arXiv:2207.01587v3 [quant-ph] 6 Jul 2023

"Proper" Shift Rules for Derivatives of Perturbed-Parametric Quantum Evolutions

Dirk Oliver Theis

Theoretical Computer Science, University of Tartu, Estonia
Thu Jul 6 17:52:07 (CEST) 2023

Banchi & Crooks (*Quantum*, 2021) have given methods to estimate derivatives of expectation values depending on a parameter that enters via what we call a "perturbed" quantum evolution $x \mapsto e^{i(xA+B)/\hbar}$. Their methods require modifications, beyond merely changing parameters, to the unitaries that appear. Moreover, in the case when the B-term is unavoidable, no exact method (unbiased estimator) for the derivative seems to be known: Banchi & Crooks's method gives an approximation.

In this paper, for estimating the derivatives of parameterized expectation values of this type, we present a method that only requires shifting parameters, no other modifications of the quantum evolutions (a "proper" shift rule). Our method is exact (i.e., it gives analytic derivatives, unbiased estimators), and it has the same worst-case variance as Banchi-Crooks's.

Moreover, we discuss the theory surrounding proper shift rules, based on Fourier analysis of perturbed-parametric quantum evolutions, resulting in a characterization of the proper shift rules in terms of their Fourier transforms, which in turn leads us to non-existence results of proper shift rules with exponential concentration of the shifts. We derive truncated methods that exhibit approximation errors, and compare to Banchi-Crooks's based on preliminary numerical simulations.

Keywords: Variational quantum evolution, gradient estimation, parameterized quantum circuit.

Contents

1 Introduction **2**

2 Technical overview **6**
2.1 Math notations, definitions and preliminary facts 6
2.2 Perturbed-parametric evolutions and proper shift rules 7
2.3 Function spaces 8
2.4 Reflection, dilation, and the space of feasible proper shift rules 11
2.5 Feasibility of the Nyquist shift rules 13
2.6 Cost concept and optimality 14
2.7 Truncation 16
2.8 Folding 17
2.9 Non-existence results 21
2.10 Numerical simulation 23

3 Fourier analysis of perturbed-parametric unitaries **26**

Dirk Oliver Theis: dotheis@ut.ee, This research was partly funded by the Estonian Center of Excellence in Computer Science (EXCITE), and by Estonian Research Agency (Eesti Teadusagentuur, ETIS) through grant PRG946.

3.1 Notations . . . 26
3.2 The Fourier-decomposition theorems . . . 27
3.3 Proof of Fourier-decomposition for unitaries . . . 27
3.4 The Fourier spectrum . . . 29

4 The derivative at 1/2 **32**
4.1 Impact of the Fourier-decomposition theorem . . . 32
4.2 Fourier-analytic characterization of derivative-computing measures, case $\mathscr{F}_*^{\oplus}$ . . . 34
4.3 Proof of the Space-of-Feasible-PSRs proposition . . . 35
4.4 Proof of feasibility of the Nyquist shift rule . . . 36

5 Details about reflection and dilation **38**
5.1 Proof of the general facts . . . 38
5.2 Derivation of the Nyquist shift rules . . . 39

6 Optimality **40**
6.1 The Weak Duality Theorem . . . 40
6.2 Proof of optimality of the Nyquist shift rules . . . 40
6.3 The support of optimal proper shift rules . . . 41

7 Truncation and folding **41**
7.1 Truncation . . . 41
7.2 Folding . . . 41

8 Sketches of proofs of non-existence results **45**
8.1 Impossibility of exponential concentration . . . 45
8.2 Shift rules with approximation error . . . 46

9 Discussion and outlook **46**

A Miscellaneous math **47**
A.1 Sums of reciprocals of squares of odd integers . . . 47
A.2 Simultaneous Diophantine approximation . . . 47
A.3 Exponential concentration . . . 49

B Finite-dimensional perturbation theory (for the theoretical computer scientist) **49**

C Fourier-analytic characterization of derivative-computing measures, general case **50**

D Additional graphs from numerical simulations **53**

1 Introduction

One paradigm in near-term quantum computing is that of variational quantum algorithms (VQAs): Quantum algorithms which contain real-number parameters and which must be trained, i.e., the parameters must be optimized — similarly to classical differentiable programming, e.g., artificial neural networks.

In fault-tolerant quantum computing, evidence has been given (see, e.g., [1] and the references therein) that the concept of quantum programs depending on parameters that are fitted to data may turn out to be an important component in applications of quantum computing, beyond the realm of machine learning and AI.

Pre-fault-tolerance, under the name of Parameterized Quantum Circuits or Variational Quantum Circuits, VQAs are at the heart of proposals for quantum-computer based simulations of molecules and condensed matter materials, quantum machine learning and quantum AI, some approaches to quantum-based combinatorial optimization, for example some uses of the Quantum

Approximate Optimization Algorithm, and even applications such as linear-system solving and factoring (e.g., [2, 3, 4, 5, 6, 7]). But the concept of parameterized quantum evolution for pre-fault-tolerant quantum simulation and computation extends beyond gate-based quantum circuits (e.g., [8, 9, 10]). In this paper, our interest is in quantum evolutions where the parameters do not enter via simple gates.

As optimization/training algorithms based on estimates of derivatives (e.g., variants of Stochastic Gradient Descent) outperform derivative-free methods in practice in the training of variational quantum algorithms (cf. [11] and the references therein), there is the need to obtain, efficiently, estimates of derivatives with respect to the parameters in a VQA. Starting with seminal works by Li et al. [12] and Mitarai et al. [3], unbiased estimators for derivatives have been obtained efficiently using so-called shift rules: If f denotes the expectation-value function dependent on, w.l.o.g., a single parameter, a shift rule is a relation

$$f'(x) = \sum_j u_j \, f(x - s_j) \tag{1a}$$

where the u_j (coefficients) and s_j (shifts) are fixed real numbers, and the equation holds for all x. This is a convolution of the expectation-value function with a finite-support measure, and by replacing "finite-support" with "finite", this extends to shift rules with a continuum of shifts:

$$f'(x) = \int f(x - s) \, d\phi(s) =: f * \phi(x), \tag{1b}$$

where ϕ is a finite measure on $\mathbb{R}$, and "$*$" denotes convolution.

The notion of a shift rule has been extended, e.g. [13] speak of a Stochastic Shift Rule for a method that involves other modifications of the quantum evolution than merely changing the parameters.[1] To avoid ambiguity, we use the term Feasible *Proper* Shift Rule (feasible PSR) for the ϕ in relation (1b).

We add the qualification "feasible" as we also study a couple of notions of approximate PSRs where (1b) holds up to an approximation error[2], leading to biased estimators.

The present paper studies PSRs for derivatives of the type of quantum evolutions which is the subject of [13]: We expect that the parameter, x enters in the following form

$$e^{i(xA+B)/\hbar}. \tag{2}$$

We refer to (2) as "perturbed-parametric" unitary, and we speak of a perturbed-parametric expectation-value function if a measurement depends on x via a perturbed-parametric unitary.

Banchi & Crooks [13] propose a stochastic shift rule for estimating derivatives of perturbed-parametric expectation-value functions, building on an equation by Feynman [14, 15]. As said above, their rule modifies the underlying quantum evolution itself: To obtain an estimate of the derivative, the unitary (2) must be replaced by a concatenation

$$e^{it_3\,(x_3A+B)/\hbar} \cdot e^{i\beta\,(x_2/\beta\cdot A+B)/\hbar} \cdot e^{it_1\,(x_1A+B)/\hbar}, \tag{3}$$

for certain values of t_*, x_*, and a small $\beta > 0$; if the B-term in the evolution can be switched off, $\beta \to 0$ effects the unitary $e^{ix_2A/\hbar}$. For $\beta > 0$, Banchi & Crooks refer to their method as Approximate Stochastic Parameter Shift Rule.

Modifying the underlying quantum evolution has some small technical disadvantages, as it requires a re-calculation of the schedule of quantum-control pulses.

[1] There is another sense in which Banchi & Crooks term shift rule differs from our "proper" shift rule in Eqn. (1b): Banchi & Crooks include the *estimator* for the resulting value into the term stochastic shift rule, while we, following Mitarai et al. keep the relation (1b) (the "proper" shift rule)) separate from an algorithm implementing an estimator of the quantity $f * \phi(x)$.

[2] Difference quotients in numerical differentiation are examples of PSRs that are "not quite feasible".

Moreover, Banchi & Crooks's shift rule has the disadvantage that it cannot give an unbiased estimate of the derivative in situations where the B-term in the perturbed-parametric unitary (2) is unavoidable: A $O(\beta)$ approximation error results.

In many situations where perturbed-parametric unitaries arise, a certain small bias in the estimate of the derivative may be allowed. Indeed, in pre-fault-tolerant settings, expressions such as (2) are approximations up to control inaccuracies.

But in the case where the B-term is unavoidable, Banchi-Crooks's Approximate Stochastic Parameter Shift Rule requires, to achieve a good approximation, a small $\beta > 0$, so the parameter value, $1/\beta$, is large. In practice, large parameter settings may not be desirable (e.g., they could result in more cross talk) or even technically possible.[3]

Moreover, in some situations where perturbed-parametric quantum evolutions arise, the factor β in the middle factor in (3) is a duration of some pulse, and effecting unitaries such as $e^{i\beta\,(x_2/\beta\cdot A+B)/\hbar}$ involves a narrow frequency band for that pulse (e.g., the resonant frequency of a qubit). In these situations, the Fourier Uncertainty Principle may make it difficult in principle to choose an arbitrarily small β.

Contributions of this paper

We present a PSR feasible for expectation-value functions where the parameter enters in the form of (2); we call it the Nyquist Shift Rule. Our method is exact, so that estimators without bias can be constructed; one such estimator, that we present below, has the same variance as Banchi & Crooks's Stochastic Parameter Shift Rule.

As the tag "Nyquist" suggests, our PSR is based on Fourier analysis. The connection between PSRs and Fourier spectra was probably first observed in [16], and exploited in [17] and elsewhere; the present paper shows ways to exploit Fourier analytic properties of perturbed-parametric expectation-value functions. Indeed, our results are based on the observation that the perturbed-parametric expectation-value function is — in finite dimension — band limited: Its Fourier spectrum (i.e., support of the Fourier transform) is contained in an interval $[-K, +K]$, where K is determined by the difference between the largest and smallest eigenvalues of A. For functions like that, immediately the Nyquist-Shannon Sampling Theorem comes to mind, here stated in the form of the Shannon-Whittaker Interpolation Formula: E.g., for functions f with Fourier spectrum contained in $[-1/2, +1/2]$

$$f = \operatorname{sinc} \ast f := \sum_{n\in\mathbb{Z}} \operatorname{sinc}(\,\square - n)\, f(n) \tag{4a}$$

where "$\square$" is placeholder for the variable, $\operatorname{sinc} := \sin(\pi\,\square)/(\pi\,\square)$ and "$\ast$" refers to sum-convolution[4] as defined with the RHS. From here, obtaining a derivative[5] seems straight forward:

$$\partial f = \partial(\operatorname{sinc} \ast f) = (\partial \operatorname{sinc}) \ast f = f \ast (\partial \operatorname{sinc}). \tag{4b}$$

The resulting PSR[6] $\operatorname{sinc}'(\,\square - n)$ $(n \in \mathbb{Z})$ decays only as $1/|n|$ for $|n| \to \infty$. This can be fixed by picking a sweet spot point x for which $\operatorname{sinc}'(x - n))$ decays as $1/|n|^2$, namely $x = 1/2$. Restarting with

$$f'(1/2) = \sum_{n\in\mathbb{Z}} \operatorname{sinc}'(1/2 - n)\, f(n) \tag{4c}$$

and then performing a reflection of f (the details are in §2.4), results in, for every $x \in \mathbb{R}$,

$$f'(x) = \sum_{a\in 1/2+\mathbb{Z}} \frac{(-1)^{a+1/2}}{\pi a^2} f(x - a), \tag{4d}$$

[3] The methods in the present paper cannot substantially alleviate the problem of large magnitudes of parameter-values, but in preliminary numerical simulations (§2.10) they appeared to be more efficient in that regard.

[4] Note: "$\ast$" $\neq$ "$*$"; we won't use "$\ast$" in the main parts of the paper.

[5] We use "∂g" and "g'" interchangeably for the derivative of a function g of a single argument.

[6] Warning: It isn't one, as the measure is not finite.

as $\operatorname{sinc}'(1/2 - n) = (-1)^{n-1}/(\pi(n - 1/2)^2)$.

Generalizing to other values of K for the interval $[-K, +K]$ containing the Fourier spectrum, one obtains the following family of PSRs, which we call Nyquist Shift Rules:

$$\phi_K := 2K \cdot \sum_{a \in 1/2+\mathbb{Z}} \frac{(-1)^{a+1/2}}{\pi a^2} \delta_{a/2K}, \text{ where } K > 0; \tag{5}$$

with δ_x denoting the Dirac point measure at the point $x \in \mathbb{R}$.

While this hand-waiving argument captures the starting point of the research presented in this paper, there are a number of problems with it. First of all, the Shannon-Whittaker formula (4a) as presented doesn't work: Plugging in $f = \cos(\pi\,\square)$ at the point $1/2$ gives infinity on the RHS.[7] Secondly, in (4b), we are taking the derivative under a sum which (even if it converged) doesn't converge absolutely, causing headaches and high blood pressure. However, if we are able to arrive at (4c), then (4d) will follow, and the feasibility of the Nyquist shift rules (5) will be established.

The contributions one by one. *(1)* Counting the observation about the Fourier spectra of perturbed-parametric expectation-value functions (see §2.3) as the first contribution of this paper, *(2)* the second contribution is a rigorous proof that the Nyquist shift rule (5) is feasible for expectation-value functions where the difference between largest and smallest eigenvalues of A is at most[8] $K/2\pi\hbar$; see §2.5.

(3) *En passant*, in §2.4, we characterize the set of all feasible PSRs (for fixed $K > 0$) in terms of their Fourier transforms — mirroring the characterization via a system of linear equations in [18] but with infinitely many equations. As a first consequence of that characterization, we can show (§2.4.b) that the space[9] of all feasible PSRs has infinite dimension (for each fixed K) — so you could definitely say that there are many of them.

(4) As a second consequence of the characterization in §2.4, we can prove some non-existence results for particularly nice feasible PSRs (§2.9), e.g., it's a one-liner to see that there is no PSR with compact support that is feasible for the frequency band $[-1/2, +1/2]$. It takes a little more effort to prove that there is no feasible PSR which is exponentially concentrated, indeed, there isn't even a family of exponentially concentrated PSRs that are, in a wide sense,"nearly feasible" (definition in Corollary 2.34): You cannot let the approximation error tend to 0 without blowing up either the tail or the "cost" (see next item) of the PSR.

(5) The "cost" of PSRs: In §2.6.a, in parallel with the results in [18], we show that the (total-variation) norm of a PSR equals the *worst-case standard deviation* of a single-shot estimator for the convolution of the expectation-value function with the PSR (see Fig. 2 below). Moreover, we prove that, for each $K > 0$, there is no feasible PSR that has smaller norm than our Nyquist shift rule ϕ_K, i.e., the Nyquist shift rules are optimal; see §2.6. The Stochastic Parameter Shift Rule of Banchi & Crooks [13] has the same worst-case standard deviation (but it isn't a "proper" shift rule).

While from a mathematical point of view, the Nyquist shift rules in (5) are the "right" ones, they require to query the expectation-value function at arbitrarily large parameter settings. Indeed, the expected magnitude of a parameter setting is ∞ as the harmonic sum diverges. The usual remedy for this situation in the context of Nyquist-Shannon-Whittaker theory is to truncate the sum, which introduces an approximation error. As this paper's contribution *(6)*, in §2.7, we discuss truncation of Nyquist shift rules to shifts of finite magnitude, and upper-bound the error: In terms of big-Os, the decay in the approximation error is comparable to the Approximate Stochastic Parameter Shift Rule from [13], namely $O(1/\Delta)$, where Δ is the maximum magnitude of a shift. *(7)* We present results from preliminary numerical simulations based on an implementation of the

[7] Indeed, for it to converge for all bounded band-limited functions f, we must require that the Fourier spectrum of f is contained in $[-1/2 + \delta, +1/2 - \delta]$, and the speed of convergence will depend on $\delta > 0$.

[8] The factor $2\pi\hbar$ comes from the choice of Fourier transform in this paper.

[9] This is an affine space of functions.

Method	Variant	Apx err	Max MoPV	Avg MoPV
Banchi-Crooks	approximate	$O(\beta)$	$4/\beta$	$4/\beta$
Nyquist	exact	0	n/a	∞
	truncated	ε	$\lvert x\rvert + 2/\varepsilon\pi + 1/2K$	$\lvert x\rvert + \ln(1/\varepsilon) + O_K(1)$
→ parameter-folded, x unconstrained		ε	$< 4\pi K/\varepsilon + p$	
→ parameter-folded, $x \in [-p, +p]$		ε	$O_K(1/\sqrt{\varepsilon})$	$< p + O_K(\ln(1/\sqrt{\varepsilon} + 2p))$

Table 1: **Comparison of methods.** The table shows the approximation error ("apx err"), maximum magnitude of parameter values ("max MoPV") and average magnitude of parameter values when simple estimators are used. "Banchi-Crooks" refers to the Approximate Stochastic Parameter Shift Rule in [13], which depends on a small parameter, $\beta > 0$, that determines the query points. The methods presented in this paper are dubbed "Nyquist"; "exact" refers to (5) with the Single-Shot Estimator, Fig. 2; "truncated" refers to the method presented in §2.7 with a variant of the Single-Shot Estimator; "parameter-folded" refers to the method presented in §2.8.c with the estimator from §2.8.b. In the parameter-folded methods, the maximum and average magnitudes of parameter values depend not on the parameter value x at which the derivative is requested; instead it is influenced by a common divisor (cf. Fn. 25), $1/p$, of the eigenvalues of A. For simplicity, we have set $\hbar = 1/2\pi$.

Approximate Stochastic Parameter Shift Rule and a simplified version of the truncated Nyquist shift rule, see §2.10; the computer code is available online[10] for experimenting.

(8) Finally, in §2.8, we discuss a more sophisticated way of confining parameter values to finite magnitudes, which we call Folding. We discuss a Parameter Folding method that exploits a deeper Fourier-analytic property of perturbed-parametric expectation-value functions, based on eigenvalue perturbation theory. Parameter Folding our Nyquist shift rule results in an approximation error $O(1/c^2)$ for points x that have magnitude $(1 - \Omega(1))c$, where c is the maximum magnitude of a parameter value.

We summarize the results about various types of proper, improper, and folded shift rules in Table 1.

Organization of the paper

In the next Section, we give a technical, mathematically rigorous overview of the results of this paper, with an emphasis on motivation and easy on the proofs. The more technical proofs are in the Sections 3–8. Section 9 discusses some questions that arise, and points to future work.

Appendices A and B hold math that is well known or easily derived, added for convenience; in Appendix C we prove, for the sake of completeness of the presentation, the characterization of feasible proper shift rules in a more general case than is needed for expectation-value functions. Appendix D has additional graphs from numerical simulations (cf. §2.10).

The author aims for the content of this paper, *from this point on*, to be fully mathematically unambiguous and rigorous, and welcomes any criticism that points out where this goal has been missed.

2 Technical overview

This section presents the results of the paper in full mathematical rigor, with an emphasis on motivation: we state the results and discuss their relationships. In terms of proofs, we give a few that help to motivate, banning the more technical or longer ones to later sections.

2.1 Math notations, definitions and preliminary facts

In this paper, a *finite measure* is a finite (not necessarily positive) regular Borel measure on $\mathbb{R}$; we use the phrase *signed measure* and *complex measure,* resp., to emphasize that only real or

[10] https://dojt.github.io/storage/nyquist/

also complex values are allowed. In this paper, finiteness is implied in the term signed/complex measure.

Complex measures have a canonical decomposition into two signed measures, which in turn have a canonical decompositions into two positive measures (Jordan decomposition). The sum of the resulting four (or 2) positive measures in the decomposition of a finite measure μ is called the *total-variation measure* and denoted by $|\mu|$. The *total-variation norm* of a finite measure μ is defined as $\|\mu\| := \int \mathbf{1}\, d|\mu|$, and satisfies

$$\|\mu\| = \sup_f \int f\, d\mu, \tag{6}$$

where the supremum ranges over all measurable functions f with $\|f\|_\infty \leq 1$. This equation shows that convergence $\mu_n \xrightarrow{n\to\infty} \mu$ in total variation implies convergence for every bounded measurable function f: $\int f\, d\mu_n \xrightarrow{n\to\infty} \int f\, d\mu$ — we use that implicitly all the time.

Convergence in the total-variation norm of the sum of measures defining the Nyquist shift rule (5) can easily be checked.

For a measurable mapping $\tau\colon \mathbb{R} \to \mathbb{R}$ and a finite measure μ on $\mathbb{R}$, we denote by $\tau(\mu)$ the image of the measure under τ (aka push-forward measure). It is defined by $\tau(\mu)(E) = \mu(\tau^{-1}(E))$ for every Borel set E, or, equivalently by $\int f\, d\tau(\mu) = \int f \circ \tau\, d\mu$ for all bounded measurable functions f. For example,

$$\tau(\delta_a) = \delta_{\tau(a)}. \tag{7}$$

We will repeatedly use the following fact, without reference to it: For every measurable $\tau\colon \mathbb{R} \to \mathbb{R}$, the mapping $\mu \mapsto \tau(\mu)$ takes complex measures to complex measures, is linear, and is continuous in the total-variation norm; in particular, if $\mu := \sum_{j\in\mathbb{N}} \alpha_j \mu_j$ is a total-variation-norm convergent series, where the μ_* are complex measures and the α_* are complex numbers, then $\tau(\mu) = \sum_j \alpha_j \tau(\mu_j)$.

The terms *Fourier transform, support,* and *Fourier spectrum* (the support of the Fourier transform) refer to the concepts for tempered distributions; in the case of finite measures, the Fourier transform coincides (up to technicalities) with the Fourier-Stieltjes transform:

$$\hat{\mu}(\xi) := \int e^{-2\pi i \xi x}\, d\mu(x) \text{ (μ finite measure on $\mathbb{R}$, $\xi \in \mathbb{R}$)};$$

note the 2π in the exponent. The inverse Fourier transform is denoted by $\check{\square}$. We also use the notations $(\dots)^\wedge$ and $(\dots)^\vee$ for the (inverse) Fourier transform of the function / measure / distribution in the place of "…".

While the results in this paper can be extended to multi-parameter functions, here, we focus solely on a single parameter. As a consequence, all function spaces will be spaces of functions on the real line. For that reason, when using standard function-space notation, we omit "$(\mathbb{R})$": I.e., L^1 (space of absolutely integrable functions on $\mathbb{R}$), L^2 (space of square integrable functions on $\mathbb{R}$), $\mathscr{C}_0$ (space of continuous functions on $\mathbb{R}$ which vanish towards $\pm\infty$), $\mathscr{C}^1$ (space of continuously differentiable functions on $\mathbb{R}$), $\mathscr{C}^1_b$ (space of bounded continuously differentiable functions on $\mathbb{R}$), $\mathcal{S}$ (space of Schwartz-functions on $\mathbb{R}$), $\mathcal{S}'$ (space of tempered distributions on $\mathbb{R}$), etc.

We use the word *smooth* to mean continuously differentiable.

2.2 Perturbed-parametric evolutions and proper shift rules

The author apologizes to all physicists for choosing the barbaric normalization[11] $h = 1$, i.e., $\hbar := 1/2\pi$, in the following definition.

[11] This normalization simplifies the link to Fourier Analysis (with our Fourier transform $\hat{f}(\xi) = \int e^{-2\pi i \xi x} f(x)\, dx$).

Definition 2.1 (Perturbed-parametric unitaries and expectation-value functions)**.** For Hermitian operators A, B on the same finite-dimensional Hilbert space, we refer to the operator-valued function

$$x \mapsto e^{2\pi i(xA+B)} \tag{8}$$

as a *perturbed-parametric unitary* or a *perturbed-parametric unitary function.* A *perturbed-parametric expectation-value function* is a function of the following form:

$$\mathbb{R} \to \mathbb{R} \colon x \mapsto \operatorname{tr}\big(M\, e^{2\pi i(xA+B)}\, \varrho\, e^{-2\pi i(xA+B)}\big) \tag{9}$$

where ϱ is a positive operator and M a Hermitian operator, both on the same Hilbert space as A, B. To avoid trivial border cases, we require that A is not a scalar multiple of the identity operator.

This definition captures the typical setting in which VQAs are used today:

1. Prepare an n-qubit system in an initial state, say $|0^n\rangle$;
2. Subject the system to a time-dependent evolution (possibly depending on other parameters, but not x);
3. Subject the system to a evolution with Hamiltonian $-(xA + B)$ for one unit of time;
4. Subject the system to further time-dependent evolution (possibly depending on other parameters, but not x);
5. Measure an observable.

Here, the state ϱ would be reached after step #2, and the observable M would be $\mathcal{E}^\dagger(M_0)$, where $\mathcal{E}$ is the quantum operation resulting from the evolution in step #4 and M_0 the observable in step #5.

In parallel[12] with [18], we make the following definitions.

Definition 2.2 (Proper shift rule)**.** A *proper shift rule (PSR)* is a finite signed Borel measure on $\mathbb{R}$; in a *complex PSR* a complex measure is allowed.

