)^\dagger$ approximates the controlled channel c-$\mathcal{V}$ as

$$\|\text{c-}\Lambda - \text{c-}\mathcal{V}\|_\diamond \leq a^2 + 2b. \tag{35}$$

Proof. To prove this, note that

$$\big\|\text{c-}U_j - \text{c-}V\big\| = \big\||1\rangle\langle 1| \otimes (U_j - V)\big\| = \|U_j - V\| \leq a, \tag{36}$$

and similarly

$$\begin{aligned}\Big\|\sum_j p_j(\text{c-}U_j) - \text{c-}V\Big\| &= \Big\||1\rangle\langle 1| \otimes \Big(\sum_j p_j U_j - V\Big)\Big\| \\ &= \Big\|\sum_j p_j U_j - V\Big\| \leq b.\end{aligned} \tag{37}$$

Evidently, the controlled unitaries c-U_j and c-V satisfy the conditions of the mixing lemma. Therefore, one can straightforwardly apply the mixing lemma to these operations to conclude that $\|\text{c-}\Lambda - \text{c-}\mathcal{V}\|_\diamond \leq a^2 + 2b$. □

In the context of stochastic QSP, this lemma establishes that stochastic QSP is compatible with controlled target operations, which are integral to many quantum algorithms (e.g., the QSP-based phase estimation algorithm in Ref. [4]). Specifically, this result shows that a controlled target operation can be approximated through stochastic QSP as a probabilistic mixture of controlled QSP operations.

III. GENERALIZATIONS OF STOCHASTIC QSP: PROOFS

In the main text, we introduced stochastic QSP and its generalizations to definite- and indefinite-parity polynomials, Taylor series, trigonometric polynomials, generalized QSP [5], and approximating the entire QSP unitary. Here we provide the proofs for the corollaries comprising these generalizations.

A. Definite and Indefinite Parity

Corollary 6. *Consider the setting of Theorem 3 in the main text, but also suppose that $F(x)$ has definite party. Then there exists an ensemble of polynomials $\{P_j\}$ with the same parity that satisfy the conditions of the theorem.*

Proof. If $F(x) = \sum_{n=0}^\infty c_n T_n(x)$ has even (odd) parity then all c_n for odd (even) n must vanish. In other words, the function is only supported on even (odd) Chebyshev polynomials. Observe that all P_j in the construction are supported only on subsets of the support of $F(x)$. Hence they must also be even (odd). □

B. Taylor Series

Corollary 7. *Suppose $\{B_n(x)\}$ are a collection of basis functions of degree n respectively, which are all bounded as $\|B_n\|_{[-1,1]} \leq 1$. Then the statement of Theorem 3 in the main text holds with $B_n(x)$ in place of $T_n(x)$.*

Proof. This follows from the fact that the only property of the Chebyshev polynomials $T_n(x)$ leveraged in the proof of Theorem 3 is that they were bounded as $\|T_n\|_{[-1,1]} = 1$. □

C. Approximating the Entire QSP Unitary

For a given target function $F(x)$ approximated by a truncated polynomial $P(x)$, QSP yields a unitary $U[P(A)]$ that block-encodes $P(A)$, and thus approximates $U[F(A)]$ and $F(A)$ respectively. While most quantum algorithms consider only $P(A)$, some algorithms use the entire unitary $U[P(A)]$. For example, QSP-based methods for ground-state preparation and energy estimation [6] consider a Hamiltonian H, find a polynomial transformation $P(H)$ that approximates a projector onto an energy range $\Pi_{\leq E} \approx P(H)$, and construct its block-encoding $U[P(H)]$. Supposing that $U[P(H)]$ has k ancilla qubits, they repeatedly perform the map:

$$|\psi\rangle \to U[P(H)] \, |0^k\rangle \, |\psi\rangle \tag{38}$$

followed by a measurement of the k ancillae. With proper interpretation of the measurement result, this approximately implements a positive operator-valued measure with operators $\{\Pi_{\leq E}, I - \Pi_{\leq E}\}$ which forms the core of the algorithms.

Amplitude amplification is another example involving the entire block-encoding unitary. Suppose we have a projector $\Pi \approx P(A)$, and our goal is to prepare the normalized state $\Pi \, |\psi\rangle \, / |\Pi \, |\psi\rangle \, |$ given a unitary that prepares $|\psi\rangle$. Then an amplitude amplification algorithm could construct the required Grover reflection operators using the unitary that prepares $|\psi\rangle$ and the block-encoding unitary $U[P(A)]$ [7].

Suppose that we want to enhance either of the above techniques with stochastic QSP. Now it no longer suffices to show that the quantum channel implemented by stochastic QSP approximates the projected transformation (involving initialization and post-selection of the ancillae to $|0\rangle$). Instead, the channel must approximate $U[F(A)]$ as a whole. We find that this is the case. Suppose that there is some function $G(x)$ accompanying our target function $F(x)$ that satisfies $|F(x)|^2 + (1 - x^2)|G(x)|^2 = 1$ for all $x \in [-1, 1]$. Then the following holds.