Definition 2.3 (Feasible proper shift rule)**.** Let $\mathscr{G}$ be a vector space of real-valued, bounded, smooth functions defined on $\mathbb{R}$. A complex PSR ϕ is called *feasible for* $\mathscr{G}$*,* if for all $f \in \mathscr{G}$ and all $x \in \mathbb{R}$, we have

$$f'(x) = f * \phi(x), \tag{10}$$

where "$*$" is convolution: For all $x \in \mathbb{R}$,

$$f * \phi(x) := \int f(x - s)\, d\phi(s).$$

Let's take as an example the symmetric difference quotient, $f'(x) \approx (f(x + \varepsilon) - f(x - \varepsilon))/2\varepsilon$, for fixed $\varepsilon > 0$: The PSR is $(\delta_{-\varepsilon} - \delta_\varepsilon)/2\varepsilon$. It is feasible for the space[13] of polynomials of degree at most 2, but it incurs an approximation error (i.e., is not feasible) on any non-constant expectation-value function.

2.3 Function spaces

In this paper, as in [18], we address questions about parameterized quantum evolutions by proving theorems about function spaces, and about membership of expectation-value functions in these spaces. In this section, we define the spaces and present the facts regarding membership of expectation-value functions. For convenient reference, Table 2 gives an overview of all spaces we use.

For convenience, and to avoid trivial border cases, we make the following definition.

[12] In [18, §2.2], this definition occurs with the (additional) restriction that the *support* of ϕ must be finite, targeting, specifically, shift rules as in (1a).

[13] Slightly cheating here: Degree 1 and 2 polynomials are, of course, not bounded functions.

Symbol	Semantic
$\mathscr{G}$	Generic function space (of smooth bounded real-valued functions)
	Space of smooth bounded real-valued functions...
$\mathscr{K}_\Xi$	... with Fourier spectrum contained in the frequency set Ξ.
$\mathscr{F}_K$	... with Fourier spectrum contained in the interval $[-K,+K]$.
$\mathscr{F}_K^\downarrow$	... in $\mathscr{F}_K$, which are of linear decay.
$\mathscr{F}_\Xi^\oplus$	... in the (non-orthogonal) direct sum $\mathscr{K}_\Xi + \mathscr{F}^\downarrow_{\max \Xi}$.

Table 2: **Overview of function spaces.** A *frequency set* is a finite, symmetric subset $\Xi \subset \mathbb{R}$ with $|\Xi| \geq 2$ (Def. 2.4). For frequency sets, $\max \Xi := \max_{\xi\in\Xi} \xi = \max_{\xi\in\Xi} |\xi|$ holds.

Definition 2.4 (Frequency set)**.** A *frequency set* is a *finite* set $\Xi \subset \mathbb{R}$ that is symmetric (i.e., $-\Xi = \Xi$), and that has at least two elements, $|\Xi| \geq 2$.

Let Ξ be a frequency set. The space of real-valued, bounded, smooth functions with Fourier spectrum contained in Ξ is denoted by[14]

$$\mathscr{K}_\Xi = \left\{ f\colon x \mapsto \sum_{\xi\in\Xi} b_\xi e^{2\pi i \xi x} \;\middle|\; b \in \mathbb{C}^\Xi \text{ with } \forall \xi \in \Xi\colon\ b_{-\xi} = b_\xi^* \right\}, \tag{11}$$

as in [18]: There, convex optimization techniques and computations are explored, based on the fact that this space has finite dimension.

The spaces in the present paper, though, are not finite dimensional: For positive real K, we denote by $\mathscr{F}_K$ the set of smooth, bounded real-valued functions on $\mathbb{R}$ with Fourier spectrum contained in $[-K,+K]$

$$\mathscr{F}_K := \left\{ f \in \mathscr{C}^1_b \;\middle|\; f \text{ real valued},\ \operatorname{Supp} \hat{f} \subseteq [-K,+K] \right\}.$$

From Paley-Wiener theory [19, Theorem IX.12] we know that every tempered distribution with compact Fourier spectrum is indeed a function which is (extendable to a) holomorphic (function) with polynomial growth (in the real part). Hence, the set $\mathscr{F}_K$ contains all tempered distributions with two constraints: *(1)* on the Fourier spectrum, and *(2)* boundedness. The boundedness is necessary, or otherwise the convolution with finite measures would not be well defined.

We find it more convenient to work with the function space, but the prime example that we are interested in are the perturbed-parametric expectation-value functions from Def. 2.1. The following proposition gives the connection. (We denote by $\mathrm{eig}_{\min}(\cdot), \mathrm{eig}_{\max}(\cdot)$ the smallest and largest, resp., eigenvalues of an operator.)

Proposition 2.5. *With the notations of Def. 2.1, set* $K := \mathrm{eig}_{\max}(A) - \mathrm{eig}_{\min}(A)$*. The expectation-value function is a bounded analytic function whose Fourier spectrum is contained in* $[-K,+K]$*, and hence it is a member of* $\mathscr{F}_K$*.*

As everybody will surmise, the proof of the proposition, in §3.4, is based on the Lie Product Formula, aka "Trotterization".

The work in [18] is based on the function space $\mathscr{K}_\Xi$, which is justified as every function in that space is an expectation-value function of a variational quantum circuit [18, Prop. 4]. In the present paper, that is not the case: There are functions in $\mathscr{F}_K$ that are not perturbed-parametric expectation-value functions[15]. Working with the spaces $\mathscr{F}_*$ is nothing else but a convenient abstraction, simplifying the reasoning for much of what we are doing in this paper — particularly the non-existence proofs. At some points, a more refined Fourier-analytic abstraction of perturbed-parametric expectation-value (and unitary) functions is more convenient or necessary. We prepare that with the following definition.

[14] The equation follows from standard tempered distribution theory, based on the boundedness of the functions in $\mathscr{K}_\Xi$; cf. Appendix A.2.a.

[15] For example, the function $x \mapsto \sum_{j=2}^\infty 1/j^2 \cos(2\pi x/j)$ is in $\mathscr{F}_{1/2}$, but does not arise from Def. 2.1.

Definition 2.6 (Linear decay)**.** For non-negative real constants c, C, we say that a continuous function f of a real variable *decays linearly* (or is of *linear decay*) *with decay constants* (c, C), if for all $x \in \mathbb{R}$ with $|x| \geq c$ we have

$$|f(x)| \leq C/|x|$$

Defining, as usual, the *difference set* of a set Λ as $\operatorname{Diff}\Lambda := \{\lambda - \lambda' \mid \lambda, \lambda' \in \Lambda\}$, textbook eigenvalue perturbation theory (in finite dimension!) will give us the following.

Proposition 2.7 (Fourier-decomposition for perturbed-parametric EVFs)**.** *With the notations of Def. 2.1, let f be a perturbed-parametric expectation value function, and denote $\Xi := \operatorname{Diff}\operatorname{Spec} A$.*

There exists a function $f_1 \in \mathscr{K}_\Xi$ such that $f_0 := f - f_1$ is of linear decay.

In Section 3, we obtain Prop. 2.7 as a consequence of a corresponding Fourier-decomposition for perturbed-parametric unitary functions (Prop. 3.1). Details and the proofs of both propositions are in §§3.2–3.3.

With the notations of Prop. 2.7, as the tempered distributions with support contained in a given set form a subspace, and since both f (by Prop. 2.5) and f_1 have Fourier spectra that are contained in $[-K, +K]$ for $K := \max \Xi$, we find that f_0 also has a Fourier spectrum that is contained in $[-K, +K]$. This motivates the following definition.

Definition 2.8. For $K > 0$, we denote by $\mathscr{F}^{\downarrow}_K$ the vector space of analytic functions that are of linear decay for some choice of decay constants, and whose Fourier spectrum is contained in the interval $[-K, +K]$.

Combining Propositions 2.5 and 2.7, we obtain the following as a direct consequence.

Corollary 2.9. *With the notations of Def. 2.1, set $\Xi := \operatorname{Diff}\operatorname{Spec} A$. Expectation-value functions are members of the (non-orthogonal) direct sum*

$$\mathscr{F}^{\oplus}_\Xi := \mathscr{K}_\Xi + \mathscr{F}^{\downarrow}_{\max \Xi}.$$

In other words, every perturbed-parametric expectation value function f can be uniquely decomposed as $f = f_1 + f_0$ where

- *$f_1 \in \mathscr{K}_\Xi$, i.e., a (finite) linear combination of $\sin(2\pi\xi\,\square)$ and $\cos(2\pi\xi\,\square)$ with frequencies $\xi \in \Xi$;*
- *$f_0 \in \mathscr{F}^{\downarrow}_{\max \Xi}$, i.e., an analytic function of linear decay whose L^2-Fourier transform is a square-integrable function with support in $[\operatorname{eig}_{\min}(A) - \operatorname{eig}_{\max}(A), \operatorname{eig}_{\max}(A) - \operatorname{eig}_{\min}(A)]$.*

The only part of Corollary 2.9 that we have not discussed yet is the directness of the sum (which is equivalent to the uniqueness of the decomposition); we refer to Prop. A.4 in Appendix A.2.

The convenience gain of the Fourier-decomposition theorem lies in the fact that functions of linear decay are square integrable. So the Fourier transform $\widehat{f_0}$ is a function (not an evil tempered distribution), and, thanks to the compact support, the inverse Fourier transform is realized simply by the integral:

$$f_0(x) = \int_{-K}^{K} e^{2\pi i \xi x} \widehat{f_0}(\xi)\, d\xi \quad \text{for all } x \in \mathbb{R}$$

— no tempered distribution theory required, not even improper integrals. The consequence is that, in some occasions, $\mathscr{F}^{\oplus}_\Xi$ is technically easier to work with in terms of the Fourier transform (e.g., see the discussion following Lemma 2.12 in §2.4.a below). In other occasions, the additional knowledge in $\mathscr{F}^{\oplus}_*$ does not seem to translate into reduced technical complexity of the proofs (e.g., the proof of Prop. 2.5 above).

2.3.a Feasibility for $\mathscr{F}_{\max \Xi}$ vs feasibility for $\mathscr{F}_\Xi^\oplus$

As

$$\mathscr{F}_\Xi^\oplus \subsetneq \mathscr{F}_{\max \Xi},$$

every PSR feasible for $\mathscr{F}_{\max \Xi}$ is also feasible for $\mathscr{F}_\Xi^\oplus$, and it is conceivable that there could be PSRs that are feasible for $\mathscr{F}_\Xi^\oplus$ but not feasible for $\mathscr{F}_{\max \Xi}$. This is not the case. Based on the characterization of PSRs feasible for $\mathscr{F}_{\max \Xi}$ in §2.4.a below, we will show the following.

Proposition 2.10. *Let Ξ be a frequency set. Every complex PSR feasible for $\mathscr{F}_\Xi^\oplus$ is also feasible for $\mathscr{F}_{\max \Xi}$.*

(The proof is in §4.1.) Hence, for every frequency set Ξ, the PSRs feasible for the smaller space $\mathscr{F}_\Xi^\oplus$ of Corollary 2.9 are exactly the same as those feasible for the larger space $\mathscr{F}_{\max \Xi}$ used in Prop. 2.5.

In terms of practical use of the PSRs, the structure of $\mathscr{F}_\Xi^\oplus$ can be exploited: In §2.8 below, we will discuss the concept of Folding which exploits the linear decay condition to achieve a shift-rule-ish method with an approximation error. The approximation error decays quickly (quadratically) with the magnitudes of the parameter values at which the perturbed-parametric expectation-value function is queried.

2.4 Reflection, dilation, and the space of feasible proper shift rules

If ϕ is a PSR feasible for a space $\mathscr{G}$ (see Def. 2.3) then, for $f \in \mathscr{G}$ and all $x \in \mathbb{R}$ we have $f'(x) = \int f(x-s)\, d\phi(s)$. Choosing $x = 0$ as an anchor point, as it were, we find $f'(0) = \int f\, d\bar{\phi}$ where $\bar{\phi} := (-\square)(\phi)$ is the image of ϕ under the measurable mapping $(-\square)\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto -x$, the reflection on 0. Conversely, any complex measure $\bar{\phi}$ satisfying that integral equation (for all f) gives rise to a feasible PSR, via $\phi := (-\square)(\bar{\phi})$. This approach leads to the characterization of the set of (finite-support) shift rules via the system of equations in [18, Eqn. (8)].

In the present paper, we choose an anchor point different from 0 — simply because it is convenient for the concrete Nyquist Shift Rules that we will work with: The anchor point will be $1/2$. Moreover, as discussed in the introduction (and as is evident in (5)), the Nyquist shift rule takes its simplest form when the Fourier spectrum is $[-1/2, 1/2]$.

For these reasons, we will consider measures μ with the following property:

$$g'(1/2) = \int g\, d\mu \qquad \text{for all } g \in \mathscr{F}_{1/2}; \tag{12}$$

the space which g is allowed to be in will be swapped out for $\mathscr{F}_\Xi^\oplus$ in some places below. In any case, we will speak of a *derivative-computing measure.*

We start by providing the details about the connection between μ's and PSRs. First of all, note that if ϕ is a PSR for a function space containing a certain function g, then

$$g'(1/2) = g * \phi(1/2) = \int g(1/2 - s)\, d\phi(s) = \int g \circ \tau\, d\phi \int g\, d\tau(\phi),$$

where $\tau\colon \mathbb{R} \to \mathbb{R}\colon s \mapsto 1/2 - s$ is the reflection on $1/2$. This means: If ϕ is a PSR feasible for $\mathscr{F}_{1/2}$, then $\mu := \tau(\phi)$ satisfies (12). Dilation then gives feasibility for $\mathscr{F}_K$ in the case $K \neq 1/2$. We summarize in the form of the following proposition (proof details in §5.1).

Proposition 2.11. *For every $K > 0$, there is a canonical isomorphism between the (real or complex, resp.) affine space[16] of (signed or complex, resp.) measures μ satisfying (12), and the affine space of (real or complex, resp.) PSRs feasible for $\mathscr{F}_K$.*

[16] As, strictly speaking, we "don't know yet" whether these spaces are empty or not, we are abusing terminology here, allowing the empty set as an affine space.

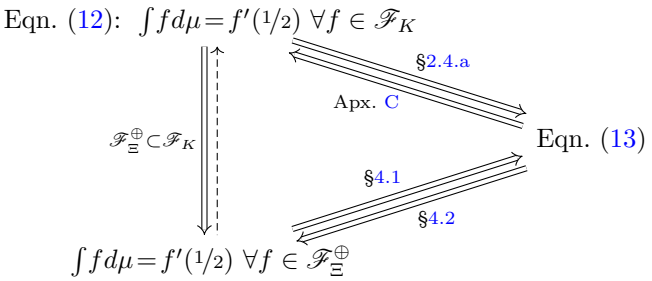


Figure 1: **Fourier-analytic characterization of derivative-computing measures.** This diagram shows the logical implications between (the versions of) the statements in Lemma 2.12: Eqn. (13), "$\hat{\mu}(\xi) = \dots$", and variants of Eqn. (12), "$\int f\, d\mu = f'(1/2)$". The labels on the arrows give the locations of the proofs.
The dashed arrow is implied by the north-east and north-west implications; it yields Prop. 2.10 (stated in §2.3.a, proof in §4.1). Lemma 2.12 as stated consists of the south-east and north-west implications: Although the south-east implication is implied by the north-east implication proven in Section 4 together with the south implication, its presentation serves as motivation in the technical overview, Section 2.

2.4.a The space of PSRs

We now come to the characterization of the set of PSRs feasible for $\mathscr{F}_K$, for any $K > 0$, and for $\mathscr{F}^{\oplus}_{\Xi}$ for any frequency set Ξ: We mentioned in Prop. 2.10 that these sets are the same, and we will now develop the machinery to prove it. By what we just summarized, the task is to characterize the derivative-computing measures, i.e., the finite measures satisfying (12).

The following lemma is the infinite-support extension of the corresponding statement on finite support [18, Eqn. (8)].

Lemma 2.12 (Fourier-analytic characterization of derivative-computing measures)**.** *Let μ be a finite measure on $\mathbb{R}$. For (12) to hold, it is necessary and sufficient that*

$$\text{for all } \xi \in [-1/2, +1/2]\colon \qquad \hat{\mu}(\xi) = -2\pi i \xi e^{-i\pi\xi}. \tag{13}$$

Recall that the Fourier transform of a tempered distribution which arises from integrating against a complex measure coincides with integrating against the Fourier-Stieltjes transform of the measure. As the Fourier-Stieltjes transform of a complex measure is a continuous function, evaluating $\hat{\mu}$ at individual frequencies ξ is well defined.

Let us consider the necessity of the condition.

Proof that (12) implies (13). For all $\xi \in [-1/2, +1/2]$, the real and imaginary parts of the function $x \mapsto e^{-2\pi i \xi x}$ lie in $\mathscr{F}_{1/2}$. Assume that μ is a complex measure satisfying (12). By the complex linearity of both the derivative at $1/2$, $f \mapsto f'(1/2)$, and of the integral $f \mapsto \int f\, d\mu$, by applying both to the real- and imaginary parts of $x \mapsto e^{-2\pi i \xi x}$, we find that

$$\begin{aligned} -2\pi i \xi e^{-i\pi\xi} &= (e^{-2\pi i \xi\, \square})'(1/2) = (\cos(-2\pi i \xi\, \square))'(1/2) + i\,(\sin(-2\pi i \xi\, \square))'(1/2) \\ &= \int \cos(-2\pi i \xi\, \square)\, d\mu + i \int \sin(-2\pi i \xi\, \square)\, d\mu = \int e^{-2\pi i \xi\, \square}\, d\mu = \hat{\mu}(\xi). \end{aligned}$$

This proves the condition (13), demonstrating its necessity in Lemma 2.12. □

As for the proof of the sufficiency-direction, it is based on Fourier Inversion. Performing that rigorously for only $\mathscr{F}^{\oplus}_{*}$ is less technical than the general case as stated in the lemma, which requires tempered distribution arguments and Paley-Wiener-Schwartz theory. As, in view of Corollary 2.9, the general case is mostly of academic interest, in §4.2, we prove the lemma only for the special case of $\mathscr{F}^{\oplus}_{*}$, and demote the general case, $\mathscr{F}_{*}$, to Appendix C.

As the proofs of the versions of Lemma 2.12 are somewhat spread out, we refer to Fig. 1 for a visual guide.

Revisiting §2.3.a, suppose $\Xi \subset [-1/2, 1/2]$ is a frequency set with $\max \Xi = 1/2$. In Lemma 4.1 of §4.1 we will prove that (13) is already implied by requiring that the equation $g'(1/2) = \int g \, d\mu$ in (12) holds only for $g \in \mathscr{F}_\Xi^\oplus$ — instead of for all $g \in \mathscr{F}_{1/2}$. This will prove Prop. 2.10 together with the reflection on $1/2$, translation, and dilation techniques laid out in §2.4 above.

2.4.b "Quantitative" view

Using Lemma 2.12, we are now ready to make a "quantitative" statement about the set of measures μ satisfying (12). At this point, as it were, we "don't know yet" whether a single such μ exists, but if a single one exists, the set is large.

Proposition 2.13 (Space of feasible PSRs)**.** *The (real) affine space of signed measures μ satisfying (13) is either empty or has infinite dimension.*

In §4.3, the proof of the proposition proceeds by adding to a single such signed measure infinitely many linearly independent signed measures whose Fourier transform vanishes on $[-1/2, +1/2]$, and invoking Lemma 2.12.

2.5 Feasibility of the Nyquist shift rules

It's about time to discuss the feasibility of the Nyquist shift rules ϕ_K, $K > 0$, from (5), for $\mathscr{F}_K$; throughout this section, ϕ_* will be as in (5).

As indicated in the introduction, we will not attempt to pursue a strategy based on the Shannon-Whittaker Interpolation Theorem. Instead, we apply Lemma 2.12 of the previous section for a suitable signed measure μ, which we will define in a moment. This μ will allow us to define $\phi_{1/2}$ via reflection, and dilation will give us all other ϕ_K, $K > 0$ (Prop. 2.11).

2.5.a The derivative-computing measure

With

$$u \colon \mathbb{Z} \to \mathbb{R} \colon n \mapsto \frac{(-1)^{n-1}}{\pi \, (n - 1/2)^2} \tag{14a}$$

we define the following signed measure on $\mathbb{R}$:

$$\mu := \sum_{n \in \mathbb{Z}} u(n) \delta_n, \tag{14b}$$

where convergence in the total-variation norm can easily be checked. One main technical piece of work in this paper is the following theorem.

Theorem 2.14. *If μ is as in (14), then the condition (13) of Lemma 2.12 holds.*

Applying Lemma 2.12 shows that the μ in (14) satisfies (12), i.e., integrating against it gives the derivative at $1/2$, provided that the integrand is a bounded smooth function with Fourier spectrum contained in $[-1/2, +1/2]$.

The proof of the theorem is in §4.4.

2.5.b From μ to ϕ_K

We now state the feasibility of the Nyquist shift rules; the purely technical proofs, based on the reflection and dilation, are in §5.2.

Corollary 2.15. *The Nyquist shift rule*

$$\phi_{1/2} \;=\; \sum_{a \in 1/2 + \mathbb{Z}} \frac{(-1)^{a+1/2}}{\pi a^2} \, \delta_a$$

(same as in (5)), is feasible for $\mathscr{F}_{1/2}$.

Corollary 2.16. *For each fixed real number $K > 0$, the Nyquist shift rule (5) is feasible for $\mathscr{F}_K$.*

With Prop. 2.5, we see that the Nyquist shift rule ϕ_K, $K > 0$, from (5), when convoluted against an expectation-value function f as in Def. 2.1 with $\mathrm{eig}_{\max}(A) - \mathrm{eig}_{\min}(A) \leq K$, gives the derivative of f.

2.6 Cost concept and optimality

So we have PSRs feasible[17] for $\mathscr{F}_K$, $K > 0$. But are there better ones? In this section, in answering that question in the negative, we understand the word "better" in a narrow technical sense to mean smaller norm. The norm of a PSR will turn out to be the worst-case standard deviation of the Single-Shot Estimator for it.

This section explains why we have somewhat emphasized real-valued PSRs: It can be seen that, as the functions in our spaces $\mathscr{F}_*$, $\mathscr{F}_*^{\oplus}$ are real-valued[18], the real part of any feasible complex PSR is a feasible PSR with smaller norm (cf. [18, Lemma 16]). Hence, there doesn't seem to be an advantage in allowing complex PSRs.

2.6.a The cost of a PSR

As in [18], we define the *cost* of a PSR ϕ simply as the (total-variation) norm of the measure, $\|\phi\|$. It is elementary that the norm coincides with the operator norm of the linear operator

$$f \mapsto f * \phi \colon \mathscr{M}_b \to \mathscr{M}_b$$

on the normed space of bounded, measurable real-valued functions of a real variable, $\mathscr{M}_b$, with the supremum norm $\|\,\square\,\|_\infty$; for convenient reference, we note two relevant inequalities in a remark.[19]

Remark 2.17. *For every real-valued, bounded, measurable function f and every signed measure ϕ, we have*

$$\int f \, d\phi \leq \|f\|_\infty \cdot \|\phi\|, \tag{15a}$$

and in particular,

$$\|f * \phi\|_\infty \leq \|f\|_\infty \cdot \|\phi\|. \tag{15b}$$

By applications of standard facts about regular measures (and continuous, differentiable, analytic functions), the statements remain true if we replace the qualification "bounded measurable" by "bounded continuous", or "bounded smooth" $\mathscr{C}_b^1$, or "bounded infinitely differentiable", or "bounded analytic".

In addition to the application-side interpretations of the operator norm as the worst-case factor by which noise (e.g., higher-frequency contributions to the input) is amplified when taking the convolution, [18] discusses two more motivations for referring to $\|\phi\|$ as the cost of the PSR ϕ, which make sense only for finite-support PSRs. Here, we discuss the following: The norm of the PSR is the standard deviation of a single-shot estimation algorithm for the convolution, in the worst case over all functions f and points x.

[17] Let it be repeated that, due to Prop. 2.10 in §2.3.a, there's no difference between feasibility for $\mathscr{F}_*$ and for $\mathscr{F}_*^{\oplus}$; we stick to $\mathscr{F}_*$ as it is the "simpler" space.

[18] So are, importantly, all expectation-value functions.

[19] Cf. (6) for the first one.

Algorithm: Single-Shot Estimator

Input : $x \in \mathbb{R}$
Plus: Access to *shot oracle* F

Output: Random number $E(x)$ with $\mathbb{E}\, E(x) = f * \phi(x)$

1 Sample a random point $A \in \mathbb{R}$ according to the probability measure $|\phi|/\|\phi\|$
2 Invoke the shot oracle for the point $x - A$ to obtain $F(x - A)$
3 **Return** $F(x - A) \cdot (d\phi/d|\phi|)(A) \cdot \|\phi\|$.

Figure 2: Algorithm **Single-Shot Estimator.** Simple unbiased estimator for the convolution with a real proper shift rule (PSR) ϕ. The expression $d\phi/d|\phi|$ stands for the Radon-Nikodym derivative of ϕ wrt $|\phi|$; note that $d\phi/d|\phi|(A) \in \{\pm 1\}$.

The Single-Shot Estimator. As we want the convenience of talking about function spaces, we introduce a theoretical computer science concept that replaces a "shot", i.e., a single measurement of the observable M from Def. 2.1 in the parameterized quantum state:

Definition 2.18 (Shot oracle)**.** A *shot oracle*, F, for a function $f\colon \mathbb{R} \to \mathbb{R}$ takes as input $s \in \mathbb{R}$ and returns a random number $F(s) \in \mathbb{R}$ such that $\mathbb{E}\, F(s) = f(s)$, with the requirement that runs of the shot oracle (for same or different inputs) return independent results.

Fig. 2 defines the Single-Shot Estimator.

Apart from abstracting from quantum evolutions and measurements, there is absolutely nothing novel here: The Single-Shot Estimator is really the direct generalization of the single shot estimator for finite-support shift rules.

Variance of the Single-Shot Estimator. For analyzing the (worst-case) standard deviation of the Single-Shot Estimator, consider the following fact.

Proposition 2.19. *Let ϕ be a PSR, and assume a shot oracle*[20] *that returns values in $\{\pm 1\}$. With input $x \in \mathbb{R}$, the Single-Shot Estimator in Fig. 2 has variance $\|\phi\|^2 - |f * \phi(x)|^2$.*

Proof. We use the notations from Fig. 2. Suppose the Single-Shot Estimator receives $x \in \mathbb{R}$ as input. By our assumption that $F(\,\square\,)$ is ± 1-valued and since $d\phi/d|\phi|$ is a function taking ± 1-values $|\phi|$-almost everywhere, we have $E(x)^2 = \|\phi\|^2$, so that $\mathbb{E}\, E(x)^2 = \|\phi\|^2$.

As advertised in Fig. 2, the Single-Shot Estimator is indeed an unbiased estimator for $f * \phi$, i.e, $\mathbb{E}\, E(x) = f * \phi(x)$, and we conclude that $\mathbb{V}\mathrm{ar}\, E(x) = \mathbb{E}\, E(x)^2 - (\mathbb{E}\, E(x))^2 = \|\phi\|^2 - (f * \phi(x))^2$ as claimed. □

For a feasible PSR, the variance is $\|\phi\|^2 - f'(x)^2$. Making a worst-case assumption on f and x, we take $f'(x) = 0$ — barren plateaus [20, 21] raising their flat heads again. In other words, we upper-bound the variance by $\|\phi\|^2$. This worst-case upper bound leads to the worst-case standard deviation of $\|\phi\|$ for the Single-Shot Estimator.

2.6.b Review of weak duality

While much of the convex optimization theory in [18] breaks down when moving from the finite dimensional vector spaces $\mathscr{K}_\Xi$ there to our infinite dimensional spaces $\mathscr{F}_K$, and moving from finite-support PSRs to infinite-support ones, the *Weak Duality Theorem*, [18, Prop. 6] goes through letter for letter. We give here a version of it that is tailored to our needs; the proof is in §6.1.

[20] In terms of Def. 2.1, this corresponds to the situation where the observable, M, has ± 1 eigenvalues only, e.g., a multi-qubit Pauli or so.

Proposition 2.20 (Weak duality theorem)**.** *Let $\mathscr{G}$ be a vector space of real-valued, bounded smooth functions, which satisfies $f(-\square) \in \mathscr{G}$ if $f \in \mathscr{G}$.*

If ϕ is a PSR feasible for $\mathscr{G}$, and $f \in \mathscr{G}$ with $\|f\|_\infty = 1$, then

$$-f'(0) \leq \|\phi\|. \tag{16}$$

Under the antecedent condition, equality holds in (16) if, and only if, inequality (15a) in Remark 2.17 is satisfied with equality for f, ϕ.

The spaces $\mathscr{F}_*$ and $\mathscr{F}_*^\oplus$ satisfy the condition on $\mathscr{G}$ in Prop. 2.20: By the usual tempered-distribution calculations we have for all tempered distributions f, $\mathrm{Supp}(f(-\square))^\wedge = -\mathrm{Supp}\,\hat{f}$, so that $f(-\square) \in \mathscr{F}_K$ if $f \in \mathscr{F}_K$.

2.6.c Optimality of the PSR

Now we are ready to prove that, in terms of our cost concept, the Nyquist shift rules (5) are optimal for $\mathscr{F}_K$.

Theorem 2.21. *Let K be a positive real number. The Nyquist shift rule (5) has smallest norm among all PSRs feasible for $\mathscr{F}_K$. The norm is $2\pi K$.*

The proof is in §6.2; it comprises guessing a function f and a calculating equality of the two sides in the Weak Duality Theorem Prop. 2.20.

2.6.d Other optimal feasible PSRs

In Section 9, the question will arise whether we can replace ϕ_K as defined in (5) by another optimal feasible PSR — maybe one whose support is more concentrated near 0. Here we note the following consequence of (the proof of) Theorem 2.21 together with Prop. 2.20.

Corollary 2.22. *Fix $K > 0$. If ϕ is a PSR feasible for $\mathscr{F}_K$, then ϕ has smallest norm (i.e., is optimal) if, and only if, $\phi = -\phi^0 + \phi^1$ where for $i = 0, 1$, ϕ^i is the restriction of $|\phi|$ to the set*

$$\big\{(1/2 + m)/2K \;\big|\; m \equiv i \bmod 2\big\} \tag{17}$$

This means that for all $K > 0$, a PSR feasible for $\mathscr{F}_K$ is optimal for $\mathscr{F}_K$ if, and only if, its support is contained in the set of numbers of the form $(1/2 + m)/2K$ where $m \in \mathbb{Z}$, and the signs alternate in the right way.

The proof of the corollary is sketched in §6.3.

2.7 Truncation

As discussed in the introduction, the Nyquist shift rules (5) require to apply the perturbed-parametric unitary from Def. 2.1 for arbitrarily large values of the parameter x in (8), which may not be physically desirable, or even possible. In this section, we discuss ways to truncate our PSRs to a bounded set of shifts, at the expense of introducing a (hopefully small) approximation error in the derivative. The next section §2.8 pursues the same goal using a technique that we call Folding. A third way to keep the values of the parameter small is described and used in the section on numerical simulations, §2.10.

Preparations. We start by discussing the approximation error.

Definition 2.23 (Near feasibility)**.** For a space of real-valued, bounded, smooth functions $\mathscr{G}$, and $\varepsilon \geq 0$, let us say that a PSR ϕ is *ε-nearly feasible for $\mathscr{G}$,* if we have

$$\|f * \phi - f'\|_\infty \leq \varepsilon \|f\|_\infty \quad \text{for all } f \in \mathscr{G}.$$

Feasibility is the same as being 0-nearly feasible, of course.

While convenient for our calculations (see the next lemma), the example of the symmetric difference quotients shows the limitations of this definition. The approximation guarantee is well-known: For at least 3-times differentiable functions f, we have (for variable ε)

$$\left| f * \big((\delta_{-\varepsilon} - \delta_{+\varepsilon})/2\varepsilon\big)(x) - f'(x) \right| = \varepsilon^2 |f'''(x)| + o_{\varepsilon\to 0}(\varepsilon^3);$$

the norm (cost) can readily be computed:

$$\big\|(\delta_{-\varepsilon} - \delta_{+\varepsilon})/2\varepsilon\big\| = 1/\varepsilon.$$

(It has been observed before (e.g., [13]) that the standard deviation of estimating based on this PSR is $\Omega(1/\varepsilon)$.) We see that improvement in approximation error is bought at the cost of increasing standard deviation of the estimator: As the sampling complexity (i.e., number of samples to reach a precision) grows quadratic in the standard deviation, cost and benefit are of equal order.

However, according to our Def. 2.23, the symmetric difference quotient is only $O(1)$-nearly feasible — which erroneously[21] suggests that it is completely useless.

The following lemma shows how near-feasibility is being used. Its proof explains why we get $\|f\|_\infty$ on the RHS.

Lemma 2.24. *Let $\phi, \tilde\phi$ be PSRs, and $\mathscr{G}$ a space of smooth, bounded, real-valued functions.*
If ϕ is feasible for $\mathscr{G}$, then $\tilde\phi$ is $\|\tilde\phi - \phi\|$-nearly feasible for $\mathscr{G}$.

Proof. As ϕ is feasible for $\mathscr{G}$, we can replace $f' = f * \phi$, use the bilinearity of convolution, and then use the inequality (15b) of Remark 2.17:

$$\|f * \tilde\phi - f'\|_\infty = \|f * (\tilde\phi - \phi)\|_\infty \le \|f\|_\infty \cdot \|\tilde\phi - \phi\|.$$

This concludes the proof. □

Truncation. Now we can discuss what happens to our ϕ_K, $K > 0$, when we truncate them. In the next proposition, no effort has been made to optimize the constant in front of the $1/N$; the proof is in §7.1.

Proposition 2.25. *Let $K > 0$, let N be a positive integer, and denote by $\phi_K^{(N)}$ the measure defined by truncating the sum defining ϕ_K in (5) to the $2N$ terms centered around 0, i.e.,*

$$\phi_K^{(N)} := 2K \cdot \sum_{\substack{a\in 1/2+\mathbb{Z} \\ -N+1/2\le a\le N-1/2}} \frac{(-1)^{a+1/2}}{\pi a^2}\, \delta_{a/2K}. \tag{18}$$

The PSR $\phi_K^{(N)}$ is ε-nearly feasible for $\mathscr{F}_K$, where $\varepsilon := \frac{4K}{\pi(N-1/2)}$.

As discussed in the introduction, usually, a small error in the derivative is acceptable. Setting $N := \lceil 4K/\varepsilon\pi + 1/2 \rceil$, Prop. 2.25 shows us that we can press the error to below ε by using shifts of magnitudes less than $2/\varepsilon\pi + 1/2K$.

2.8 Folding

For truncating a PSR, we have taken an interval and treated specially all shifts that didn't fall into that interval: We simply discarded them. We discarded shifts, not parameter values: With a truncation interval $[-\Delta, \Delta]$, for the derivative at a point $x \in \mathbb{R}$, the parameter values at which the expectation-value function is queried to approximate the derivative at x were in $[x - \Delta, x + \Delta]$.

[21] Cf. the non-existence result in §2.9.d, though: There, a much weaker concept of "ε-nearly feasible" is used (Corollary 2.34) which includes difference quotients.

For folding a PSR, we can also take an interval, and treat specially either shifts (in shift folding) or parameter values (in parameter folding) that don't fall into that interval. Instead of simply discarding, though, we will replace the shift/parameter value by one that falls into the interval.

It will make sense, in this subsection, to switch our attention from $\mathscr{F}_*$ to $\mathscr{F}_\Xi^\oplus := \mathscr{K}_\Xi + \mathscr{F}^\downarrow_{\max \Xi}$, i.e., we always think of the expectation-value function f as being decomposed into $f = f_1 + f_0$, as in Corollary 2.9.

2.8.a Folding with zero perturbation, i.e., $[A, B] = 0$

We start by clarifying the relationship between this paper's Nyquist shift rule (5) which has infinite support and is feasible for $\mathscr{F}_\Xi^\oplus$, and the known PSR which has finite support and is feasible [17] for $\mathscr{K}_\Xi$.

For every positive integer L, there exists a finite-support PSR feasible for $\mathscr{K}_{\{-L,\dots,L\}}$ [17]:

$$\begin{aligned} &\frac{1}{2L}\sum_{j=-L}^{L-1} u_j\, \delta_{(j+1/2)/2L}, \qquad \text{where} \\ u_j &:= \frac{\pi\,(-1)^{j+1}}{\sin^2(\pi\,(j+1/2)/2L)} \quad \text{for } j = -L,\dots,L-1; \end{aligned} \tag{19}$$

the expression in the 2nd line is the first derivative of the modified Dirichlet kernel at $(j+1/2)/2L$. The following remark makes this compatible with the notation in this paper, by dilation.

Remark 2.26 ([17])**.** *Let K, ξ_1 be positive real numbers with the property that $L := K/\xi_1 \in \mathbb{N}$, and define*

$$\Xi := \Big\{ \ell \cdot \xi_1 \;\Big|\; \ell = -L, -L+1, \dots, L-1, L \Big\}.$$

The following is a PSR feasible for $\mathscr{K}_\Xi$ with support contained in[22] $[-1/2\xi_1, +1/2\xi_1[$ *and with norm*[23] $2\pi K$*:*

$$\psi := 2K \cdot \sum_{\substack{a \in 1/2+\mathbb{Z} \\ -L+1/2 \le a \le L-1/2}} \frac{\pi}{(2L)^2} \cdot \frac{(-1)^{a+1/2}}{\sin^2(\pi a/2L)}\, \delta_{a/2K}.$$

For a positive real number p and every $x \in \mathbb{R}$ we will use the following notations:

$$\left.\begin{matrix} x \,\%\, p \\ x \,\%_+\, p \\ x \,\%_-\, p \end{matrix}\right\} := \text{ the unique } y \in \left\{\begin{matrix} [-p/2, +p/2[\\ [0, p[\\ ]-p, 0] \end{matrix}\right\} \text{ with } x = y \bmod p\mathbb{Z}. \tag{20}$$

(Recall the definition of $x = y \bmod p\mathbb{Z}$, which is: $y - x \in p\mathbb{Z}$.) With the notations of the previous remark, the mapping $x \to x \,\%\, (1/\xi_1)$ folds its argument into a fundamental region of periodicity of the functions in $\mathscr{K}_\Xi$.

For a fixed real number $p > 0$, by *shift-folding* a PSR *with mod-p*, we mean taking the image of the measure under the mapping $(\square \,\%\, p)\colon \mathbb{R} \to [-p/2, +p/2[$. We will give a general definition of shift folding in Def. 2.28 below; here we are only interested in taking the remainder mod-p. With the notations of Remark 2.26, by shift-folding with mod p where $p = 1/\xi_1$, we compile the measure of every real number onto the corresponding point in the fundamental region $[-1/2\xi_1, +1/2\xi_1[$ of periodicity of the function space $\mathscr{K}_\Xi$.

Now we are ready to establish the relationship between the infinite-support Nyquist shift rules (5) feasible for $\mathscr{F}_\Xi^\oplus$ and known finite-support PSRs feasible for $\mathscr{K}_\Xi$: With the notation of Remark 2.26, if we take a Nyquist shift rule and shift-fold with mod-$1/\xi_1$, then we obtain the known shift rules feasible [17] and optimal [18, Theorem 18] for $\mathscr{K}_\Xi$.

[22] Note that the functions in $\mathscr{K}_\Xi$ are $1/\xi_1$-periodic.

[23] The norm is discussed in [18], but it also follows from Theorem 2.27 below.

Theorem 2.27. *For all K, ξ_1, ψ as in Remark 2.26, and with ϕ_K the Nyquist shift rule (5), we have $(\square \,\%\, (1/\xi_1))(\phi_K) = \psi$.*

The proof in §7.2.a is based on partial fraction decomposition of $1/\sin^2$.

This theorem is relevant not only to make the point that, just as the PSRs ψ from Remark 2.26 are the right[24] PSRs for $\mathscr{K}_\Xi$, our Nyquist shift rules (5) are, mathematically speaking, the right PSRs for $\mathscr{F}_\Xi^\oplus$. The theorem also motivates our general concept of folding, as defined in the next subsection, §2.8.b.

We have taken the approach here to present the special case where Ξ consists of equi-spaced frequencies with no gaps. It should be clear to the attentive reader that Theorem 2.27 applies more generally in the case of frequency sets Ξ that have a common divisor[25]. The frequency set Ξ having a common divisor is equivalent to the functions in $\mathscr{K}_\Xi$ being periodic (the periods are the reciprocals of the common divisors). In [18] we prove that in the case of frequency sets with gaps, the finite-support shift rule ψ from Remark 2.26 is still feasible and optimal, but there is a wealth of optimal finite-support shift rules, notably some with smaller support.

2.8.b General definitions for shift and parameter folding

We treat folding abstractly. Here's the definition.

Definition 2.28 (Folding function)**.** For a positive real number p, we call a measurable function $\tau\colon \mathbb{R} \to \mathbb{R}$ a *folding function* or a *p-folding,* if τ is equal to the identity modulo $p\mathbb{Z}$, i.e.,

$$\text{for all } s \in \mathbb{R}\colon \quad \tau(s) = s \bmod p\mathbb{Z};$$

in remainder calculus: $\tau(s) \,\%\, p = s \,\%\, p$ for all s.

The intuition behind a folding function is that in the decomposition $f = f_1 + f_0$ from Corollary 2.9, the folding function with $p := 1/\xi_1$ (notations from Remark 2.26) banks on the p-periodicity of the f_1-term to recover the known shift rules [17] for f_1. It ignores the f_0-term, introducing an approximation error. The hope is that as f_0 decays towards $\pm\infty$, the approximation error introduced through the folding can be made small. That hope can be brought to fruition, as the next subsection shows.

The formal definitions are in the following Lemma 2.29: we refer to Item (a) as *shift folding,* and to Item (b) as *parameter folding.* Fig. 3 shows the estimator for parameter-folded PSRs that corresponds to the Single-Shot Estimator for unfolded ones, Fig. 2.

Lemma 2.29 (Fundamental folding-lemma)**.** *Let Ξ be a frequency set, and $p > 0$ a real number such that $1/p$ is a common divisor of Ξ. Let ϕ be a PSR and τ a p-folding.*

If ϕ is feasible for $\mathscr{K}_\Xi$, then the following hold.

(a) The PSR $\tau(\phi)$ is feasible for $\mathscr{K}_\Xi$, i.e., for all $f \in \mathscr{K}_\Xi$ and all $x \in \mathbb{R}$ we have

$$f'(x) = \int f(x - \tau(s))\, d\phi(s);$$

and

(b) For all $f \in \mathscr{K}_\Xi$ we have $f' = (f \circ \tau) * \phi$, i.e., for all $x \in \mathbb{R}$,

$$f'(x) = \int f(\tau(x - s))\, d\phi(s).$$

The proof of Lemma 2.29 is given in §7.2.b, as it illuminates the relationships between the technical aspects of Def. 2.28 and shift/parameter folding.

[24] Case of Ξ's without gaps, otherwise see [18].

[25] A *common divisor* of a set $S \subset \mathbb{R}$ is a positive real number γ_1 such that $S \subset \gamma_1\mathbb{Z}$.

Algorithm: Simple Folding Estimator

Input : $x \in \mathbb{R}$
Plus: Access to *shot oracle* F.

Output: Random number $E(x)$ with $\mathbb{E}\, E(x) = (f \circ \tau) * \phi$

1 Sample a random point $A \in \mathbb{R}$ according to the probability measure $|\phi|/\|\phi\|$
2 Apply the folding function to obtain $B := \tau(x - A)$
3 Invoke the shot oracle for the point B to obtain $F(B)$
4 **Return** $F(B) \cdot (d\phi/d|\phi|)(A) \cdot \|\phi\|$.

Figure 3: Algorithm **Simple Folding Estimator.** Simple unbiased estimator for the parameter-folded PSR ϕ with folding function τ. Cf. Fig. 2.

2.8.c An example of parameter folding with quadratic decay in approximation error

There are many ways to chose the folding function. Here we present one of them, for which we can prove its effectiveness in parameter folding using our Fourier-decomposition toolkit. The definition of the folding function is inside the following lemma. (The proof is mere arithmetic and therefore omitted.) We use the %-notations defined in (20) above.

Lemma 2.30. *For positive real numbers p, c satisfying $p|c$ (i.e., $c/p \in \mathbb{N}$), define the following function:*

$$\tau_{p,c} \colon \mathbb{R} \to \mathbb{R} \colon s \mapsto \begin{cases} -c + s \,\%_{-}\, p, & \text{if } s \leq -c - p; \\ s, & \text{if } s \in \,]-c-p, +c+p[\,; \\ +c + s \,\%_{+}\, p, & \text{if } s \geq +c + p. \end{cases} \tag{21}$$

The function $\tau_{p,c}$ is a p-folding function with image $]-c-p, +c+p[$. □

For the remainder of this section, and in §7.2.c (which contains the proofs), we use the following big-O notation: We write $g(c) = O(h(c))$ for functions $g = g(c)$, $h = h(c) \geq 0$, to indicate the existence of an absolute constant C such that $|g(c)| \leq C \cdot h(c)$ for all allowed c; if the constant is not absolute but depends on, say, "K", we write $O_K(\boldsymbol{.})$. We also use the corresponding big-Ω notation.

Corollary 2.9 motivates the conditions in the following proposition, which quantifies the approximation error, operator norm (i.e., cost), and maximum parameter values.

Proposition 2.31. *Let Ξ be a frequency set, p a positive real number such that $1/p$ is a common divisor of Ξ, and c a positive real number with $p|c$; let $\tau_{p,c}$ as in Lemma 2.30.*

The Nyquist shift rule ϕ_K parameter-folded by $\tau_{p,c}$ as in Lemma 2.29(b),

$$f \mapsto (f \circ \tau_{p,c}) * \phi_K \tag{22}$$

has the following properties.

(a) If $f = f_1 + f_0 \in \mathscr{F}_\Xi^\oplus$ where $f_1 \in \mathscr{K}_\Xi$ and f_0 decays linearly with constants (c, C) for some C, then

$$\left|(f \circ \tau_{p,c}) * \phi_K(x) - f'(x)\right| \leq \begin{cases} \dfrac{8KC}{\pi c(2Kc-1)}, & \text{for } x \in [-p, +p]; \\ \dfrac{8\pi KC}{c}, & \text{for all } x \in \mathbb{R}. \end{cases}$$

(b) The operator norm of the linear operator[26] $\mathscr{M}_b \to \mathscr{M}_b$ defined in (22) is smaller than the norm of ϕ_K, i.e., $2\pi K$ (cf. Theorem 2.21).

[26] On the space of measurable bounded real-valued functions with $\|\boldsymbol{.}\|_\infty$, cf. page 14.

(c) The maximum magnitude of a parameter value at which f is evaluated in (22) is less than $c+p$.

(d) In the Simple Folding Estimator, Fig. 3, with input $x \in [-p,+p]$, the expected magnitude of parameter values at which the shot oracle is queried, is at most

$$|x| + \frac{\ln(K(c+2p)) + O(1)}{\pi^2 K}.$$

To parse the expressions involving products of p or c with K (in denominators or under the logarithm), note that

$$cK \geq pK \geq 1. \tag{23}$$

The second inequality follows from the fact that $1/p$ divides the elements of Ξ one of which is K, and $K > 0$ holds as Ξ is a frequency set (Def. 2.4).

To summarize, first note that the expression on the RHS in Item (a) in the case "$x \in [-p,+p]$" is $O_{K,C}(1/c^2)$, i.e., we have quadratic decay of the approximation error in terms of the largest parameter value queried.

As for the value of p, in the situation of Def. 2.1, a greatest common divisor of the difference set of the spectrum of A would be expected to be known, and a smaller common divisor can be chosen to cover a larger interval in the order-$1/c^2$ case of Item (a). The number c is the main quantity to play with: In the interval $[-p,+p]$, the approximation error goes down quadratically in c, while the maximum magnitude of a parameter value increases linearly, and the expected magnitude of a parameter value increases only logarithmically. Note that increasing c might allow to decrease C, as for each fixed f, C decays linearly as $c \to \infty$.

The strict condition "$x \in [-p,+p]$" in Item (a) is merely to make the proof less onerous: The interested reader will, upon inspection of the proof in §7.2.c, realize that the quadratic decay of the approximation error holds if $|x| \leq p + (1-\Omega(1))c$, and hence in particular for $|x| = o(c)$. Generally speaking, it should be understood that the approximation error decreases as $|x| \ll c+p$ and increases as $|x|$ approaches $c+p$. A look at the proof reveals that a more fine grained analysis would have in the denominator the term $c+p-|x|$ instead of $2Kc$, provided that $|x| \leq c+p-\Omega_K(1)$; this would lead to decays between $(1/c)^1$ and $(1/c)^2$ in that region of x's.

2.9 Non-existence results

In this section, we briefly discuss a few results of non-existence of the optimist's feasible PSRs, e.g., those with compact support. As non-existence is so frustrating, we won't spend too much time on it; the proofs are mostly only sketched.

Before we start, two things must be emphasized.

Firstly, the non-existence results are only for *feasible* (as opposed to approximate) PSRs: As we already know from the discussions about truncation in §2.7, for every $\varepsilon > 0$ there exists an ε-almost-feasible PSR with compact support (and decent norm/cost). Having said that, note that the size of the support in these examples grows beyond all bounds as ε tends towards 0; we will prove below that that is unavoidable.

Secondly, recalling the discussion in §2.3.a, the reader should be aware: While the results of this section pertain to feasibility of PSRs for $\mathscr{F}_{1/2}$, with proofs making use of all frequencies $\xi \in [-1/2,+1/2]$, by Prop. 2.10, non-existence of the described types of PSRs feasible for $\mathscr{F}_\Xi^\oplus$ is implied for all frequency sets Ξ with $\max \Xi = 1/2$. (And from there, of course, by dilation, for all frequency sets.)

We start with some observations involving slight modifications of our Nyquist shift rules.

2.9.a Other optimal PSRs

An inspection of the proof of Theorem 2.14 in §4.4 shows that, for all $K > 0$, the Nyquist shift rule ϕ_K is the unique complex PSR that is feasible for $\mathscr{F}_K$ and whose support is contained in the

support of ϕ_K. Indeed, the function u in (14) is the inverse periodic Fourier transform of the RHS of equation (13) from Lemma 2.12. (The Fourier Inversion Theorem applies as both u and $\hat{u}$ are absolutely integrable.)

In particular, there is no other complex PSR that is *optimal* for $\mathscr{F}_K$, as, by Corollary 2.22, the support of such a PSR would have to be contained in the support of ϕ_K.

2.9.b PSRs with compact support

By standard Paley-Wiener theory, a complex measure μ with compact support will have a Fourier transform that is a holomorphic function on a connected domain containing the real line. From standard complex variable function theory we know that two holomorphic functions on a connected domain must be equal if they coincide on a set that contains an accumulation point. This means that Lemma 2.12 restricts the choices for $\hat{\mu}$ to only a single one:

$$\hat{\mu} = \left(\mathbb{R} \to \mathbb{C} \colon \xi \mapsto \hat{\mu}(\xi) = -2\pi i \xi \, e^{-i\pi\xi}\right),$$

so that μ is the tempered distribution "for every Schwartz function, take the derivative at $1/2$". But this is a contradiction, as that particular tempered distribution is not a finite measure.[27]

Hence, there is no compactly supported complex PSR feasible for $\mathscr{F}_*$.