Corollary 8. *In the setting of Theorem 3 in the main text, consider the ensemble of quantum circuits that implement stochastic QSP (i.e. the QSP circuits that implement $\{P_j(A)\}$), upon leaving the QSP ancilla qubit(s) unmeasured. Denoting this ensemble by $\{U_j\}$, the quantum channel $\rho \to \sum_j p_j U_j \rho U_j^\dagger$ approximates the map $\rho \to U[F(A)]\rho U[F(A)]^\dagger$ to error $O(\epsilon)$ in diamond norm.*

Proof. The precise form of $U[F(A)]$ in the most general setting for possibly non-Hermitian and possibly non-square A can be reviewed in [7]. In the basis presented in their Table 2, each of its matrix elements either vanish or equal $\pm F(\sigma_i)\sigma_i$ or $G(\sigma_i)\sqrt{1 - \sigma_i^2}$ where σ_i is a singular value of A.

By the conditions of QSP (see Eq. (8) in the main text), each polynomial of the ensemble $P_j(A)$ has a corresponding polynomial $Q_j(x)$ that satisfies

$$|P_j(x)|^2 + (1 - x^2)|Q_j(x)|^2 = 1 \tag{39}$$

for all $x \in [-1, 1]$. Then, for each unitary $U[P_j(A)]$ in the ensemble, the matrix elements in the same basis as considered for $U[F(A)]$ are respectively proportional to either $\pm P_j(\sigma_i)\sigma_i$ or $Q_j(\sigma_i)\sqrt{1 - \sigma_i^2}$.

It follows that we can express the accuracy of the ensemble in terms of the following quantities. We define for some fixed j:

$$\alpha(x) := P_j(x) - F(x), \text{ and} \tag{40}$$
$$\beta(x) := Q_j(x) - G(x). \tag{41}$$

Then:

$$\Big|U[P_j(A)] - U[F(A)]\Big| \leq \sup_x \left[\alpha(x)^2 + \beta(x)^2(1 - x^2)\right] \tag{42}$$

Now we calculate:

$$\begin{aligned} 1 &= |F|^2 + (1 - x^2)\,|G|^2 \\ &= |P_j + \alpha|^2 + (1 - x^2)\,|Q_j + \beta|^2 \\ &= (|P_j|^2 + \alpha P_j^* + \alpha^* P_j + |\alpha|^2) \\ &\quad + (1 + x^2)(|Q_j|^2 + \beta Q_j^* + \beta^* Q_j + |\beta|^2) \\ &= 1 + (\alpha P_j^* + \alpha^* P_j + |\alpha|^2) \\ &\quad + (1 + x^2)(\beta Q_j^* + \beta^* Q_j + |\beta|^2), \end{aligned} \tag{43}$$

where * denotes the complex conjugate here. Thus:

$$(1 + x^2)|\beta Q_j^* + \beta^* Q_j + |\beta|^2| = |\alpha P_j^* + \alpha^* P_j + |\alpha|^2|. \tag{44}$$

We have:

$$\begin{aligned} (1 + x^2)|\beta Q_j^* + \beta^* Q_j + |\beta|^2| &\geq |\beta|^2 \\ |\alpha P_j^* + \alpha^* P_j + |\alpha|^2| &\leq 2|\alpha| + |\alpha|^2 \leq 3|\alpha| \end{aligned} \tag{45}$$

So $|\beta|^2 \leq 3|\alpha| \leq 6\sqrt{\epsilon}$. Having already established that $|\alpha(x)| = |P_j(x) - F(x)| \leq 2\sqrt{\epsilon}$ in the proof of Theorem 3 in the main text, we conclude the quantity in Eq. (42) is bounded by $\sqrt{40\epsilon}$.

Considering $\bar{\alpha}(x) := \sum_j p_j P_j(x) - F(x)$ and $\bar{\beta}(x) := \sum_j p_j Q_j(x) - G(x)$, we observe:

$$\left|\sum_j p_j U[P_j(A)] - U[F(A)]\right| \leq \sup_x \left[\bar{\alpha}(x)^2 + \bar{\beta}(x)^2(1 - x^2)\right] \tag{46}$$

An identical calculation yields the bound $|\bar{\beta}|^2 \leq 3|\bar{\alpha}|$, hence bounding the quantity in Eq. (46) by $\sqrt{10\epsilon} \leq 4\epsilon$.

Now applying Lemma 2 of the main text with $a = \sqrt{40\epsilon}$ and $b = 4\epsilon$, we obtain an error in diamond norm of $a^2 + 2b = 48\epsilon = O(\epsilon)$. Naturally the constant 48 can be tightened significantly using a more careful analysis. □

[1] Yitzhak Katznelson, *An introduction to harmonic analysis* (Cambridge University Press, 2004).

[2] John P Boyd, *Chebyshev and Fourier spectral methods* (Courier Corporation, 2001).

[3] Earl Campbell, "Shorter gate sequences for quantum computing by mixing unitaries," Physical Review A **95** (2017).

[4] John M Martyn, Zane M Rossi, Andrew K Tan, and Isaac L Chuang, "Grand unification of quantum algorithms," PRX quantum **2**, 040203 (2021).

[5] Danial Motlagh and Nathan Wiebe, "Generalized quantum signal processing," PRX Quantum **5**, 020368 (2024).

[6] Yulong Dong, Lin Lin, and Yu Tong, "Ground-state preparation and energy estimation on early fault-tolerant quantum computers via quantum eigenvalue transformation of unitary matrices," PRX Quantum **3** (2022).

[7] András Gilyén, Yuan Su, Guang Hao Low, and Nathan Wiebe, "Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics," in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC '19 (ACM, 2019).