2.9.c Exponential concentration

Now, using nothing but standard tools, we show that complex PSRs with exponential concentration cannot exist.

Definition 2.32 (Exponential concentration)**.** We say that a complex measure μ on the real line is *exponentially concentrated* if there exist $C, r > 0$ such that

$$|\mu|(\mathbb{R} \setminus [-T, +T]) \leq C \cdot e^{-rT} \quad \text{for all } T \geq 0. \tag{24}$$

With a similar argument as in the case of compact support, we will prove the following in §8.1.

Theorem 2.33. *Let $\Xi \subseteq [-1/2, +1/2]$ be an infinite set. There exists no complex PSR ϕ with exponential concentration that is feasible for the function space $\mathscr{G}$ spanned by the functions* $\cos(2\pi\xi\,\square)$, $\sin(2\pi\xi\,\square)$, $\xi \in \Xi$.

We note that for $\Xi := [-1/2, +1/2]$, the space $\mathscr{F}_{1/2}$ contains the space $\mathscr{G}$ in Theorem 2.33, and hence no exponentially concentrated PSR is feasible for $\mathscr{F}_{1/2}$.

2.9.d Exponentially concentrated nearly feasible PSRs

We can also give impossibility results for complex PSRs with non-zero approximation error.

For this context, the concept of Nearly Feasible from Def. 2.23 is too limiting. Instead we simply take convergence of the approximations of the derivative in the point 0 to the correct value. The proof is sketched in 8.2.

Corollary 2.34. *Let Ξ and $\mathscr{G}$ be as in Theorem 2.33.*

The following does not exist: Numbers $C, r > 0$ and a norm-bounded family ϕ_j, $j \in \mathbb{N}$, of complex PSRs all satisfying (24) from the definition of exponential concentration (all with the given r), such that $f * \phi_j(0) \xrightarrow{j\to\infty} f'(0)$ for every $f \in \mathscr{G}$.

Note that the boundedness condition, $\|\phi_j\| = O(1)$ $(j \to \infty)$ is needed, otherwise the symmetric difference quotient would be a counterexample.

[27] The equality $[\varphi \mapsto \partial_{1/2}\varphi] = \int \square \, d\mu$ for a finite measure μ would imply, for all positive integers k, $2\pi k = \int \sin(2\pi k(x - 1/2)) \, d\mu(x)$, which would contradict the finiteness of the measure.

2.10 Numerical simulation

To make an attempt at understanding the practical utility of the Nyquist shift rule vs Banchi-Crooks's method, the author has designed a small *Pluto*[28]-notebook containing code in the mathematical computation programming language *Julia*[29], which simulates the two methods numerically and allows us to produce colorful pictures.

The *Approximate Stochastic Parameter Shift Rule* (ASPSR) has been implemented as in [13], accommodating our idiosyncratic choice of $\hbar = 1/2\pi$. It is compared to a version of the truncated Nyquist shift rule (STNySR), tailored especially for the comparison to Banchi-Crooks: A positive real number T can be provided by the user which has the effect of limiting to the interval $[-T, +T]$ all parameter values in queries to the expectation-value function.

As the stochastic properties of the two estimators are identical, the numerical simulation ignores that aspect. It allows to query

- In the case of the STNySR: Directly the expectation-value function as in (2.1) at a given parameter value $x \in \mathbb{R}$;
- In the case of the ASPSR: For given s, x, ε, directly the expectation value

$$\mathrm{tr}\big(M\, U_{\pm}(s,x,\varepsilon)\varrho\, U_{\pm}(s,x,\varepsilon)^{\dagger}\big), \qquad \text{where}$$
$$U_{\pm}(s,x,\varepsilon) := e^{2\pi i s(xA+B)} e^{2\pi i\varepsilon(\pm A/8\varepsilon+B)} e^{2\pi i(1-s)(xA+B)}.$$

The truncation parameter, T, mentioned above is set to $1/8\varepsilon$, to ensure that both methods use parameter values in the same interval.

The *Julia* code produces random perturbed-parametric unitary instances as follows: A random matrix M with ± 1 eigenvalues; a random positive-semidefinite trace-1 matrix ϱ; a random matrix A with ± 1-eigenvalues; a random standard Gaussian Hermitian matrix B.

The Pluto notebook makes it is easy to make numerical simulations to compare absolute and relative errors between ASPSR and STNySR graphically in plots. While in the following we discuss some typical features and noteworthy behavior based on plots created with the code, the reader is invited to play with the notebook[10] and judge for her or himself.

Fig. 4, shows a single expectation-value function randomly constructed in that way, and the absolute errors of ASPSR and the STNySR. Typical features are that, *(1)* due to the cutoff at $T := 1/8\varepsilon$ of parameter values in the STNySR, near the ends of the interval $[-T, +T]$ the approximation errors explode; and *(2)* STNySR is better than ASPSR when there is sufficient gap between the parameter value x and the cut-off $\pm 1/8\varepsilon$.

As A has ± 1 eigenvalues we have $K = 2$, and hence the query points of the Nyquist shift rules are $1/4$ apart. Sub-figure (c) of Fig. 4 shows break points in the green STNySR-line, which are caused by changes in the set of shifts that are queried.

Fig. 5 shows the differences of relative errors between the two methods. Positive values mean STNySR is better. It can be seen (positive green 10th-percentile points) that in about 90% of the random instances, STNySR was better than the ASPSR, at least where the point x at which the derivative is requested is sufficiently far away from the cut-off $1/8\varepsilon = 12.5$.

As the query points of the Nyquist shift rules are $1/4$ apart, if $x > 12.25$, no query point of STNySR is to the right of x — leading to a noise-only "approximation" of the derivative.

It can be seen (positive blue median points) that for $x \leq 12$, i.e., when there are at least 3 query points to the right of x, STNySR gives at least as good an approximation of the derivative as ASPSR, in at least half of the cases. It should be noted that the mean (not plotted) lies above the median, indicating that the advantage is, on average, substantial.

For a considerable region of the parameter, STNySR is roughly at least as good ASPSR in 99% of the instances (magenta 1st percentile data points hugging zero).

[28] plutojl.org

[29] julialang.org

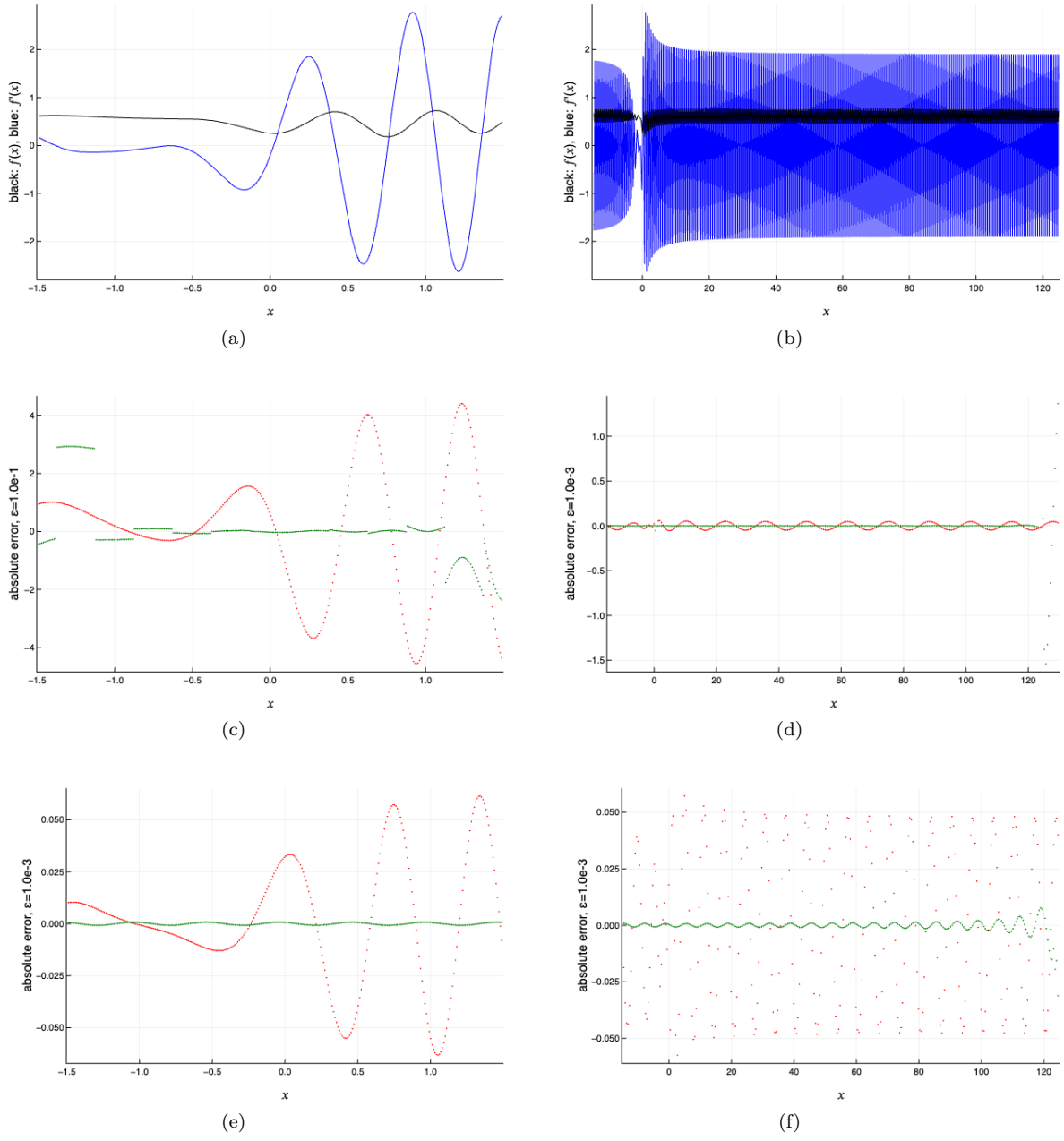


Figure 4: **Example of approximations provided by this paper's STNySR and BC's ASPSR.** Data from a randomly created (and arbitrarily selected) expectation-value function f is presented (with matrices of size 4×4). Graphs (a),(b) show f (black) and f' (blue), on different parameter intervals. In (c,d,e,f) absolute errors of the ASPSR (red) and the STNySR (green) are plotted. In (c) we took $\varepsilon = 10^{-1}$ (leading to the cut-off $T = 1.25$ for the STNySR) and in (d), (e), (f)) we took $\varepsilon = 10^{-3}$ (leading to the cut-off $T = 125$).

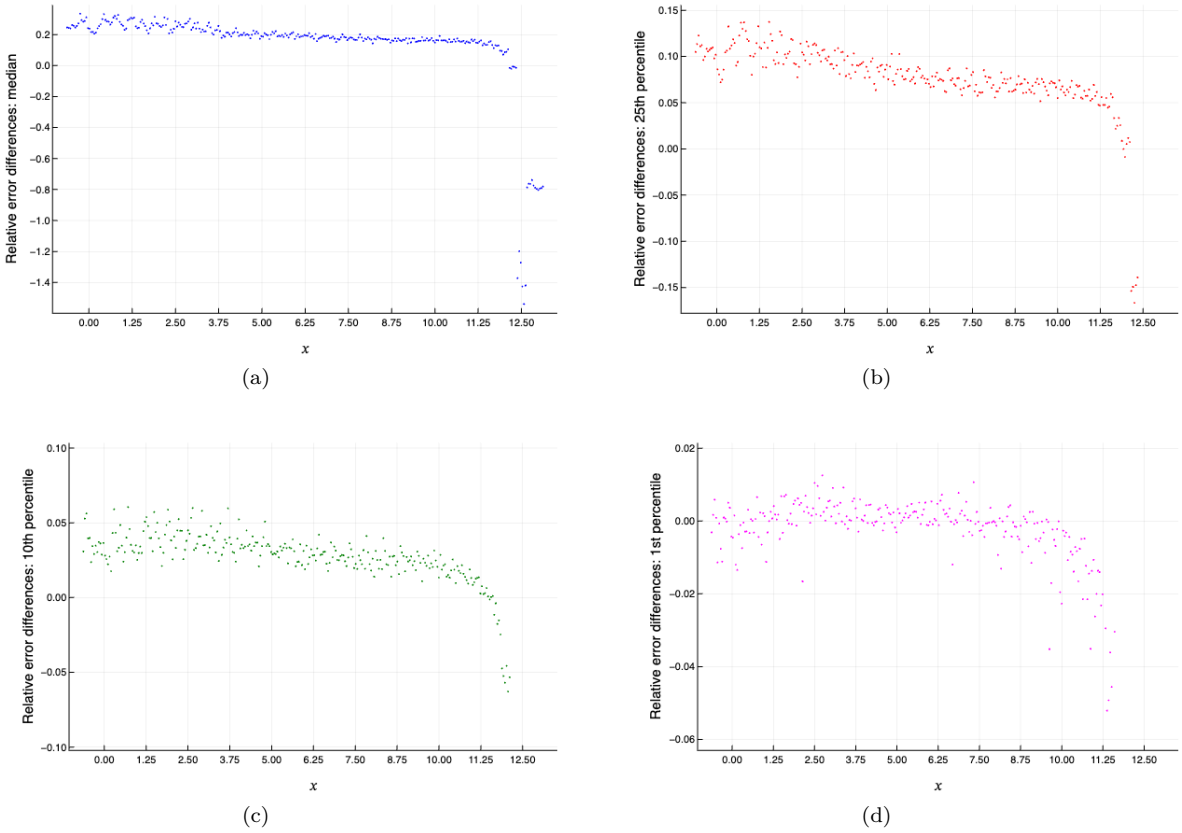


Figure 5: **Comparison of relative errors between STNySR and BC's ASPSR.** Plots of 200 random instances of expectation-value functions f chosen randomly as described in the text (with 8×8 matrices), considered at 300 parameter values, x. For each combination of parameter value x and expectation-value function f, the difference between the relative error of the ASPSR-approximation (with $\varepsilon = 10^{-2}$) of $f'(x)$ minus the relative error of the STNySR-approximation of $f'(x)$ (with parameter cut-off $1/8\varepsilon = 12.5$) is computed. For each parameter value x, the median (blue), 25th percentile (red), 10th percentile (green), 1st percentile (magenta) of the 200 random instances are plotted, in different scales.

Appendix D has plots of the results of more numerical simulations. The reader should understand that the presented numerical results are preliminary, and that more refined and extensive numerical simulations and statistical analysis are necessary for a comprehensive comparison of the methods. In particular, which (proper or not) shift rules are preferable in which parameter regions when the parameter values are constrained to an interval might be a topic of future research.

3 Fourier analysis of perturbed-parametric unitaries

In this section, we will deal with the Fourier-analytic properties of the functions in Def. 2.1: Perturbed-parametric unitary functions and expectation-value functions. We start by reviewing the (standard) notation we use.

3.1 Notations

Section 3 is somewhat demanding in terms of notations, as we have to deal with operator-valued tempered distributions.

3.1.a Tempered distributions

We denote the space of (complex-valued) Schwartz functions on $\mathbb{R}$ by $\mathcal{S}$, and the tempered distributions on $\mathbb{R}$ by $\mathcal{S}'$. We denote the duality between Schwartz functions (on the left) and tempered distributions (on the right) by "$\langle \centerdot : \centerdot \rangle$"; e.g., for a Borel measure μ of at most polynomial growth we have $\langle \varphi : \mu \rangle := \int \varphi(x)\, d\mu(x)$, and for a measurable function f of at most polynomial growth we have $\langle \varphi : f \rangle := \int \varphi(x) f(x)\, dx$.

As we are using angle-brackets "$\langle \centerdot : \centerdot \rangle$" for tempered distributions, we revert to parentheses "$(\centerdot \mid \centerdot)$" for the Hilbert-space inner product; it is *linear in the right argument*, anti-linear in the left argument.

3.1.b Operator-valued functions

The space of linear operators on a finite-dimensional Hilbert space $\mathcal{H}$ is $\mathbb{L}(\mathcal{H})$.

In all of Section 3, we denote by $\|\!|\!| \square |\!|\!|$ the operator / spectral norm (of operators on a Hilbert space or of square matrices, resp.). If T is a set and F is an operator-valued function defined on T, we let $\|\!|\!| F |\!|\!|_{\infty,T} := \sup_{x\in T} \|\!|\!| F(x) |\!|\!|$; for $T = \mathbb{R}$ we omit the subscript T on the norm.

For the concepts of linear decay and square integrability of operator-valued functions we use $\|\!|\!| \square |\!|\!|$ (although they are norm independent in finite dimension). Integration of operator-valued functions against complex-valued Schwartz functions and the Fourier transform are defined as usual: element-wise. This means, for example, the Fourier transform $\hat{F}$ of an operator-valued function F is defined through the condition: For all vectors ϕ, ψ in the underlying Hilbert space,

$$(\phi \mid \hat{F}\psi) = \Big(x \mapsto (\phi \mid F(x)\psi)\Big)^{\wedge}$$

We will also need a small addition to the "$\mathscr{K}_\Xi$" notation: Denoting the underlying Hilbert space by $\mathcal{H}$, for an arbitrary finite non-empty set $\Lambda \subset \mathbb{R}$, the complex vector space of bounded functions with values in $\mathbb{L}(\mathcal{H})$ whose Fourier spectrum is contained in Λ is denoted by

$$\mathscr{K}_\Lambda^{\mathrm{op}} = \Big\{ f \colon x \mapsto \sum_{\lambda\in\Lambda} e^{2\pi i \lambda x} B_\lambda \;\Big|\; B \in \mathbb{L}(\mathcal{H})^\Lambda \Big\}; \tag{25}$$

We emphasize that Λ is not required to be symmetric.

3.2 The Fourier-decomposition theorems

Now we are ready to formulate the Fourier-decomposition theorem.

Proposition 3.1 (Fourier-decomposition for perturbed-parametric unitaries)**.** *With the notations of Def. 2.1, let* $U\colon x \mapsto e^{2\pi i(xA+B)}$*, and* $\Lambda := \operatorname{Spec} A$*.*

There exists a unitary-operator-valued function $\tilde{U}$ *with the following properties.*

(a) $\tilde{U} \in \mathscr{K}_\Lambda^{op}$ *and* $U - \tilde{U}$ *is of linear decay.*
(b) If A*,* B *commute, then* $U = \tilde{U}$*.*
(c) For any finite-Fourier-spectrum, bounded, operator-valued function $\tilde{V}$*: If* $lim_{x\to\infty}(U(x) - \tilde{V}(x)) = 0$*, then* $\tilde{V} = \tilde{U}$*.*

Item (c) is a corollary to Item (a), and serves as a sanity check: It establishes that $\tilde{U}$ is the *unique* bounded, finite-Fourier-spectrum operator-valued function that approximates U for all large parameter values.

The proof of Prop 3.1 is eigenvalue perturbation theory in finite dimension; the details fill §3.3; now we derive Prop. 2.7 from Prop. 3.1.

3.2.a Proof of Prop. 2.7 from Prop. 3.1

We now derive the Fourier-decomposition theorem for perturbed-parametric expectation-value functions (§2.3) from Fourier-decomposition for perturbed-parametric unitaries.

Proof of Prop. 2.7. We simply expand the definition of the perturbed-parametric expectation-value function in Def. 2.1. With $D\colon x \mapsto U(x) - \tilde{U}(x)$, for all $x \in \mathbb{R}$,

$$\operatorname{tr}\Big(M\,U(x)\varrho U(x)^\dagger\Big) = \operatorname{tr}\Big(M\,(\tilde{U}(x)+D(x))\varrho(\tilde{U}(x)^\dagger + D(x)^\dagger)\Big) =$$
$$\operatorname{tr}\Big(M\,\tilde{U}(x)\varrho\tilde{U}(x)^\dagger\Big) + \operatorname{tr}\Big(M\,\tilde{U}(x)\varrho D(x)^\dagger\Big) + \operatorname{tr}\Big(M\,D(x)\varrho\tilde{U}(x)^\dagger\Big) + \operatorname{tr}\Big(M\,D(x)\varrho D(x)^\dagger\Big).$$

We set $f_1 := \operatorname{tr}\Big(M\,\tilde{U}(\,\square\,)\varrho\tilde{U}(\,\square\,)^\dagger\Big)$. This is a bounded function with Fourier spectrum contained in $\operatorname{Diff}\operatorname{Spec} A$ (e.g., [16, 17], or see the proof of Lemma 3.5 below). By standard estimates, we find that each of the remaining terms has linear decay. As sums of functions of linear decay have linear decay, this completes the proof of the proposition. □

3.3 Proof of Fourier-decomposition for unitaries

Notational conventions for §3.3.∗. This section is somewhat hard on the alphabet. For that reason, we adopt the following notation convention: We denote matrices by typewriter-font letters $\mathtt{A}, \mathtt{B}, \mathtt{C}, \ldots$. Moreover, for operators denoted by $A, B, U, \ldots$, when an orthonormal basis has been fixed, we denote the matrix corresponding to the operator by the typewriter version of the letter, e.g., $A \leftrightarrow \mathtt{A}$, $B \leftrightarrow \mathtt{B}$, ... We denote the identity matrix by $\mathtt{I}$.

The proof of Prop. 3.1 is by degenerate (i.e., eigenvalues are not necessarily simple) eigenvalue perturbation theory, as everybody has learned it in their quantum mechanics class. To enable a mathematically rigorous proof, Appendix B reviews the mathematical foundation, Rellich's theorem, based on which it also summarizes textbook perturbation theory in matrix notation (Corollary B.2), for convenient reference.

3.3.a Proof of Prop. 3.1 (a),(b)

This subsection holds the proof of Fourier-decomposition, Prop. 3.1. We make some definitions and we will work with them throughout the subsection.

We use the notations of Def. 2.1, and denote by d the dimension of the underlying Hilbert space. Define $U(x) := \exp(2\pi i(xA + B))$ for $x \in \mathbb{R}$.

We use the following matrix big-O notation: By $\mathfrak{O}^{\geq R}(1/x^2)$ we refer to an unnamed square-matrix valued function $x \mapsto \mathtt{M}(x)$ defined on $\mathbb{R} \setminus [-R, +R]$, with the property that

$$\sup_{x \in \mathbb{R} \setminus [-R,+R]} \left\| x^2 \mathtt{M}(x) \right\| \quad < \infty.$$

(Recall that $\|\cdot\|$ is the spectral norm, but the property is independent of the norm, as we are in finite dimension.)

Lemma 3.2. *We use the notations above. There exists an orthonormal basis of eigenvectors of A, and a bounded d-by-d matrix-valued function $\mathtt{C}$ on $\mathbb{R}$ with the following property.*

Denoting $\mathtt{A}, \mathtt{B}, \mathtt{U}$ the $(d \times d)$-matrices of the operators A, B, U, resp., wrt the basis, and setting

$$\widetilde{\mathtt{U}}(x) := \exp\bigl(2\pi i (x\mathtt{A} + \mathtt{Diag}(\mathtt{B}))\bigr)$$

we have for all x (with $|x| \geq 1$),

$$\mathtt{U}(x) = \widetilde{\mathtt{U}}(x) + \frac{\mathtt{C}(x)}{x} + \mathfrak{O}^{\geq 1}(1/x^2).$$

If A, B commute, then the basis can be chosen to consist of eigenvectors of B, so that $U = \tilde{U}$.

Proof. The remark about the case when A, B commute is trivial.

In the general case, we apply Corollary B.2 from Appendix B, and use the notation therein. In view of Items (f), (g) and (h) of Corollary B.2 we define

$$\begin{aligned} \mathtt{E}_{\mathtt{W}}(z) &:= \mathtt{W}(z) - \mathtt{I} - z\mathtt{W}^{[1]} \\ \mathtt{E}_{\Lambda}(z) &:= \Lambda(z) - \mathtt{A} - z\,\mathtt{Diag}(\mathtt{B}) - z^2\Lambda^{[2]}; \end{aligned}$$

and note that $z \mapsto \mathtt{E}_{\mathtt{W}}(z)/z^2$ and $z \mapsto \mathtt{E}_{\Lambda}(z)/z^3$ are bounded on[30] $[-r/2, +r/2]$, by Items (b) and (d), resp., of Corollary B.2.

We set $R := \max(1, 2/r)$ and calculate, for $|x| > R$,

$$\begin{aligned} \mathtt{U}(x) &= \exp(2\pi i(x\mathtt{A} + \mathtt{B})) \\ &= \exp(x 2\pi i(\mathtt{A} + \tfrac{1}{x}\mathtt{B})) \\ &= \mathtt{W}(1/x)\, \exp(x 2\pi i \Lambda(1/x))\, \mathtt{W}(1/x)^\dagger. \end{aligned} \tag{$*$}$$

From here, we treat the terms separately. We find:

$$\begin{aligned} \mathtt{W}(1/x) &= \mathtt{I} + \tfrac{1}{x}\mathtt{W}^{[1]} + \mathtt{E}_{\mathtt{W}}(1/x) \\ \mathtt{W}(1/x)^\dagger &= \mathtt{I} - \tfrac{1}{x}\mathtt{W}^{[1]} + \mathtt{E}_{\mathtt{W}}(1/x)^\dagger \end{aligned}$$

where the first equation is just the definition of $\mathtt{E}_{\mathtt{W}}$, and the second follows from the first using the fact the $\mathtt{W}^{[1]} = \mathtt{W}'(0)$ is anti-hermitian, by Corollary B.2(c), as $\mathtt{W}(0) = \mathtt{I}$.

For the middle factor in ($*$), we find

$$\begin{aligned} \exp(x 2\pi i \Lambda(1/x)) &= \exp\Bigl(x 2\pi i \bigl(\mathtt{A} + \tfrac{1}{x}\,\mathtt{Diag}(\mathtt{B}) + \tfrac{1}{x^2}\Lambda^{[2]} + \mathtt{E}_{\Lambda}(1/x)\bigr)\Bigr) \\ &= \exp\Bigl(2\pi i \bigl(x\mathtt{A} + \mathtt{Diag}(\mathtt{B}) + \tfrac{1}{x}\Lambda^{[2]} + x\mathtt{E}_{\Lambda}(1/x)\bigr)\Bigr) \\ &= \widetilde{\mathtt{U}}(x) \cdot \exp(2\pi i \Lambda^{[2]}/x) \cdot \exp(2\pi i (x\mathtt{E}_{\Lambda}(1/x)) \end{aligned}$$

Items (e) and (c) of Corollary B.2 imply that all entries of all $\Lambda^{[*]}$ are real numbers, and hence, for the spectral norm of the matrix exponentials, we find

$$\begin{aligned} \exp(2\pi i \Lambda^{[2]}/x) - \mathtt{I} &= 2\pi i \Lambda^{[2]}/x + \mathfrak{O}^{\geq 1}(1/x^2) \\ \left\| \exp(2\pi i x \mathtt{E}_{\Lambda}(1/x)) - \mathtt{I} \right\| &\leq 2\pi \left\| \mathtt{E}_{\Lambda}(1/x) \right\| \cdot x = O^{\geq R}(1/x^2), \end{aligned}$$

[30] The factor 1/2 ensures that the power series are bounded as $|z| \to r/2$.

where in the first line we just use the Taylor remainder bound of the exponential function, whereas in the second line we use the bound on $\mathtt{E}_\Lambda$ mentioned above.

Now, continuing (∗), we calculate

$$\begin{aligned}\mathtt{U}(x) = \Big(\mathtt{I} + \tfrac{1}{x}\mathtt{W}^{[1]} + \mathtt{E}_W(1/x)\Big) \\ \cdot \widetilde{\mathtt{U}}(x) \cdot \big(\mathtt{I} + (\exp(2\pi i \Lambda^{[2]}/x) - \mathtt{I})\big) \cdot \big(\mathtt{I} + (\exp(2\pi i x \mathtt{E}_\Lambda(1/x)) - \mathtt{I})\big) \\ \cdot \Big(\mathtt{I} - \tfrac{1}{x}\mathtt{W}^{[1]} + \mathtt{E}_W(1/x)^\dagger\Big)\end{aligned}$$

Collecting terms, we conclude

$$\mathtt{U}(x) = \widetilde{\mathtt{U}}(x) + \frac{\mathtt{C}(x)}{x} + \mathfrak{O}^{\geq R}(1/x^2),$$

with

$$\mathtt{C}(x) := [\mathtt{W}^{[1]}, \widetilde{\mathtt{U}}(x)] + 2\pi i \widetilde{\mathtt{U}}(x)\, \Lambda^{[2]}. \tag{26}$$

This completes the proof of Lemma 3.2, noting that $\mathfrak{O}^{\geq 1}$ arises from $\mathfrak{O}^{\geq R}$ by the continuity of the three summands on $\mathbb{R} \setminus \{0\}$. □

We can now derive Prop. 3.1 directly from Lemma 3.2.

3.3.b Proof of Prop. 3.1(c)

We have to show that

$$U(x) - \tilde{V}(x) \xrightarrow{|x|\to\infty} 0$$

implies $\tilde{V} = \tilde{U}$. But by Item (a) of the proposition, we see that $D := \tilde{U} - \tilde{V}$ is a bounded, finite-Fourier-spectrum, operator-valued function which satisfies

$$D(x) = \xrightarrow{|x|\to\infty} 0.$$

This implies that $D = 0$, by invoking, e.g., Prop. A.4 in Appendix A.2.

This completes the proof of Prop. 3.1, Fourier-decomposition for perturbed-parametric unitaries.

3.4 The Fourier spectrum

3.4.a Operator-valued tempered distributions

We set out to prove Prop. 3.4. The annoyingly technical proof is based on the Lie Product Formula

$$e^{2\pi i(xA+B)} = \lim_{n\to\infty}\big(e^{2\pi i x A/n} e^{2\pi i B/n}\big)^n, \qquad \text{for all } x \in \mathbb{R};$$

we emphasize that the limit is taken pointwise.

We will prove convergence of the Fourier transforms of the finite products to the Fourier transform of the RHS. The technical nuisance is that the Fourier transforms of the finite products are linear combinations of Dirac measures, while the infinite product has a continuous, square-integrable component. (We know that from the Fourier-decomposition theorem, Prop. 3.1.) Unfortunately, that means that the convergence can only happen in the sense of tempered distributions. Moreover, as we reason about the unitary-valued function (i.e., not the expectation-value function), we need operator-valued tempered distributions. The author wishes to emphasize that they don't add difficulty, just abstraction.

Operator-valued tempered distributions are, basically, matrices with entries in $\mathcal{S}'$; but while speaking of matrices requires fixing an orthonormal basis of the Hilbert space, we give the equivalent coordinate-free definition: An operator-valued tempered distribution τ is a sesqui-linear[31] mapping

$$\mathcal{H} \times \mathcal{H} \to \mathcal{S}' \colon (\phi, \psi) \mapsto \tau_{\phi,\psi}.$$

[31] Linear on the right side, anti-linear on the left.

For every given $\varphi \in \mathcal{S}$, as $(\phi, \psi) \mapsto \langle \varphi : \tau_{\phi,\psi} \rangle$ is a sesqui-linear form on $\mathcal{H}$, there exists a unique linear operator, denoted by $\langle \varphi : \tau \rangle$, that satisfies, for all ϕ, ψ

$$(\phi \mid \langle \varphi : \tau \rangle \psi) = \langle \varphi : \tau_{\phi,\psi} \rangle;$$

the mapping $\varphi \to \langle \varphi : \tau \rangle$ is linear. Moreover, the set of operator-valued tempered distributions is a complex vector space with the standard arithmetic operations.

With this machinery, we can now get to work.

3.4.b Fourier spectra of perturbed-parametric unitary functions

The Fourier spectrum of the expectation-value function (9) relates to that of the unitary function (8) as in the following lemma. We sketch the proof for the sake of completeness.

Lemma 3.3. *Let $n \in \mathbb{N}$ and F, $F_1, \dots, F_n$ bounded smooth operator-valued functions, and $M_0, \dots, M_n$ arbitrary operators.*

(a) The Fourier spectrum of the complex-number valued function $x \mapsto \operatorname{tr}(F(x))$ is contained in $\operatorname{Supp} \hat{F}$;

(b) The Fourier spectrum of the operator-valued function $x \mapsto F(x)^\dagger$ is

$$-\operatorname{Supp} \hat{F} := \{-\xi \mid \xi \in \operatorname{Supp} \hat{F}\};$$

(c) If each of $F_1, \dots, F_n$ has compact Fourier spectrum, then the Fourier spectrum of the operator-valued function $x \mapsto M_0 F_1(x) M_1 \dots F_n(x) M_n$ is contained in

$$\operatorname{Supp} \widehat{F_1} + \dots + \operatorname{Supp} \widehat{F_n},$$

(where the sum of sets $S_1, \dots, S_n$ is defined as $S_1 + \dots + S_n := \{s_1 + \dots + s_n \mid s_i \in S_i, i = 1, \dots, n\}$).

Sketch of proof.

- Item (a) follows from the fact that $\operatorname{tr} F(\square)$ is a sum of matrix elements, each with Fourier spectrum contained in $\operatorname{Supp} \hat{F}$.
- For Item (b), we consider the matrix elements: For $\phi, \psi \in \mathcal{H}$ we have, for all x,

$$(\phi \mid F(x)^\dagger \psi) = (\psi \mid F(x)\phi)^*,$$

and the statement follows from standard Fourier analysis: The equation $\widehat{f^*} = \hat{f}^*(-\square)$ holds also for the Fourier transform of tempered distributions f; hence $\langle \varphi : \widehat{f^*} \rangle = 0$ for all Schwartz functions φ with $\operatorname{Supp} \varphi \subseteq \mathbb{R} \setminus (-\operatorname{Supp} \hat{f})$.

- For Item (c), consider the function of n variables

$$Y \colon (x_1, \dots, x_n) \mapsto M_0 F_1(x_1) M_1 \dots F_n(x_n) M_n;$$

we are interested in the Fourier spectrum of the operator-valued function $x \mapsto Y(x, \dots, x)$. The matrix elements of Y are linear homogeneous polynomials of degree n in the matrix elements of the F_j, $j = 1, \dots, n$, i.e., they are of the form $g_1 \cdot \dots \cdot g_n$, where $\operatorname{Supp} \widehat{g_j} \subseteq \operatorname{Supp} \widehat{F_j}$, $j = 1, \dots, n$, and all these supports are compact.

The statement now follows from standard Fourier analysis. Indeed, from [22], Def. 6.36, following remarks, and Theorems 6.37 and 7.19, if $g_1, \dots, g_n$ are $\mathbb{C}$-valued bounded measurable functions with compact Fourier spectrum, then

$$(g_1 \cdot \dots \cdot g_n)^\wedge = \widehat{g_1} * \dots * \widehat{g_n};$$

the support of the last distribution is $\operatorname{Supp} \widehat{g_1} + \dots + \operatorname{Supp} \widehat{g_n}$ ([22, Theorem 6.37(b)], note that all distributions that occur are tempered).

□

The central mini-result of §3.4 is the following fact.

Proposition 3.4 (Fourier spectrum of perturbed-parametric unitary)**.** *With the notations of Def. 2.1, the Fourier spectrum of the perturbed-parametric unitary function is contained in the interval* $[\mathrm{eig}_{\min}(A), \mathrm{eig}_{\max}(A)]$.

Before we prove the proposition, let us derive Prop. 2.5 (§2.3) from it.

Proof of Prop. 2.5. Firstly, boundedness and analyticity of the expectation-value function are obvious.

Secondly, we consider the Fourier spectrum. With the notations of Definition 2.1, denote the unitary function by $U\colon x \mapsto e^{2\pi i(xA+B)}$, and the expectation-value function by $f\colon x \mapsto \mathrm{tr}(MU(x)\varrho U(x)^\dagger)$. By Prop. 3.4, the Fourier spectrum of U is contained in $[\mathrm{eig}_{\min}(A), \mathrm{eig}_{\max}(A)]$. Lemma 3.3 then gives that the Fourier spectrum of f is contained in

$$[\mathrm{eig}_{\min}(A), \mathrm{eig}_{\max}(A)] - [\mathrm{eig}_{\min}(A), \mathrm{eig}_{\max}(A)] \\ = [\mathrm{eig}_{\min}(A) - \mathrm{eig}_{\max}(A), \mathrm{eig}_{\max}(A) - \mathrm{eig}_{\min}(A)].$$

This concludes the proof. □

3.4.c Proof of Prop. 3.4

We pull the following lemma out for easier readability.

Lemma 3.5. *With the notations of Def. 2.1, letting, for* $n \in \mathbb{N}$,

$$U_n\colon x \mapsto \left(e^{2\pi i x A/n} e^{2\pi i B/n}\right)^n,$$

we have $\operatorname{Supp} \widehat{U_n} \subseteq [\mathrm{eig}_{\min} A, \mathrm{eig}_{\max} A]$.

Proof. Same calculation as always: Let P_λ be the spectral projection onto the eigenspace for eigenvalue λ of A. The equation

$$e^{2\pi i \,\square\, A/n} = \sum_{\lambda \in \mathrm{Spec}\, A} P_\lambda \cdot e^{2\pi i \lambda/n \cdot \square}$$

implies, by the linearity of the Fourier transform,

$$\left(e^{2\pi i \,\square\, A/n}\right)^\wedge = \sum_{\lambda \in \mathrm{Spec}\, A} P_\lambda \cdot \delta_{\lambda/n},$$

and hence $\operatorname{Supp}\left(\left(e^{2\pi i \,\square\, A/n}\right)^\wedge\right) = \mathrm{Spec}(A/n)$.

Now, by Lemma 3.3(c), we find that

$$\operatorname{Supp} \widehat{U_n} \subseteq \underbrace{\mathrm{Spec}(A/n) + \cdots + \mathrm{Spec}(A/n)}_{n \text{ terms}}.$$

Noting that $\mathrm{Spec}(A/n) = (\mathrm{Spec}\, A)/n$, the proof of the lemma is concluded by observing that the sum of sets on the RHS is contained in the interval $[\mathrm{eig}_{\min} A, \mathrm{eig}_{\max} A]$. □

Proof of Prop. 3.4. With the notations of Def. 2.1, denote the unitary function by $U\colon x \mapsto e^{2\pi i(xA+B)}$, and let U_n, $n \in \mathbb{N}$, as in Lemma 3.5. By the Lie Product Formula, for every fixed $x \in \mathbb{R}$, the sequence of operators $(U_n(x))_{n\in\mathbb{N}}$ converges to $U(x)$, i.e., the sequence of operator-valued functions $(U_n)_n$ converges pointwise to U.

Claim. The sequence converges to U in the tempered-distribution topology.

This is a known fact, modulo technicalities arising from the functions being operator-valued; the standard arguments are below, for the sake of completeness.

From the claim, as the Fourier transform is continuous, we find that $\hat{U}_n \xrightarrow{n\to\infty} \hat{U}$, and hence

$$\operatorname{Supp}\hat{U} \subseteq \bigcup_{n\in\mathbb{N}} \operatorname{Supp}\widehat{U_n}; \tag{$*$}$$

indeed: if φ is a Schwartz function with support disjoint from $\bigcup_{n\in\mathbb{N}} \operatorname{Supp}\widehat{U_n}$, then $\langle\varphi : \widehat{U_n}\rangle = 0$ for all n, and hence $\langle\varphi : \widehat{U}\rangle = 0$.

From ($*$) and Lemma 3.5, we arrive at $\operatorname{Supp}\hat{U} \subseteq [\mathrm{eig}_{\min}(A), \mathrm{eig}_{\max}(A)]$.

Proof of the claim. Let $\varphi\colon \mathbb{R} \to \mathbb{C}$ be a Schwartz function. We have to show that the sequence of operators $(\langle\varphi : U_n\rangle)_n$ converges to $\langle\varphi : U\rangle$. It is sufficient (finite-dimensionality of $\mathcal{H}$) to show convergence for every matrix element, i.e., for every $\phi, \psi \in \mathcal{H}$, we have to show $(\phi \mid \langle\varphi : U_n\rangle\psi) \xrightarrow{n\to\infty} (\phi \mid \langle\varphi : U\rangle\psi)$. By definition,

$$(\phi \mid \langle\varphi : U_n\rangle\psi) = \langle\varphi : x \mapsto (\phi \mid U_n(x)\psi)\rangle = \int (\phi \mid U_n(x)\psi)\varphi(x)\,dx;$$

similarly for U in place of U_n. As we have $\big|(\phi \mid U_n(x)\psi)\varphi(x)\big| \leq \|\phi\|\|\psi\| \cdot |\varphi(x)|$ for all x and $|\varphi|$ is integrable, the condition in the dominated convergence theorem applies, and we have $(\phi \mid \langle\varphi : U_n\rangle\psi) \xrightarrow{n\to\infty} (\phi \mid \langle\varphi : U\rangle\psi)$. This completes the proof of the claim, and hence of Prop. 3.4. □

4 The derivative at $1/2$

We continue using the tempered-distributions notations defined in §3.1.

In this section, we discuss the Fourier-analytic characterization of measures computing the derivative at $1/2$ of functions in $\mathscr{F}_*$ and of functions in $\mathscr{F}_*^{\downarrow}$; cf. (12). We derive consequences for the space of feasible shift rules, and, last not least, we'll prove that the Nyquist shift rules are feasible.

We start by discussing the potential consequences of the Fourier-Decomposition theorem in the form of Corollary 2.9, i.e., the "$\mathscr{F}_*$ vs. $\mathscr{F}_*^{\oplus}$".

4.1 Impact of the Fourier-decomposition theorem

In the next subsection §4.2, we will prove the characterization, in terms of $\hat{\mu}$, of the finite complex measures μ which satisfy (12): Integrating a against μ gives the derivative at $1/2$ for all functions in $\mathscr{F}_{1/2}$ (Lemma 2.12). The current section proves that restricting to the smaller space $\mathscr{F}_*^{\oplus}$ of Corollary 2.9 does not yield any additional feasible PSRs, i.e., we prove Prop. 2.10.

Fix a frequency set $\Xi \subset [-1/2, +1/2]$ with $\max\Xi = 1/2$. As both taking the derivative at $1/2$ and integration against μ are linear, a characterization of derivative-computing measures for $\mathscr{F}_\Xi^{\oplus}$ (cf. (12)) must consist of the following two parts:

(1.) For all $\xi \in \Xi$, we need $\int e^{2\pi i\xi x}, d\mu(x) = 2\pi i\xi e^{i\pi\xi}$;
(2.) For all $f_0 \in \mathscr{F}_{1/2}^{\downarrow}$, we need $\int f_0(x)\,d\mu(x) = f_0'(1/2)$.

The first part takes care of f_1 in the notation of the Fourier-decomposition theorem, Prop. 2.7. It corresponds to condition (13) in Lemma 2.12, but it is now a condition on only a finite number of frequencies, exactly like for the unperturbed theory in Eqn. (8) of [18] (only that we have chosen the anchor point $1/2$ here, instead of 0 in [18]).

As for part *(2.)*, the hope would be that, even though in the functions in $\mathscr{F}_{1/2}^{\downarrow}$ all Fourier frequencies in $[-1/2, +1/2]$ are allowed to occur, the linear decay condition would result in a larger set of feasible PSRs.

Unfortunately, that does not seem to be the case: Not even restricting the space in part *(2.)* to only translated sinc-functions,

$$f_0 = \operatorname{sinc}(\,\square - x_0) \text{ for some } x_0 \in \mathbb{R},$$

increases the space of feasible PSRs, as the following lemma shows.

Lemma 4.1. *Let μ be a complex measure. If*

$$\operatorname{sinc}'(1/2 - x_0) = \int \operatorname{sinc}(\,\square - x_0)\, d\mu \qquad \textit{for all } x_0 \in \mathbb{Z}, \tag{weak-12}$$

then μ satisfies the condition (13) *of Lemma 2.12.*

Clearly, (12) implies (weak-12); Lemma 4.1 together with Lemma 2.12 show that the two conditions are in fact equivalent.

Proof of Lemma 4.1. We use a basic fact [23, Theorem 2.1] about Fourier transforms of periodic functions in the following form: If $h_1, h_2 \colon \mathbb{R} \to \mathbb{C}$ are two continuous functions such that, for all $x_0 \in \mathbb{Z}$,

$$\int_{-1/2}^{1/2} e^{2\pi i \xi x_0} h_1(\xi)\, d\xi = \int_{-1/2}^{1/2} e^{2\pi i \xi x_0} h_2(\xi)\, d\xi, \tag{$*$}$$

then h_1, h_2 agree on $[-1/2, +1/2]$.

For h_2 we take the RHS in condition (13): $h_2(\xi) = -2\pi i \xi e^{-i\pi\xi}$. With μ satisfying Condition (weak-12), for h_1 we take $\hat{\mu}$, recalling that the Fourier-Stieltjes transform of a finite measure is continuous.

That's it, now we verify ($*$) by direct calculation. For every $x_0 \in \mathbb{R}$, we have ($\check{\square}$ is inverse Fourier transform):

$$\begin{aligned}
\int_{-1/2}^{1/2} e^{2\pi i \xi x_0} \hat{\mu}(\xi)\, d\xi &= \int_{-1/2}^{1/2} e^{2\pi i \xi x_0} \int_{\mathbb{R}} e^{-2\pi i \xi x}\, d\mu(x)\, d\xi \\
&= \int_{\mathbb{R}} \int_{-1/2}^{1/2} e^{2\pi i \xi x_0}\, e^{-2\pi i \xi x}\, d\xi\, d\mu(x) && \text{[finite measures, bounded functions]} \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-2\pi i \xi x} \cdot e^{2\pi i \xi x_0} \mathbf{1}_{[-1/2,+1/2]}(\xi)\, d\xi\, d\mu(x) \\
&= \int_{\mathbb{R}} \operatorname{sinc}(\,\square - x_0)\, d\mu && [(\mathbf{1}_{[-1/2,+1/2]})^{\wedge} = \operatorname{sinc}] \\
&= \operatorname{sinc}'(1/2 - x_0) && \text{[condition (weak-12)]} \\
&= \int_{\mathbb{R}} e^{-2\pi i \xi (1/2 - x_0)} \cdot (-2\pi i \xi) \mathbf{1}_{[-1/2,+1/2]}\, d\xi && [\operatorname{sinc}' = \left(-2\pi i\, \square\, \mathbf{1}_{[-1/2,+1/2]}\right)^{\wedge}] \\
&= \int_{-1/2}^{1/2} e^{2\pi i \xi x_0} (-2\pi i \xi) e^{-i\pi\xi}\, d\xi.
\end{aligned}$$

This completes the proof of Lemma 4.1. □

We can now finish off the proposition in §2.3.a.

Proof of Prop. 2.10. The claimed feasibility for $\mathscr{F}_{\max \Xi}$ follows from the feasibility for $\mathscr{F}_{\Xi}^{\oplus}$ by Lemma 4.1 and Prop. 2.11. □

4.2 Fourier-analytic characterization of derivative-computing measures, case $\mathscr{F}_*^{\oplus}$

In §2.4.a, we have proven one direction of Lemma 2.12. Here, we prove the other direction in the special case described in Corollary 2.9: We replace, in condition (12), the qualification "$g \in \mathscr{F}_{1/2}$" by "$g \in \mathscr{F}_\Xi^{\oplus}$", for arbitrary frequency set $\Xi \subset [-1/2, +1/2]$.

As a matter of fact, we will take a condition that is both more general and allows for a lazier proof — here is the statement (cf. Lemma 2.12):

Proposition 4.2. *Let μ be a complex measure on $\mathbb{R}$. If (13) holds, then $g'(1/2) = \int g\, d\mu$ holds for all functions $g \in \mathscr{F}_{1/2}$ which can be decomposed as $g = g_0 + g_1$ where*

- *g_0 is smooth and square integrable; and*
- *g_1 has finite Fourier spectrum, i.e.[32], there exists a finite $\Xi \subseteq [-1/2, +1/2]$ with $g_1 \in \mathscr{K}_\Xi$.*

Proof. Let μ, g, g_0, g_1, Ξ as described. There exist complex numbers b_ξ, $\xi \in \Xi$, such that

$$\begin{aligned} g &= g_0 + \sum_{\xi\in\Xi} b_\xi e^{2\pi i \xi \,\square}, \\ \hat{g} &= \widehat{g_0} + \sum_{\xi\in\Xi} b_\xi \delta_\xi. \end{aligned} \tag{$*$}$$

Note that $\widehat{g_0}$ is the usual L^2-Fourier transform of g_0, in particular, it is an L^2-function (not an evil tempered distribution).

By linearity, it suffices to prove the derivative-computing measure equation for each term of ($*$) separately:

$$\text{For the } L^2\text{-part:} \qquad g_0'(1/2) = \int g_0(x)\, d\mu(x); \tag{A}$$

$$\text{For each } \xi \in \Xi\text{:} \qquad 2\pi i\xi\, e^{i\pi\xi} = \int e^{2\pi i\xi x}\, d\mu(x). \tag{B}$$

The equations (B) correspond to the condition (13), evaluated in $-\xi$, so there is nothing to be done.

For the L^2-part (A), we argue based on basic Fourier analysis. As g_0 is a smooth L^2-function whose Fourier spectrum is contained in $[-1/2, +1/2]$, we have $\hat{g}_0 \in L^1([-1/2, +1/2])$, and the following equations:

$$g_0(x) = \int_{-1/2}^{1/2} e^{2\pi i\xi x}\, \hat{g}_0(\xi)\, d\xi \qquad \text{for all } x \in \mathbb{R}, \text{ and}$$

$$g_0'(x) = \int_{-1/2}^{1/2} 2\pi i\xi\, e^{2\pi i\xi x}\, \hat{g}_0(\xi)\, d\xi \qquad \text{for all } x \in \mathbb{R}.$$

[32] Note that boundedness of g_1 is implied by the smoothness of g_0.

Using these, we calculate:

$$\begin{aligned}
\int g_0(x)\,d\mu(x) &= \int \int_{-1/2}^{1/2} e^{2\pi i\xi x}\,\widehat{g_0}(\xi)\,d\xi\,d\mu(x) \\
&= \int_{-1/2}^{1/2} \int e^{2\pi i\xi x}\,\widehat{g_0}(\xi)\,d\mu(x)\,d\xi \\
&= \int_{-1/2}^{1/2} \int e^{2\pi i\xi x}\,d\mu(x)\widehat{g_0}(\xi)\,d\xi \\
&= \int_{-1/2}^{1/2} (2\pi i\xi)e^{i\pi\xi}\widehat{g_0}(\xi)\,d\xi \qquad \text{[condition (13)]} \\
&= g_0'(1/2).
\end{aligned}$$

This completes the proof. □

As noted in §2.4.a, the proof of the general case of Lemma 2.12, i.e., taking derivatives of all $\mathscr{F}_{1/2}$-functions, is in Appendix C.

4.3 Proof of the Space-of-Feasible-PSRs proposition

We now prove Prop. 2.13: The set of all signed measures μ satisfying (13) is either empty or an infinite-dimensional (real) affine space. If it is empty, there is nothing to prove. Otherwise let μ be such a measure.

As indicated in §2.4.a, we will exhibit a set of measures ν_a, $a \in \mathbb{N}$, that are linearly independent[33] and that satisfy $(\operatorname{Supp}\nu_a) \cap [-1/2, +1/2] = 0$. Each of the measures $\mu + \nu_a$, $a \in \mathbb{N}$, satisfies the condition (13) in Lemma 2.12, as μ satisfies it.

For the ν_a we take the finite signed measures $\cos(2\pi a\,\square)\operatorname{sinc}^2$ (density on Lebesgue measure). We prove the conditions in the following two lemmas, but first we need a remark.

Remark 4.3. *Denote by* $\Lambda\colon \mathbb{R} \to \mathbb{R}\colon \xi \mapsto \max(0, 1 - |\xi|)$ *the triangle function. It is a standard fact that* $\hat{\Lambda} = \operatorname{sinc}^2$*, and this equation easily implies*

$$\check{\Lambda} = \operatorname{sinc}^2 .$$

Lemma 4.4. *For all* $a \in \mathbb{R}$*, we have* $\big(\cos(2\pi a\,\square)\operatorname{sinc}^2\big)^{\wedge} = \big(\Lambda(\square - a) + \Lambda(\square + a)\big)/2$.

Proof. By Remark 4.3, we have $\big(\Lambda(\square - a)\big)^{\vee} = e^{2\pi i a\,\square}\operatorname{sinc}^2$, so that $\Big(\big(\Lambda(\square - a) + \Lambda(\square + a)\big)/2\Big)^{\vee} = \cos(2\pi a\,\square)\operatorname{sinc}^2$.

The statement about the Fourier transform of $\cos(2\pi a\,\square)\operatorname{sinc}^2$ follows from the Fourier Inversion Theorem (e.g., [24, Theorem 4.11]). □

Lemma 4.5. *The set of functions*

$$\Big\{\cos(2\pi(3a + 1/2)\,\square)\ \operatorname{sinc}^2 \;\Big|\; a \in \mathbb{N}\Big\}$$

is linearly independent.

[33] I.e., for every finite(!) linear combination of ν_*'s, if the result equals the 0-measure, then all coefficients in the linear combination must be 0.

Before we prove the lemma, we finish the proof of Prop. 2.13. Let

$$\nu_a := \cos(2\pi(3a + 1/2)\,\square\,)\,\mathrm{sinc}^2 \cdot \lambda,$$

where λ is the Lebesgue measure. By Lemma 4.4, we have

$$\operatorname{Supp} \widehat{\nu_a} = [-3a - 3/2, -3a + 1/2] \cup [3a - 1/2, 3a + 3/2] \tag{28}$$

and these sets are disjoint from $[-1/2, +1/2]$. The proof of Prop. 2.13 is completed by proving Lemma 4.5: the linear independence of the ν_a, $a \in \mathbb{N}$.

Proof of Lemma 4.5. The functions $\widehat{\nu_a}$, $a \in \mathbb{N}$ are linearly independent as their supports are disjoint; see (28). As the inverse Fourier transform is injective and Fourier Inversion holds, the measures ν_a, $a \in \mathbb{N}$, are linearly independent, too. □

4.4 Proof of feasibility of the Nyquist shift rule

In this section, we prove Theorem 2.14: The measure μ defined in (14) satisfies the condition (12) of Lemma 2.12.

The function $u\colon \mathbb{Z} \to \mathbb{R}$ in (14) plays a special role. In the next subsection, §4.4.a, we will prove the following.

Proposition 4.6. *With u as in (14a), we have,*

$$\text{for all } \xi \in [-1/2, +1/2]\text{:} \quad \hat{u}(\xi) = -2\pi i\xi\, e^{-i\pi\xi}. \tag{29}$$

Here, we are using the Fourier transform on $\mathbb{Z}$: For all absolutely summable sequences $v\colon \mathbb{Z} \to \mathbb{C}$,

$$\hat{v}(\xi) := \sum_{n\in\mathbb{Z}} e^{-2\pi i\xi n} v(n); \tag{30}$$

the result will be a 1-periodic[34] continuous function.[35]

Completing the proof of the theorem is now a *piece of cake.*

Proof of Theorem 2.14. Let μ be as in (14), and $\xi \in [-1/2, +1/2]$. A swift calculation:

$$\begin{aligned}
\hat{\mu}(\xi) &= \int e^{-2\pi i\xi s}\mu(s) \\
&= \int e^{-2\pi i\xi s}\, d\big(\textstyle\sum_n u(n)\delta_n\big)(s) && \text{[def. of } \mu \text{ in (14)]} \\
&= \sum_{n\in\mathbb{Z}} \int e^{-2\pi i\xi s} u(n)\, d\delta_n(s) \\
&= \sum_{n\in\mathbb{Z}} u(n) e^{-2\pi i\xi n} \\
&= \hat{u}(\xi) \\
&= -2\pi i\xi\, e^{-i\pi\xi} && \text{[by Prop. 4.6.]}
\end{aligned}$$

And the proof is completed. □

[34] More accurately, a continuous function on $\mathbb{R}/\mathbb{Z}$.

[35] And, wait for it — the RHS of (29) indeed evaluates to $-i\pi$ for both $\xi = 1/2$ and $\xi = -1/2$.

4.4.a The Fourier transform of $u \in \ell^1(\mathbb{Z})$

To compute the $\mathbb{Z}$-Fourier transform of u, basically, you have to take the $\mathbb{Z}$-Fourier transform of $n \mapsto 1/n^2$ and work from there.

We denote by Li_2 the *dilogarithm* or *Spence's function* [25]. It is a holomorphic function on $\mathbb{C} \setminus [1, \infty[$ with continuous extensions to $\mathbb{C} \setminus]1, \infty[$. The eponymous equation

$$\mathrm{Li}_2(z) = \sum_{n=1}^{\infty} \frac{z^n}{n^2} \text{ for all } z \in \mathbb{C} \text{ with } |z| \leq 1 \tag{31a}$$

holds (Eqn. (3.1) in [26], or beginning of Chapter I in [25], plus Abel's theorem). Moreover we have the functional equation (§I.2 in [25], Eqn. (3.2) in [26]):

$$\mathrm{Li}_2(z) + \mathrm{Li}_2(1/z) = \tfrac{-1}{2} \operatorname{Log}^2(-z) - \pi^2/6 \quad \text{for all } z \notin [0, \infty[, \tag{31b}$$

where Log is the principle branch of the complex logarithm, i.e., defined on $\mathbb{C} \setminus]-\infty, 0]$.

We recycle the notation "%"; see (20).

Lemma 4.7. *For all $\xi \in \mathbb{R}$ we have*

$$\mathrm{Li}_2(e^{2\pi i\xi}) + \mathrm{Li}_2(e^{-2\pi i\xi}) = \frac{-\big(2\pi i \cdot ((\xi + 1/2) \,\%\, 1)\big)^2}{2} - \pi^2/6.$$

Proof. The functional equation (31b) of the dilogarithm gives the equation for all $\xi \notin \mathbb{Z}$.

We verify it by hand for $\xi \in \mathbb{Z}$. In that case, on the LHS we have $2 \cdot \mathrm{Li}_2(1) = \pi^2/3$ by (31a) and the Basel Problem (or §I.1 in [25]).

On the RHS, for integral ξ, we find that $(\xi + 1/2) \,\%\, 1 = -1/2$, and hence

$$\frac{-\big(2\pi i \cdot ((\xi + 1/2) \,\%\, 1)\big)^2}{2} = \frac{-(-i\pi)^2}{2} = \pi^2/2,$$

from which we have to subtract $\pi^2/6$ to arrive at the RHS; so LHS and RHS coincide also for integral ξ. □

We define for all $\xi \in \mathbb{R}$,

$$\begin{aligned} g(\xi) &:= \mathrm{Li}_2(e^{2\pi i\xi}) + \mathrm{Li}_2(e^{-2\pi i\xi}) && \text{(32a)} \\ &= \sum_{n \in \mathbb{Z}\setminus\{0\}} \frac{e^{-2\pi i\xi n}}{n^2} && \text{[by (31a)]} \\ &= \frac{-\big(2\pi i \cdot ((\xi + 1/2) \,\%\, 1)\big)^2}{2} - \pi^2/6 && \text{[by Lemma 4.7.]} \\ g_0(\xi) &:= g(2\xi)/4 && \text{(32b)} \\ &= \sum_{\substack{n \in \mathbb{Z}\setminus\{0\}, \\ n \text{ even}}} \frac{e^{-2\pi i\xi n}}{n^2} \\ g_1(\xi) &:= g(\xi) - g_0(\xi) && \text{(32c)} \\ &= \sum_{\substack{n \in \mathbb{Z}, \\ n \text{ odd}}} \frac{e^{-2\pi i\xi n}}{n^2}. \end{aligned}$$

(The equation for g_0 follows from the one for g by applying it for 2ξ.)

Lemma 4.8. *For all $\xi \in \mathbb{R}$,*

$$\hat{u}(\xi) = \tfrac{-4i}{\pi} \cdot e^{-i\pi\xi} \cdot g_1((\xi - 1/2)/2).$$

Proof. This is a direct calculation:

$$\begin{aligned}
\hat{u}(\xi) &= \sum_{n\in\mathbb{Z}} \frac{-(-1)^n e^{-2\pi i\xi n}}{\pi\,(n-1/2)^2} \\
&= \sum_{n\in\mathbb{Z}} \frac{-(-1)^n e^{-2\pi i\xi n}}{\pi\,(2n-1)^2/4} \\
&= \sum_{n\in\mathbb{Z}} \frac{-e^{-2\pi i(\xi-1/2)n}}{\pi\,(2n-1)^2/4} \\
&= \tfrac{-4}{\pi} \sum_{n\in\mathbb{Z}} \frac{e^{-2\pi i(\xi-1/2)n}}{(2n-1)^2} \\
&= \tfrac{-4}{\pi} \sum_{n\in\mathbb{Z}} \frac{e^{-2\pi i(\xi/2-1/4)(2n)}}{(2n-1)^2} \\
&= \tfrac{-4}{\pi} \cdot e^{-2\pi i(\xi/2-1/4)} \cdot \sum_{n\in\mathbb{Z}} \frac{e^{-2\pi i(\xi/2-1/4)(2n-1)}}{(2n-1)^2} \\
&= \tfrac{-4i}{\pi} \cdot e^{-i\pi\xi} \cdot g_1((\xi-1/2)/2),
\end{aligned}$$

where in the last step we have used the equation for g_1 above. □

We can now complete the proof of the proposition.

Proof of Prop. 4.6. As $\hat{v}$ is 1-periodic for every absolutely summable $v\colon \mathbb{Z}\to\mathbb{Z}$, we fix $\xi \in [-1/2, +1/2[$. We have

$$\begin{aligned}
g_1((\xi-1/2)/2) &= g((\xi-1/2)/2) - g_0((\xi-1/2)/2) \\
&= g((\xi-1/2)/2) - g(\xi-1/2)/4 \\
&= \frac{-\big(2\pi i\cdot((\xi/2+1/4)\,\%\,1)\big)^2}{2} - \frac{-\big(2\pi i\cdot(\xi\,\%\,1)\big)^2}{2}\cdot\frac{1}{4} \\
&\quad - \pi^2/6 + \pi^2/24 \\
&\underset{(*)}{=} 2\pi^2\Big((\xi/2+1/4)^2 - \xi^2/4\Big) - \pi^2/8 \\
&= \tfrac{\pi^2}{2}(\xi+1/4) - \pi^2/8.
\end{aligned}$$

In equation marked with an asterisk, omitting the "%1" is allowed as $\xi \in [-1/2, +1/2[$ and hence $\xi/2 + 1/4 \in [-1/2, +1/2[$.

Now starting from Lemma 4.8, we compute, for every $\xi\in\mathbb{R}$,

$$\begin{aligned}
\hat{u}(\xi) &= \tfrac{-4i}{\pi}\cdot e^{-i\pi\xi}\cdot g_1((\xi-1/2)/2) \\
&= \tfrac{-4i}{\pi}\cdot e^{-i\pi\xi}\cdot\big(\tfrac{\pi^2}{2}(\xi+1/4) - \pi^2/8\big) \\
&= -2\pi i\cdot e^{-i\pi\xi}\cdot(\xi+1/4-1/4) \\
&= -2\pi i\xi\cdot e^{-i\pi\xi}.
\end{aligned}$$

This completes the proof of the proposition. □

5 Details about reflection and dilation

5.1 Proof of the general facts

We prove Prop. 2.11 stated in §2.4. The main work is in the following lemma.

Lemma 5.1. *Let μ be a finite measure.*

(a) With $\tau\colon \mathbb{R} \to \mathbb{R}\colon s \mapsto 1/2 - s$ the reflection on $1/2$: The finite measure μ satisfies (12), if and only if $\phi := \tau(\mu)$ is feasible for $\mathscr{F}_{1/2}$.

(b) For fixed K_1, K_2, with $\tau\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto K_1 x/K_2$ the dilation by K_1/K_2: If the PSR ϕ is feasible for $\mathscr{F}_{K_1}$, then the PSR $\psi := K_2/K_2 \cdot \tau(\phi)$ is feasible for $\mathscr{F}_{K_2}$.

Proof. Proof of Item (a): Let $x \in \mathbb{R}$ and $f, g \in \mathscr{F}_{1/2}$ such that $g = f(\,\square + x - 1/2)$, i.e., $f = g(\,\square - x + 1/2)$; note that $\mathscr{F}_{1/2}$ is invariant under translations. We write:

$$\begin{aligned} f'(x) = g'(1/2) \quad \underset{(?)}{=} \quad \int g(s)\, d\mu(s) &= \int f(x - (1/2 - s))\, d\mu(s) \\ &= \int f(x - \square) \circ \tau \, d\mu \\ &= \int f(x - \square)\, d\phi \\ &= f * \phi(x), \end{aligned}$$

where all equations except the one tagged with the question mark hold, and (?) is equivalent to $f'(x) = f * \phi(x)$. As (?) is (12), this proves Item (a).

Proof of Item (b): For $f \in \mathscr{F}_{K_2}$ and $x \in \mathbb{R}$, noting that $f(K_1\,\square/K_2) \in \mathscr{F}_{K_1}$, we calculate, using in $(*)$ that ϕ is feasible for $\mathscr{F}_{K_1}$:

$$\begin{aligned} f'(x) &= \tfrac{K_2}{K_1}\, \big(f(K_1\,\square/K_2)\big)'\,(K_2\,x/K_1) \\ &\underset{(*)}{=} \tfrac{K_2}{K_1}\, (f(K_1\,\square/K_2)) * \phi\,(K_2\,x/K_1) = \tfrac{K_2}{K_1} \int f(K_1(K_2\,x/K_1 - s)/K_2) d\phi(s) \\ &= \tfrac{K_2}{K_1} \int f(x - K_1\,s/K_2) d\phi(s) = \int f(x - \square) \circ \tau\, \tfrac{K_2}{K_1}\, d\phi = \int f(x - \square)\, d\Big(\tfrac{K_2}{K_1}\tau(\phi)\Big) \\ &= \int f(x - \square)\, d\psi = f * \psi(x), \end{aligned}$$

which proves Item (b). □

As a direct corollary of the lemma we obtain the isomorphisms required in Prop. 2.11, noting that: *(1)* taking images of measures is a $\mathbb{C}$-linear operation; *(2)* reflection and dilation map signed measures to signed measures; *(3)* reflection is self-inverse, and the inverse of dilation by α is dilation by $1/\alpha$.

This completes the proof of Prop. 2.11.

5.2 Derivation of the Nyquist shift rules

Proof of Corollary 2.15. Applying Theorem 2.14, Lemma 2.12, and Lemma 5.1(a) to the μ defined in (14), we find that the measure $\phi := \tau(\mu)$, for $\tau\colon x \mapsto 1/2 - x$ the reflection on $1/2$, is a PSR that is feasible for $\mathscr{F}_{1/2}$. We calculate, by linearity and norm-continuity of $\nu \mapsto \tau(\nu)$ and Eqn. (7),

$$\begin{aligned} \phi = \tau(\mu) = \sum_{n\in\mathbb{Z}} u(n)\tau(\delta_n) = \sum_{n\in\mathbb{Z}} u(n)\delta_{\tau(n)} &= \sum_{n\in\mathbb{Z}} u(n)\delta_{1/2-n} \\ = \sum_{n\in\mathbb{Z}} \frac{(-1)^{n-1}}{\pi\,(n-1/2)^2}\,\delta_{1/2-n} &\underset{(*)}{=} \sum_{a\in 1/2+\mathbb{Z}} \frac{(-1)^{a+1/2}}{\pi\,a^2}\,\delta_a = \phi_{1/2} \end{aligned}$$

where in the equation marked with the asterisk we perform the change of variables $a := 1/2 - n$. This proves the corollary. □

Proof of Corollary 2.16. From the previous corollary, we know that $\phi_{1/2}$ is a PSR that is feasible for $\mathscr{F}_{1/2}$. Fix $K > 0$. Applying Lemma 5.1(b), we find that the measure $2K \cdot (\square/2K)(\phi_{1/2})$ is a PSR feasible for $\mathscr{F}_K$. Again invoking linearity, continuity, and Eqn. (7), we calculate

$$2K \cdot (\square/2K)(\phi_{1/2}) = 2K \sum_{a \in 1/2+\mathbb{Z}} \frac{(-1)^{a+1/2}}{\pi\, a^2} (\square/2K)(\delta_a) = 2K \sum_{a \in 1/2+\mathbb{Z}} \frac{(-1)^{a+1/2}}{\pi\, a^2} \delta_{a/2K} = \phi_K,$$

and the proof of the corollary is complete. □

6 Optimality

This section has parallels with [18].

6.1 The Weak Duality Theorem

Compared to [18, Prop. 6], the version of the Weak Duality Theorem in the current paper has the additional statement about the inequality being an equation.

Proof of Prop. 2.20. Let ϕ be a PSR feasible for $\mathscr{G}$ and $f \in \mathscr{G}$. We calculate:

$$\begin{aligned}
-f'(0) &= (f(-\square))'(0) \\
&= (f(-\square)) * \phi(0) && \text{[by feasibility of } \phi \text{ and } f(-\square) \in \mathscr{G}\text{]} \\
&= \int f(s)\, d\phi(s) \\
&\underset{(*)}{\leq} \|f\|_\infty \cdot \|\phi\| && \text{[(15a) in Remark 2.17]} \\
&= \|\phi\|. && \text{[as } \|f\|_\infty \leq 1\text{]}
\end{aligned}$$

In the second equation, we have used the condition that $f(-\square) \in \mathscr{G}$ if $f \in \mathscr{G}$.

The inequality $-f'(0) \leq \|\phi\|$ — that we have just proven — is satisfied with equality if and only if the inequality $(*)$ holds with equality. □

6.2 Proof of optimality of the Nyquist shift rules

Here we give the proof of Theorem 2.21.

Proof of Theorem 2.21. In view of the Weak Duality Theorem, Prop. 2.20, we need to present a function $f^\star \in \mathscr{F}_K$ such that

$$-\partial f^\star(0) = \|\phi_K\|. \tag{$*$}$$

Clearly, the function $f^\star \colon x \mapsto -\sin(2\pi K x)$ is in $\mathscr{F}_K$, and we have $-\partial f^\star(0) = 2\pi K$.

To compute $\|\phi_K\|$, first recall the well-known fact (Lemma A.1 in Appendix A.1) that the sum of all reciprocals of the squares of all positive odd integers is $\pi^2/8$.

Regarding ϕ_K, we have

$$\begin{aligned}
\|\phi_K\| = 2K \cdot \sum_{a \in 1/2+\mathbb{Z}} \frac{1}{\pi\, a^2} = 2K \cdot \sum_{n \in \mathbb{Z}} \frac{1}{\pi\,(n+1/2)^2} = 2K \cdot \sum_{n \in \mathbb{Z}} \frac{1}{\pi\,(2n+1)^2/4} \\
= 2K \cdot \frac{4}{\pi} \cdot \sum_{\substack{m \in \mathbb{Z} \\ m \text{ odd}}} \frac{1}{m^2} = 2K \cdot \frac{4}{\pi} \cdot 2\pi^2/8 = 2\pi K.
\end{aligned}$$

This establishes the second part of $(*)$, and completes the proof of the theorem. □

6.3 The support of optimal proper shift rules

We now sketch the proof of Corollary 2.22.

Let $K > 0$ and assume ϕ is a PSR feasible for $\mathscr{F}_K$. Moreover, let $f^\star := -\sin(2\pi K\,\square)$ be the dual optimal solution established in the proof of Theorem 2.21 above.

As we already know from the proof of Theorem 2.21 that $-\partial f^\star(0)$ is equal to the norm of the optimal shift rule, by the Weak Duality Theorem applied to $\phi, f^\star$, equality in the inequality (16) there is sufficient and necessary for ϕ to be optimal. The last sentence of Prop. 2.20 states that that is equivalent to inequality (15a) of Remark 2.17 holding with equality.

Now, it is elementary (and left to the reader) to check that the condition in the statement of Corollary 2.22 is equivalent to the inequality (15a) of Remark 2.17 holding with equality for ϕ.

7 Truncation and folding

7.1 Truncation

We extract the technical computations as a lemma, for easier readability.

Lemma 7.1. *With $K, N, \phi_K, \phi_K^{(N)}$ as in Prop. 2.25, we have*

$$\left\|\phi_K^{(N)} - \phi_K\right\| = \frac{4K}{\pi} \cdot \sum_{n=N}^{\infty} \frac{1}{(n+1/2)^2} \leq \frac{4K}{\pi(N-1/2)}.$$

Proof. The difference between the two measures is

$$2K \cdot \sum_{n=N}^{\infty} \frac{(-1)^{n+1}}{\pi\,(n+1/2)^2}\,\delta_{(-1/2-n)/2K} + 2K \cdot \sum_{n=N}^{\infty} \frac{(-1)^{n+1}}{\pi\,(n+1/2)^2}\,\delta_{(1/2+n)/2K}.$$

Taking the norm, we obtain

$$\frac{4K}{\pi} \cdot \sum_{n=N}^{\infty} \frac{1}{(n+1/2)^2}.$$

Now some high-school math: We lazily upper-bound the sum by the integral

$$\int_{N-1/2}^{\infty} \frac{dx}{x^2} = \frac{1}{N-1/2},$$

which completes the proof of the lemma. □

Proof of Prop. 2.25. By Lemma 2.24, $\phi_K^{(N)}$ is $\|\phi_K^{(N)} - \phi_K\|$-nearly feasible, and Lemma 7.1, the norm is upper-bounded by ε. □

7.2 Folding

7.2.a Folding with zero perturbation ($[A, B] = 0$)

Here we prove Theorem 2.27, which states that, for all positive real numbers K, ξ_1 with $L := K/\xi_1 \in \mathbb{N}$, ϕ_K the Nyquist shift rule (5) and ψ as in Remark 2.26, shift-folding ϕ_K with mod-$1/\xi_1$ gives exactly ψ.

The proof relies on the *partial fraction decomposition of* $1/\sin^2$*:*

$$\text{For all } z \in \mathbb{C} \setminus \mathbb{Z}\text{:} \quad \frac{\pi^2}{\sin^2(\pi z)} = \sum_{\ell \in \mathbb{Z}} \frac{1}{(z+\ell)^2}; \tag{33}$$

this is a textbook identity, see, e.g., [27, Exercise 12 in Chapter 3].

We note the following elementary fact about modulo-arithmetic for reference:

$$\text{For all } \gamma > 0,\ x \in \mathbb{R}\text{:} \quad x \,\%\, \gamma = \gamma \cdot \big((x/\gamma) \,\%\, 1\big). \tag{\%}$$

Proof of Theorem 2.27. Let K, ξ_1, $L = K/\xi_1$, and ψ as in Remark 2.26, and let ϕ_K as in (5).

Applying Eqn. (%) twice, first with $\gamma := 1/\xi_1$ and second with $\gamma := 2L$, we find that for all $j \in \mathbb{Z}$,

$$\begin{aligned}((j+1/2)/2K) \,\%\, (1/\xi_1) &= (((j+1/2)/2L) \,\%\, 1)/\xi_1 \\ &= ((j+1/2) \,\%\, (2L))/2L\xi_1 = ((j+1/2) \,\%\, (2L))/2K. \quad (*)\end{aligned}$$

Now we calculate

$$\begin{aligned}(\square \,\%\, (1/\xi_1))(\phi_K) &= 2K \cdot \sum_{j\in\mathbb{Z}} \frac{(-1)^{j+1}}{\pi\,(j+1/2)^2}\, \delta_{((j+1/2)/2K)\%(1/\xi_1)} \\ &= 2K \cdot \sum_{j\in\mathbb{Z}} \frac{(-1)^{j+1}}{\pi\,(j+1/2)^2}\, \delta_{((j+1/2)\%(2L))/2K} && \text{[by } (*)\text{]} \\ &= 2K \cdot \sum_{\substack{a\in 1/2+\mathbb{Z} \\ -L+1/2\le a\le L-1/2}} \delta_{a/2K} \cdot \sum_{\substack{j\in\mathbb{Z} \\ (j+1/2)\%(2L)=a}} \frac{(-1)^{j+1}}{\pi\,(j+1/2)^2}.\end{aligned}$$

Comparing the last expression with ψ, we find that we have to show, for all $a \in 1/2 + \mathbb{Z}$ with $-L + 1/2 \le a \le L - 1/2$, that the following holds:

$$\sum_{\substack{j\in\mathbb{Z} \\ (j+1/2)\%(2L)=a}} \frac{(-1)^{j+1}}{\pi\,(j+1/2)^2} \overset{!}{=} \frac{\pi}{(2L)^2} \cdot \frac{(-1)^{a+1/2}}{\sin^2(\pi a/2L)}. \qquad (**)$$

Replacing the implicit condition under the sum by an explicit one (i.e., we define ℓ through $j + 1/2 = a + \ell \cdot 2L$), we obtain, on the LHS of (**),

$$\sum_{\ell\in\mathbb{Z}} \frac{(-1)^{a+1/2+\ell\cdot 2L}}{\pi\,(a+\ell\cdot 2L)^2} = \sum_{\ell\in\mathbb{Z}} \frac{(-1)^{a+1/2}}{\pi\,(a+\ell\cdot 2L)^2}.$$

Replacing this on the LHS of (**) and, isolating $\pi^2/\sin^2(\pi a/2L)$ on the RHS of (**), on the LHS we are left with

$$\sum_{\ell\in\mathbb{Z}} \frac{(-1)^{2(a+1/2)}}{(a+\ell\cdot 2L)^2/(2L)^2} = \sum_{\ell\in\mathbb{Z}} \frac{1}{(a/2L+\ell)^2}.$$

We recognize the partial fraction decomposition of $\pi^2/\sin^2(\pi z)$, for $z := a/2L$, as in (33) above, which completes the proof. □

7.2.b Proof of the fundamental folding lemma

Proof of Lemma 2.29. Writing out as integrals, we calculate for $f \in \mathscr{K}_\Xi$, $x \in \mathbb{R}$, for Item (a),

$$\begin{aligned}f(x - \tau(s)) &= f(x - \tau(s) \,\%\, p) && \text{[as } f(x-\square) \text{ is } p\text{-periodic]} \\ &= f(x - s \,\%\, p) && \text{[as } \tau \text{ is a } p\text{-folding]} \\ &= f(x - s) && \text{[as } f(x-\square) \text{ is } p\text{-periodic;]}\end{aligned}$$

and for Item (b),

$$\begin{aligned}f(\tau(x - s)) &= f(\tau(x - s) \,\%\, p) && \text{[as } f \text{ is } p\text{-periodic]} \\ &= f((x - s) \,\%\, p) && \text{[as } \tau \text{ is a } p\text{-folding]} \\ &= f(x - s) && \text{[as } f \text{ is } p\text{-periodic.]}\end{aligned}$$

The statements about the derivative then follows if ϕ is feasible for $\mathscr{K}_\Xi$. □

7.2.c Proof of Prop. 2.31

Throughout this section we use the notation of Prop. 2.31. For easier readability, we abbreviate $\phi := \phi_K$ for the Nyquist shift rule (5), and $\tau := \tau_{p,c}$.

Proof of Item (a) of Prop. 2.31. Let f, f_0, f_1 and (c, C) as stated. By linearity of (22) and the derivative, followed by the triangle inequality we find, for all $x \in \mathbb{R}$,

$$\begin{aligned}\big|(f\circ\tau)*\phi(x) - f'(x)\big| &\\ \leq \big|(f_0\circ\tau)*\phi(x) - f_0'(x)\big| &+ \big|(f_1\circ\tau)*\phi(x) - f_1'(x)\big| \\ = \big|(f_0\circ\tau)*\phi(x) - f_0'(x)\big|,& \quad (*)\end{aligned}$$

where the last equation arises from the Fundamental Folding-Lemma 2.29(b).

Hence, we only have to bound the error of the linearly decaying part, f_0, and we use a refined version of the inequalities in Remark 2.17, in a setting similar to the proof of Lemma 2.24: As $f_0 \in \mathscr{F}_K$, by Corollary 2.16, we can replace f_0' by $f_0 * \phi$, replacing and upper-bounding the last expression in $(*)$ as follows

$$\big|(f_0\circ\tau - f_0)*\phi(x)\big| \leq \int |f_0(\tau(x-s)) - f_0(x-s)|\, d|\phi|(s). \qquad (**)$$

We split the last integral into a sum $I_0 + I_- + I_+$ by distinguishing three non-overlapping cases on $x - s$ as follows (using Lemma 2.30):

Symbol	$x-s$ interval	$\tau(x-s)$ interval	s interval
I_+	$]-\infty, -c-p]$	$]-c-p, -c]$	$[x+(c+p), +\infty[$
I_0	$]-c-p, +c+p[$	$]-c-p, +c+p[$	$]x-c-p, x+c+p[$
I_-	$[c+p, \infty[$	$[c, c+p[$	$]-\infty, x-(c+p)]$.

We go through them one by one, starting with the easiest: On the interval $]-c-p, +c+p[$, we have $f_0 \circ \tau = f_0$, so that $I_0 = 0$.

I_-, i.e., assume $x - s \geq c + p$. We start by bounding

$$\int_{-\infty}^{x-(c+p)} |f_0(\tau(x-s)) - f_0(x-s)|\, d|\phi|(s) \leq \|f_0\circ\tau - f_0\|_{\infty,[c+p,\infty[} \cdot |\phi|(]-\infty, x-(c+p)])$$

As f_0 is a (c, C)-linearly decaying function, we can bound

$$\|f_0\circ\tau - f_0\|_{\infty,[c+p,\infty[} \leq 2C/c.$$

For the $|\phi|$-part, we distinguish between $x \in [-p, +p]$ and arbitrary x.

- For arbitrary x, we upper-bound by $|\phi|(]-\infty, x-(c+p)]) \leq |\phi|(\mathbb{R}) = \|\phi\| = 2\pi K$ by Theorem 2.21.
- For $x \in [-p, +p]$, we can do more. Noting that $x - (c+p) \leq -c$ as $x \leq p$, we find by the definition of ϕ in (5),

$$\begin{aligned}&|\phi|(]-\infty, x-(c+p)]) \\ &\qquad \leq |\phi|(]-\infty, -c]) = 2K \sum_{\substack{a\in 1/2+\mathbb{Z}\\ a/2K\leq -c}} \frac{1}{\pi\, a^2} \leq \frac{2K}{\pi} \int_{2Kc-1}^{\infty} \frac{dx}{x^2} = \frac{2K}{\pi\,(2Kc-1)},\end{aligned}$$

where we have relied on $cK \geq 1$, Eqn. (23).

I_+**:** The case is symmetric, yielding the same bound.

Adding the three terms we find that the RHS in (⁎⁎) is

$$I_0 + I_- + I_+ \leq 2 \cdot \frac{2C}{c} \cdot \begin{cases} \frac{2K}{\pi\,(2Kc-1)}, & \text{if } x \in [-p,+p]; \\ 2\pi K, & \text{otherwise.} \end{cases}$$

This concludes the proof of Prop. 2.31(a). □

Proof of Item (b) of Prop. 2.31. We have to find the supremum of the number $\|(f\circ\tau)*\phi\|_\infty$, where f ranges over all real-valued measurable functions with $\|f\|_\infty \leq 1$.

We will apply Remark 2.17, to the restriction $\phi^{(c)}$ of ϕ to the image of τ, which, by Lemma 2.30, is $]-c-p,+c+p[$.

$$\begin{aligned}\|(f\circ\tau)*\phi\|_\infty &= \|(f\circ\tau)*\phi^{(c)}\|_\infty \\ &\leq \|f\circ\tau\|_\infty \cdot \|\phi^{(c)}\| \leq \|f\|_\infty \cdot \|\phi^{(c)}\| \leq \|\phi^{(c)}\| = \|\phi\| - \Theta_K(1/c).\end{aligned}$$

where we have used the notation $\Theta(\centerdot) = O(\centerdot) \cap \Omega(\centerdot)$.

This concludes the proof of Prop. 2.31(b). □

Proof of Item (c) of Prop. 2.31. This follows directly from the statement about the support of τ in Lemma 2.30. □

Proof of Item (d) of Prop. 2.31. We have to give an upper bound to the quantity

$$E(x) := \frac{1}{\|\phi\|} \int \big|\tau(x-s)\big|\, d|\phi|(s),$$

over $x \in [-p,+p]$.

We split the value into three summands by distinguishing the non-overlapping cases

$$x - s \leq -c - p, \qquad -c-p < x-s < c+p, \qquad c+p \leq x-s,$$

and upper-bound each separately.

For the case $x - s \leq -c - p$ we upper bound as follows (recall from Theorem 2.21 that $\|\phi\| = 2\pi K$):

$$\begin{aligned}\frac{1}{\|\phi\|} \cdot \int\limits_{x+(c+p)}^{\infty} \big|\tau(x-s)\big|\, d|\phi|(s) &\leq \frac{c+p}{\|\phi\|} \cdot \int\limits_{x+(c+p)}^{\infty} d|\phi| \\ &\leq \frac{c+p}{\|\phi\|} \cdot \int\limits_{c}^{\infty} d|\phi| && [x \in [-p,+p]] \\ &= \frac{c+p}{2\pi K} \cdot 2K \sum_{\substack{a\in 1/2+\mathbb{Z} \\ a \geq 2Kc}} \frac{1}{\pi\, a^2} \\ &\leq \frac{c+p}{2\pi K} \cdot \frac{2K}{\pi\,(2Kc-1)} \\ &= \frac{1+p/c}{2\pi^2 K - \pi^2/c)} \\ &\leq \frac{1+p/c}{(2\pi^2 K - \pi^2 K)} && [cK \geq 1, \text{ Eqn. (23)}] \\ &= \frac{1+p/c}{\pi^2 K} \\ &= \frac{2}{\pi^2 K}. && [p \leq c]\end{aligned}$$

The case $c+p \le x-s$ is symmetrical and gives the same bound.

For the case $-c-p < x-s < +c+p$ we upper bound as follows:

$$\frac{1}{\|\phi\|}\cdot\int \mathbf{1}_{\{-c-p<x-\square<c+p\}}(s)\big|\tau(x-s)\big|\,d|\phi|(s) \le \frac{1}{\|\phi\|}\cdot\int_{x-(c+p)}^{x+c+p}|x-s|\,d|\phi|(s)$$
$$\le \frac{1}{\|\phi\|}\cdot\int_{x-(c+p)}^{x+c+p}|x|\,d|\phi|(s)+\frac{1}{\|\phi\|}\cdot\int_{x-(c+p)}^{x+c+p}|s|\,d|\phi|(s).$$

The first integral we upper-bound by $|x|\cdot\int d|\phi|$, so the first summand is upper bounded by $|x|$. For the second summand, invoking Theorem 2.21 again, we continue

$$\frac{1}{\|\phi\|}\cdot\int_{x-(c+p)}^{x+c+p}|s|\,d|\phi|(s) = \frac{1}{2\pi K}\cdot 2K \sum_{\substack{a\in 1/2+\mathbb{Z}\\ x-(c+p)\le a/2K\le x+c+p}}\frac{a/2K}{\pi\,a^2}$$
$$= \frac{1}{2\pi^2 K}\cdot\sum_{\substack{a\in 1/2+\mathbb{Z}\\ x-(c+p)\le a/2K\le x+(c+p)}}\frac{1}{a} \underset{(*)}{\le} \frac{1}{2\pi^2 K}\cdot\sum_{\substack{a\in 1/2+\mathbb{Z}\\ -c-2p\le a/2K\le c+2p}}\frac{1}{a}$$
$$\le \frac{1}{2\pi^2 K}\cdot 2\cdot\big(\ln(2K(c+2p))+2\big) = \frac{\ln(K(c+2p))+2+\ln 2}{\pi^2 K}.$$

In the inequality marked with the asterisk, we have used the assumption $x\in[-p,+p]$.

Adding everything up, we get, as the upper bound on the expected parameter value

$$2\cdot\frac{2}{\pi^2 K}+|x|+\frac{\ln(K(c+2p))+2+\ln 2}{\pi^2 K} = |x|+\frac{\ln(K(c+2p))+6+\ln 2}{\pi^2 K},$$

as claimed. □

This concludes the proof of Prop. 2.31.

8 Sketches of proofs of non-existence results

8.1 Impossibility of exponential concentration

8.1.a Background on exponential concentration

We review some technicalities about exponential concentration. The proof of the following lemma is purely technical and given in Appendix A.3 for the sake of completeness.

Lemma 8.1. *For complex measures μ the following are equivalent:*

(a) μ is exponentially concentrated;

(b) There exists an $r>0$ with $\int e^{r|x|}\,d|\mu|(x)<\infty$.

The following is easily derived from Lemma 8.1 by tracing the proof of Theorem 3 in [27].

Lemma 8.2. *Let μ be a complex measure on the real line. If μ is exponentially concentrated then $\hat{\mu}$ can be extended to a holomorphic function defined in a neighborhood of the real line.*

8.1.b Proof of Theorem 2.33

Theorem 2.33 is proven with the same complex analysis argument as the result for compact support in §2.9.b, but replacing Paley-Wiener with Lemma 8.2.

Hence, the only challenging part of the proof is that we do not assume that Ξ contains an accumulation point, i.e., an element $\xi \in \Xi$ that is in the closure of $\Xi \setminus \{\xi\}$.

But with Ξ the set $\bar{\Xi}$ also satisfies the condition of the theorem. This can be seen in two ways. *(1)* By first extending the proof of Lemma 2.12 to the case where the condition "$\xi \in [-1/2, +1/2]$" of equation (13) is replaced with "$\xi \in \Xi$", and then noting that both sides of equation (13) are continuous functions of ξ, so the equality extends to limit points. *(2)* Directly from (12), with firstly $g := \cos(2\pi\xi\,\square)$, $\xi \in \Xi$, and secondly $g := \sin(\pi\xi)$, $\xi \in \Xi$, by again noting that in all cases both sides of the equation are continuous functions of ξ.

8.2 Shift rules with approximation error

Sketch of proof of Corollary 2.34. For a proof by contradiction, assume that C, r, ϕ_* exist.

We invoke Alaoglu's theorem: "The dual unit ball is weak* compact." The space of complex measures is the dual of $\mathscr{C}_0$, the space of (complex valued) continuous functions vanishing in $\pm\infty$, and the total-variation norm is the operator norm on the dual (e.g., [28, Theorem 6.19]).

As $\{\phi_j \mid j \in \mathbb{N}\}$ is norm bounded, by passing to a sub-sequence of $(\phi_j)_j$, there exists a complex measure ϕ such that

$$\int h\, d\phi = \lim_{j\to\infty} \int h\, d\phi_j \qquad \text{for all } h \in \mathscr{C}_0.$$

The proof of the corollary is completed by showing that (a) ϕ is exponentially concentrated, and (b) ϕ is feasible for $\mathscr{G}$. The combination of (a) and (b) contradicts Theorem 2.33.

The arguments for (a,b) are entertaining exercises in calculus and measure theory, and we leave them to the bored reader. □

9 Discussion and outlook

We have given a "proper" shift rule for analytic derivatives of "perturbed-parametric" quantum evolutions. The support of the measure is non-compact, and the tail, i.e., the fraction of the total variation that falls outside a given interval $[-T, +T]$, $T \to \infty$, decays only linearly, as in a Cauchy-type probability distribution.

The question arises whether proper shift rules for analytic derivatives of perturbed-parametric quantum evolutions exist that have lighter tails. Given that we have also shown that the space of proper shift rules that give analytic derivatives has infinite dimension, at first sight, that doesn't appear impossible.

However, we have also shown that proper shift rules (for perturbed-parametric quantum evolutions) with exponentially decaying tails do not exist. Indeed, the negative result holds for norm-bounded (i.e., worst-case variance bounded) families of proper shift rules that approximate the derivative with an approximation error that tends to 0.

On the positive side, the best result we have is a $O(1/T^2)$ approximation error, for parameter values that are restricted to $[-T, +T]$ (this, by nature, cannot be "proper" shift rule).

In future research, it might be interesting to either expand the non-existence results, or to give shift rules with faster decaying tails.

Acknowledgments

The author would like to thank the anonymous referees for their careful reading of the manuscript and valuable suggestions for improving it. The author also thanks L. Savanna for discussions about the draft paper. This research was partly funded by the Estonian Center of Excellence in

Computer Science (EXCITE), and by Estonian Research Agency (Eesti Teadusagentuur, ETIS) through grant PRG946.

APPENDIX

A Miscellaneous math

A.1 Sums of reciprocals of squares of odd integers

The following is a direct consequence of the *Basel Problem*, $\sum_{n=1}^{\infty} \frac{1}{n^2} = \pi^2/6$.

Lemma A.1. $\displaystyle\sum_{j=0}^{\infty} \frac{1}{(2j+1)^2} = \pi^2/8$

Proof. In the following equations, the first and last equations are the Basel Problem:

$$\pi^2/6 = \sum_{n=1}^{\infty} \frac{1}{n^2} = \sum_{n=1}^{\infty} \frac{1}{(2n+1)^2} + \sum_{n=1}^{\infty} \frac{1}{(2n)^2} = \sum_{n=1}^{\infty} \frac{1}{(2n+1)^2} + \frac{1}{4} \cdot \pi^2/6.$$

Collecting the non-sum terms on the LHS gives the statement of the lemma. □

A.2 Simultaneous Diophantine approximation

Recall Dirichlet's simultaneous diophantine approximation theorem.

Theorem A.2 (SDA)**.** *For all positive integers d, m and real numbers $\xi_1, \ldots, \xi_d$, there exists a positive integer $q \geq m$ and integers $k_1, \ldots, k_d$ such that*

$$\left| \xi_j - \frac{k_j}{q} \right| \leq \frac{1}{q\, m^{1/d}}$$

We use it to prove the following fact.

Lemma A.3. *Let*

$$f \colon x \mapsto \sum_{j=1}^{d} e^{2\pi i \xi_j x}\, b_j$$

be a bounded real-valued, complex-valued, or operator-valued function with finite Fourier spectrum, i.e., $\xi \in \mathbb{R}^d$ and $b_1, \ldots, b_d$ are real or complex numbers, or operators, respectively.

If $\lim_{|x|\to\infty} f(x) = 0$, then $f = 0$.

Proof. The lemma is trivial if the frequencies $\xi_1, \ldots \xi_d$ have a common divisor, as then the function f is periodic.

In the general case, let f be as described, and set

$$c := \|f\|_{\infty,[-1,0]} := \sup_{x\in[-1,0]} \|f(x)\|,$$

where $\|.\|$ stands for any norm.

Assume $f \neq 0$; this implies $c > 0$ (as f can be extended to a holomorphic function on the complex plane). We prove that for every $T > 0$ we have $\|f\|_{\infty,[T,\infty[} \geq c/2$. This implies that $\lim_{|x|\to\infty} f(x) = 0$ does not hold.

For given $T > 0$, let $m \geq T$ be a integer with $m^{1/d} \geq 8\pi \|b\|_1 / c$ (where $\|b\|_1 := \sum_{j=1}^{d} \|b_j\|$).

Now apply Theorem A.2 to the numbers $\xi_1, \ldots, \xi_d$, and m, d as just defined, and let q be the denominator and $k_1, \ldots, k_d$ the numerators in the Diophantine approximations. We find that, for all $j = 1, \ldots, d$,

$$2\pi \cdot \left| \xi_j - \tfrac{k_j}{q} \right| \cdot q \leq \frac{2\pi}{m^{1/d}} \leq \frac{c}{4\|b\|_1}, \tag{34}$$

by our choice of m. It follows that, for all $x \in [-1, 0]$, we have

$$\begin{aligned}\|f(x) - f(x+q)\| &\leq \sum_{j=1}^{d} \left| e^{2\pi i \xi_j x} - e^{2\pi i \xi_j (x+q)} \right| \cdot \|b_j\| \\ &\leq \max_{j=1,\dots,d} \left| e^{2\pi i \xi_j x} - e^{2\pi i \xi_j (x+q)} \right| \cdot \|b\|_1 \end{aligned} \tag{$*$}$$

For $j = 1, \dots, d$, we continue

$$\begin{aligned}&\left| e^{2\pi i \xi_j x} - e^{2\pi i \xi_j (x+q)} \right| \\ &\qquad \leq \left| e^{2\pi i \xi_j x} - e^{2\pi i \frac{k_j}{q} x} \right| + \left| e^{2\pi i \xi_j (x+q)} - e^{2\pi i \frac{k_j}{q} (x+q)} \right| + \left| e^{2\pi i \frac{k_j}{q} x} - e^{2\pi i \frac{k_j}{q} (x+q)} \right| \end{aligned} \tag{$**$}$$

The third term on the RHS in ($**$) is 0.

We treat the first two terms in ($**$) simultaneously by setting $s := x$ or $s := x+q$, respectively, noting $|s| \leq q$. We find, using (34),

$$\left| e^{2\pi i \xi_j s} - e^{2\pi i \frac{k_j}{q} s} \right| \leq 2\pi \cdot \left| \xi_j - \tfrac{k_j}{q} \right| \cdot q \leq c/4 \|b\|_1.$$

Hence, we can upper-bound the RHS in ($**$) by $c/2\|b\|_1$. We continue ($*$):

$$\|f(x) - f(x+q)\| \leq c/2 \|b\|_1 \cdot \|b\|_1 \leq c/2,$$

and hence, for some $x \in [-1, 0]$ we have $\|f(x+q)\| \geq c/2$. □

Proposition A.4. *The following sum of vector spaces is direct: (1) the vector space of all bounded (real, complex or operator-valued) functions with finite Fourier spectrum (each function can have its own Fourier spectrum, but it must be finite) and (2) the vector space of (real, complex or operator-valued, resp.) functions that vanish in $\pm\infty$.*

In other words, if f_1, f_0, g_1, g_0 are (real, complex or operator-valued) functions which satisfy

- $f_1 + f_0 = g_1 + g_0$;
- $\lim_{|x|\to\infty} f_0(x) = 0$, $\lim_{|x|\to\infty} g_0(x) = 0$;
- *Each of f_1, g_1 has a finite Fourier spectrum, both are bounded;*

then $f_1 = g_1$ and $f_0 = g_0$ hold.

Proof. The first statement a direct consequence of Lemma A.3; the second statement follows by grouping terms. □

A.2.a Bounded functions with finite Fourier spectrum

We have used the following well-known fact all the time.

Proposition A.5. *Let f be a measurable (complex-valued) function. If f is bounded and has finite Fourier spectrum, then f is a (finite) linear combination of Fourier characters $e^{2\pi i \xi \,\square}$, $\xi \in \mathbb{R}$.*

The tempered-distribution fact behind it is the following: If a tempered distribution has finite support $\Xi \subset \mathbb{R}$, then τ is a linear combination of derivatives[36] $\partial_\xi^{\alpha_\xi}$, $\xi \in \Xi$, with $\alpha \in (\mathbb{Z}_+)^\Xi$. For $\tau := \hat{f}$, that makes f have the form

$$f(x) = \sum_{\xi \in \Xi} x^{\alpha_\xi} \cdot e^{2\pi i \xi x} \cdot b_\xi.$$

Elementary considerations about polynomials plus an argument similar to the one in the proof of Lemma A.3 show that such a function is bounded only if $\alpha_\xi = 0$ for all $\xi \in \Xi$.

[36] Defined traditionally as the αth derivative of the Dirac distribution: $\langle \varphi : \partial_a^\alpha \rangle = (-1)^\alpha \langle \partial^\alpha \varphi : \delta_a \rangle$ for all Schwartz functions φ.

A.3 Exponential concentration

Proof of Lemma 8.1. Let $\mu := |\nu|$ (or, simply, any finite positive measure).

Firstly, suppose we have $C, r > 0$ with $\nu(\mathbb{R} \setminus [-T, +T]) \leq C \cdot e^{-rT}$ for all $T \geq 0$. Recall from standard measure theory (e.g., [29, Satz 23.8]) that for all measurable functions $f \geq 0$ we have

$$\int f \, d\nu = \int_0^\infty \nu(\{f \geq y\}) \, dy.$$

From this, with $f := e^{r|\,\square\,|}$, we find, with $r' := r/2$:

$$\int e^{r'|x|} \, d\nu = \int_0^\infty \nu(\{e^{r'|\,\square\,|} \geq y\}) \, dy$$
$$\leq \|\nu\| + \int_1^\infty \nu(\{|\,\square\,| \geq \ln(y)/r'\}) \, dy \leq \|\nu\| + \int_1^\infty C y^{-2} \, dy < \infty.$$

Secondly, suppose we have $r > 0$ such that $C := \int e^{r|\,\square\,|} \, d\nu < \infty$. We calculate

$$\nu(\mathbb{R} \setminus [-T, +T]) = \int \mathbf{1}_{\mathbb{R}\setminus[-T,+T]} \, d\nu = e^{-rT} \int e^{rT} \mathbf{1}_{\mathbb{R}\setminus[-T,+T]} \, d\nu \leq e^{-rT} \int e^{r|\,\square\,|} \, d\nu \leq C \cdot e^{-rT}.$$

This completes the proof of the lemma. □

B Finite-dimensional perturbation theory (for the theoretical computer scientist)

We use the notations laid out in §3.3; in addition, in this section, e is just another letter (and not Euler's number $\exp(1)$).

The mathematical cornerstone of finite-dimensional perturbation theory is *Rellich's Theorem* [30, Theorem 1]. We give here a much simplified version adapted to our needs.

We employ the terminology of holomorphic functions on a complex disc: For $r > 0$, we denote by $\mathbb{D}_r$ the open disc of radius r around 0 in $\mathbb{C}$.

Theorem B.1 (Rellich)**.** *Let A, B be Hermitian operators on a Hilbert space $\mathcal{H}$ of dimension $d \in \mathbb{N}$, and consider $z \mapsto C(z) := A + zB$; where $z \in \mathbb{C}$. Denote the (distinct) eigenvalues of A by $\lambda_1^{(0)}, \ldots, \lambda_e^{(0)}$, and by h_ℓ the multiplicity of eigenvalue $\lambda_\ell^{(0)}$ of A, $\ell = 1, \ldots, e$.*

There exists a radius $r > 0$ such that the following holds.

- *For each $\ell = 1, \ldots, e$, $j = 1, \ldots, h_\ell$, there exists a holomorphic function $\lambda_{(\ell,j)} \colon \mathbb{D}_r \to \mathbb{C}$, such that*
  - *(a) For each $z \in \mathbb{D}_r$, the eigenvalues of $C(z)$, listed with repetitions according to algebraic multiplicities, are $\lambda_{\ell,j}(z)$, $\ell = 1, \ldots, e$, $j = 1, \ldots, h_\ell$.*
  - *(b) For each ℓ, j, we have $\lambda_{\ell,j}(0) = \lambda_\ell^{(0)}$.*
- *For each $\ell = 1, \ldots, e$, $j = 1, \ldots, h_\ell$, there exists a holomorphic (vector-valued) function $w_{\ell,j} \colon \mathbb{D}_r \to \mathcal{H}$, such that*
  - *(c) For each $z \in \mathbb{D}_r$, for each ℓ, j, the vector $w_{\ell,j}(z)$ is eigenvector for the eigenvalue $\lambda_{\ell,j}(z)$ of $C(z)$, i.e., $C(z)w_{\ell,j}(z) = \lambda_{\ell,j}(z)w_{\ell,j}(z)$ (and $w_{\ell,j}(z) \neq 0$);*
  - *(d) For each $z \in \mathbb{D}_r \cap \mathbb{R}$, the vectors $w_{\ell,j}(z)$, where $\ell = 1, \ldots, e$, $j = 1, \ldots, h_\ell$, form an orthonormal basis of $\mathcal{H}$.*

Based on Theorem B.1 the usual undergrad quantum physics perturbation theory can be executed mathematically rigorously in finite dimension; the result is shown in the following Corollary B.2 — in matrix notation (i.e., for the theoretical computer scientist).

Corollary B.2. *Let A, B be Hermitian operators on a Hilbert space $\mathcal{H}$ of finite dimension $d \in \mathbb{N}$. There exists an $r > 0$ and an orthonormal basis of eigenvectors of A, along with d-by-d matrices $\mathtt{A}, \mathtt{B}$, $\mathtt{W}^{[k]}$, $k \in \mathbb{Z}_+$, and diagonal d-by-d matrices $\Lambda^{[k]}$, $k \in \mathbb{Z}_+$, such that the following holds.*

(a) $\mathtt{A}$ ($\mathtt{B}$, resp.) is the matrix of the operator A (B, resp.) wrt the ONB;
(b) The matrix-valued power series $\mathtt{W}(z) = \sum_{k=0}^{\infty} z^k \mathtt{W}^{[k]}$ has convergence ratio at least r;
(c) $\mathtt{W}(z)$ is unitary for all $z \in]-r, +r[$;
(d) The diagonal-matrix-valued power series $\Lambda(z) = \sum_{k=0}^{\infty} z^k \Lambda^{[k]}$ has convergence ratio at least r;
(e) $\mathtt{A} + z\mathtt{B} = \mathtt{W}(z)\Lambda(z)\mathtt{W}(z)^\dagger$ holds for all $z \in]-r, +r[$;
(f) $\mathtt{W}^{[0]} = \mathtt{W}(0) = \mathtt{I}$ (identity matrix);
(g) $\Lambda^{[0]} = \Lambda(0) = \mathtt{A}$;
(h) $\Lambda^{[1]} = \mathtt{Diag}\, \mathtt{B}$;

The proof of Corollary B.2 is well known; we sketch it. The matrix-valued function $\mathtt{W}$ is defined, for $\ell', \ell = 1, \dots, e$, $j' = 1, \dots, h_{\ell'}$, $j = 1, \dots, h_\ell$, and all small enough $|z|$, by

$$\mathtt{W}_{(\ell',j'),(\ell,j)}(z) := (w_{\ell,j}(0) \mid w_{\ell',j'}(z));$$

where $(. \mid .)$ is the inner product in the Hilbert space $\mathcal{H}$, and the $w_*(.)$ are from Rellich's Theorem B.1. In other words, the columns of $\mathtt{W}(z)$ are the orthonormal eigenstates at z, wrt the basis at 0. The diagonal-matrix valued fuction Λ has on its diagonal the eigenvalue-valued functions λ from Rellich's theorem, in the corresponding ordering. Now, just expand the power series in z on both sides of the equation

$$(\mathtt{A} + z\mathtt{B})\mathtt{W}(z) = \mathtt{W}(z)\Lambda(z)$$

and collect terms. Hence, there is no need to repeat the proof here.

C Fourier-analytic characterization of derivative-computing measures, general case

In §2.4.a, we have proven one direction of Lemma 2.12, we now prove the remaining one in full generality: That (13) implies (12).

Apart from standard calculations with Fourier(-Stieltjes) transforms and tempered distributions (most of which we are writing out in this section), the proof is based on the technique in the Paley-Wiener-Schwartz theorem that is contained in the following fact.

Lemma C.1 (Paley-Wiener-Schwartz [22, 19])**.** *Let τ be a tempered distribution with compact support, and ψ a Schwartz function with compact support such that $\psi = 1$ on $\operatorname{Supp}\tau$. The tempered distribution $\check{\tau}$ (inverse Fourier transform) is given by integrating against the following function*[37]

$$x \mapsto \langle \psi \cdot e^{2\pi i \,\square\, x} : \tau \rangle,$$

which has at most polynomial growth and is analytic.

In this section we arbitrarily fix a Schwartz function ψ with compact support satisfying

$$\begin{aligned} 0 \le \psi \le 1 \quad & \text{everywhere,} \\ \psi = 1 \quad & \text{on } [-1/2, +1/2], \text{ and} \\ \psi = 0 \quad & \text{outside of } [-1, +1]. \end{aligned}$$

Let us pull out some lemmas for easier readability.

[37] Note that $\psi \cdot e^{2\pi i \,\square\, x}$ is a Schwartz function with compact support, for every $x \in \mathbb{R}$.

Lemma C.2. *Let f be a smooth function of at most polynomial growth which has compact Fourier spectrum. For all $x \in \mathbb{R}$ we have*

$$f(x) = \langle \psi \cdot e^{2\pi i \,\square\, x} : \hat{f} \rangle.$$

Proof. From Lemma C.1 and the Fourier Inversion Theorem for tempered distributions, we know that the RHS defines a function of x, integration against which defines a tempered distribution that is equal to the tempered distribution defined by integrating against f. As both f and (again by Lemma C.1) the function defined by the RHS are continuous, pointwise equality follows. □

Lemma C.3. *If f is a smooth function of at most polynomial growth, such that $\hat{f}$ has compact support, then f' has at most polynomial growth. Moreover,*

(a) $f'(x) = \langle \psi \cdot e^{2\pi i \,\square\, x} : \widehat{f'} \rangle$ *for all $x \in \mathbb{R}$, and*
(b) $\widehat{f'} = (2\pi i \,\square)\, \hat{f}$.

Proof. Let f as stated, and σ be the tempered distribution "integrating against f". The tempered distribution $\partial\hat{\sigma}$ has support contained in $\operatorname{Supp} \hat{\sigma}$. Indeed, if φ is a Schwartz function with support in the complement of $\operatorname{Supp} \hat{\sigma}$, when so is $\partial\varphi$, so that

$$\langle \varphi : \partial\hat{\sigma} \rangle = -\langle \partial\varphi : \hat{\sigma} \rangle = 0.$$

So Lemma C.1 is applicable: There exists a continuous function g of at most polynomial growth such that $\partial\sigma$ is given by integrating against g. But for all compactly supported Schwartz functions φ we have

$$\int \varphi f' = -\int \partial\varphi f = -\langle \partial\varphi : \sigma \rangle = \langle \varphi : \partial\sigma \rangle. = \int \varphi g,$$

so that $f' = g$ holds pointwise. This proves that f' is of at most polynomial growth, and we can apply Lemma C.2 to f', which gives us ((a)).

With ((a)), the equation ((b)) is purely a relation between tempered distributions and follows from $(\partial\sigma)^\wedge = (2\pi i \,\square)\, \hat{\sigma}$. □

Lemma C.4. *Let μ be a finite measure. For every Schwartz-function φ, we have*

(a) For all $x \in \mathbb{R}$,

$$\left|\langle \varphi : \psi e^{-2\pi i \,\square\, x} \rangle\right| \leq \|\varphi\|_\infty \cdot \|\psi\|_1,$$

and hence $x \mapsto \langle \varphi : \psi e^{-2\pi i \,\square\, x} \rangle$ is μ-integrable;

(b) $\langle \varphi : \psi\hat{\mu} \rangle = \int \langle \varphi : \psi e^{-2\pi i \,\square\, x} \rangle \, d\mu(x)$.

Proof. For (a), we simply calculate

$$\begin{aligned}
\left|\langle \varphi : \psi e^{-2\pi i \,\square\, x} \rangle\right| &= \left| \int \varphi(\xi)\psi(\xi) e^{-2\pi i \xi x} \, d\xi \right| \\
&\leq \int |\varphi(\xi)|\, |\psi(\xi)|\, |e^{-2\pi i \xi x}| \, d\xi \\
&\leq \|\varphi\|_\infty \int |\psi(\xi)| \, d\xi,
\end{aligned}$$

which gives the bound. For integrability, as μ is finite, in addition to the boundedness of the integrand that we have just proven, we only require measurability, which follows from the usual facts about parameterized integrals (e.g., [28]).

For (b), we simply calculate

$$\begin{aligned}\int \langle \varphi : \psi e^{-2\pi i \,\square\, x}\rangle \, d\mu(x) &= \int\int \varphi(\xi)\,\psi(\xi)\, e^{-2\pi i\xi x}\, d\xi\, d\mu(x)\\ &= \int\int \varphi(\xi)\,\psi(\xi)\, e^{-2\pi i\xi x}\, d\mu(x)\, d\xi \qquad \text{[Fubini with (a)]}\\ &= \int \varphi(\xi)\ \psi(\xi)\cdot \int e^{-2\pi i\xi x}\, d\mu(x)\ d\xi\\ &= \langle \varphi : \psi\hat{\mu}\rangle,\end{aligned}$$

and the proof of the lemma is completed. □

Now we address the task of this subsection.

Proof of sufficiency in Lemma 2.12. Let μ be a finite measure, and $f \in \mathscr{F}_{1/2}$. We have to prove that if μ satisfies the condition (13) in Lemma 2.12, then f, μ satisfy the equation in (12): $f'(1/2) = \int f\, d\mu$.

Define a sequence of Schwartz functions $f_j := \psi(\square/2j)\cdot f$, $j \in \mathbb{N}$, and note that, for all Schwartz functions φ, we have

$$\big|\langle \varphi : f - f_j\rangle\big| \le \int\limits_{\mathbb{R}\setminus[-j,+j]} |f|\cdot|\varphi| \xrightarrow{j\to\infty} 0,$$

i.e., $f_j \xrightarrow{j\to\infty} f$ as tempered distributions (i.e., in the $\sigma(\mathcal{S}',\mathcal{S})$ locally-convex topology), and hence, by continuity of the Fourier transform, $\widehat{f_j} \xrightarrow{j\to\infty} \hat{f}$ as tempered distributions.

By Lemma C.4(b), we have, for all j,

$$\langle \widehat{f_j} : \psi\hat{\mu}\rangle = \int \langle \widehat{f_j} : \psi e^{-2\pi i\,\square\, x}\rangle \, d\mu(x). \tag{$*$}$$

We take the limit $j\to\infty$ on both sides of ($*$).

The LHS of ($*$) goes to $\langle \psi\hat{\mu} : \hat{f}\rangle$ as $j\to\infty$, because $f_j \xrightarrow{j\to\infty} f$ as tempered distributions, and $\langle \widehat{f_j} : \psi\hat{\mu}\rangle = \langle \psi\hat{\mu} : \widehat{f_j}\rangle$ (both sides are Schwartz functions).

Taking the limit of the RHS of ($*$), we first note that the expression under the μ-integral goes to $\langle \psi e^{-2\pi i\,\square\, x} : \hat{f}\rangle$ for every $x\in\mathbb{R}$ (same argument as for the LHS). We can take the limit under the μ-integral by the Dominated Convergence Theorem, as, by Lemma C.4(a), we have the $|\mu|$-integrable upper bound

$$\|f\|_\infty\cdot\|\psi\|_1 \ge \|f_j\|_\infty\cdot\|\psi\|_1 \ge \big|\langle \widehat{f_j} : \psi e^{-2\pi i\,\square\, x}\rangle\big| \text{ (for all } x,j).$$

Thus, after taking the limit $j\to\infty$ on both sides of ($*$), we find that

$$\langle \psi\hat{\mu} : \hat{f}\rangle = \int \langle \psi e^{-2\pi i\,\square\, x} : \hat{f}\rangle \, d\mu(x) \tag{$**$}$$

holds.

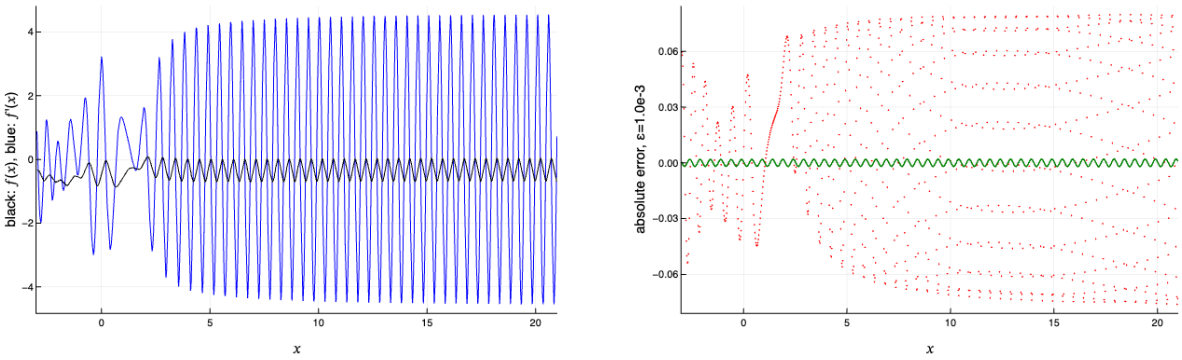
If μ satisfies the condition (13) in the statement of the lemma we are proving, we simply calculate:

$$\begin{aligned}\int f(x)\, d\mu(x) &= \int \langle \psi e^{2\pi i\,\square\, x} : \hat{f}\rangle \, d\mu(x) && \text{[by Lemma C.2]}\\ &= \langle \psi\cdot\hat{\mu}(-\square) : \hat{f}\rangle && \text{[by ($**$)]}\\ &= \langle \psi\cdot 2\pi i\,\square\cdot e^{i\pi\,\square} : \hat{f}\rangle && \text{[condition on } \mu\text{, (13)]}\\ &= \langle \psi\cdot e^{i\pi\,\square} : (2\pi i\,\square)\,\hat{f}\rangle\\ &= \langle \psi\cdot e^{i\pi\,\square} : \widehat{f'}\rangle && \text{[LemmaC.3((b))]}\\ &= \langle \psi\cdot e^{2\pi i\,\square\cdot 1/2} : \widehat{f'}\rangle\\ &= f'(1/2). && \text{[LemmaC.3((a))]}\end{aligned}$$

This completes the proof of Lemma 2.12. □

D Additional graphs from numerical simulations

We give some more graphs of numerical simulations. Figs. 6 and 7 are of the same type as Fig. 4, in that each shows data based on a single expectation-value function. Fig. 8 is of the same type as Fig. 5, in that they show statistics for a collection of 200 random instances.



(a) (b)

Figure 6: **Second example of approximations provided by this paper's STNySR and BC's ASPSR.** Data from a randomly created (and selected because it looks pretty) expectation-value function f is presented (with matrices of size 4×4). Graph (a) shows f (black) and f' (blue), and (b) shows the absolute errors of the ASPSR (red) and the STNySR (green), with $\varepsilon = 10^{-3}$. The shown interval is well within the cut-off interval $[-125, +125]$ of STNySR; in this situation, STNySR typically gives a better approximation than ASPSR.

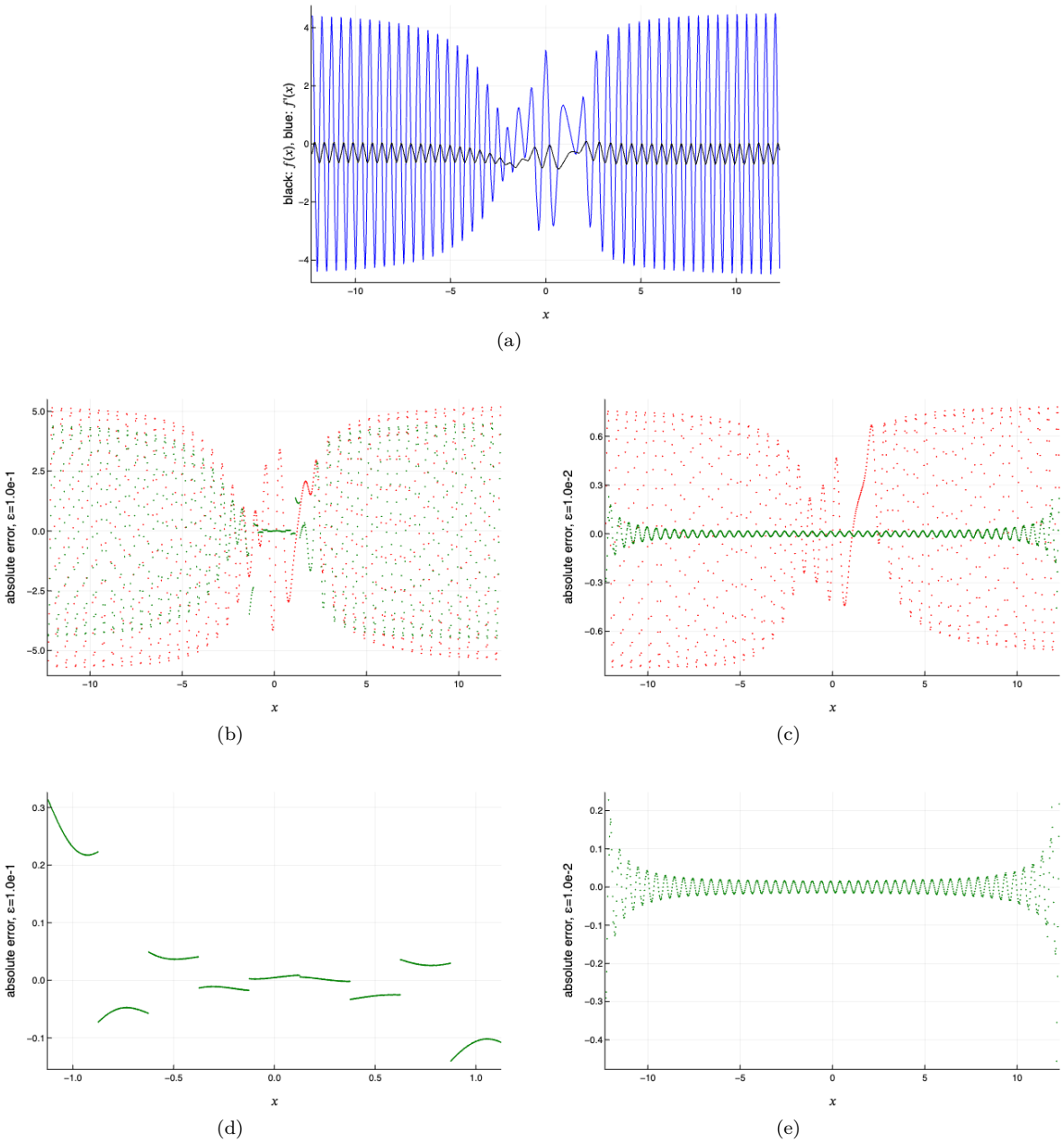


Figure 7: **Second example of approximations provided by this paper's STNySR and BC's ASPSR.** Data from the same expectation-value function f from Fig. 6. Graph (a) shows f (black) and f' (blue), on a larger interval than in Fig. 6. In (b,c,d,e) absolute errors of the ASPSR (red) and the STNySR (green) are plotted, on different parameter intervals. In (b), (d), we took $\varepsilon = 10^{-1}$ (leading to the cut-off $T = 1.25$ for the STNySR) and in (c), (e) we took $\varepsilon = 10^{-2}$ (leading to the cut-off $T = 12.5$).
For $\varepsilon = 0.1$, ASPSR is noise-only, while STNySR delivers what may be called an approximation to the derivative within the interval $[-1, +1] \subset [-T, +T]$.
For $\varepsilon = 0.01$, STNySR provides a clearly superior approximation to the derivative, at almost every point. Graph (e) shows the effect that the error in STNySR blows up near the end-points of the cut-off interval $[-12.5, +12.5]$.

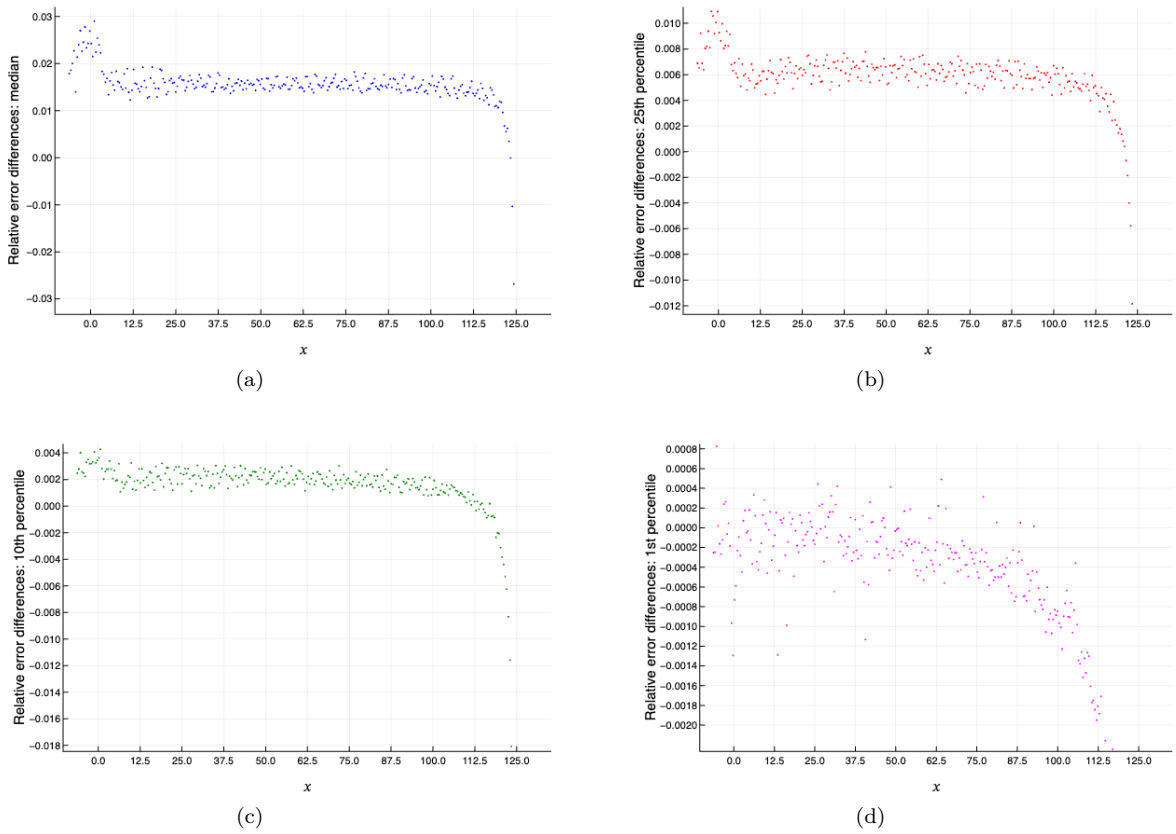


Figure 8: **Another comparison of relative errors between STNySR and BC's ASPSR.** Plots of 200 random instances of expectation-value functions f chosen randomly as described in the text (with 8×8 matrices), considered at 300 parameter values, x. For each combination of parameter value x and expectation-value function f, the difference between the relative error of the ASPSR-approximation (with $\varepsilon = 10^{-3}$) of $f'(x)$ minus the relative error of the STNySR-approximation of $f'(x)$ (with parameter cut-off $1/8\varepsilon = 125$) is computed. For each parameter value x, the median (blue), 25th percentile (red), 10th percentile (green), 1st percentile (magenta) of the 200 random instances are plotted, in different scales. For 90% of the instances, STNySR beats ASPSR by a clear margin over the all but the outside 10% of the parameter region.

References

[1] Jarrod R McClean, Nicholas C Rubin, Joonho Lee, Matthew P Harrigan, Thomas E O'Brien, Ryan Babbush, William J Huggins, and Hsin-Yuan Huang. "What the foundations of quantum computer science teach us about chemistry". The Journal of Chemical Physics **155**, 150901 (2021).

[2] Xiao Yuan, Suguru Endo, Qi Zhao, Ying Li, and Simon C Benjamin. "Theory of variational quantum simulation". Quantum **3**, 191 (2019).

[3] Kosuke Mitarai, Makoto Negoro, Masahiro Kitagawa, and Keisuke Fujii. "Quantum circuit learning". Phys. Rev. A **98**, 032309 (2018).

[4] Marcello Benedetti, Erika Lloyd, Stefan Sack, and Mattia Fiorentini. "Parameterized quantum circuits as machine learning models". Quantum Science and Technology **4**, 043001 (2019).

[5] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. "A quantum approximate optimization algorithm". Preprint (2014).

[6] Eric R Anschuetz, Jonathan P Olson, Alán Aspuru-Guzik, and Yudong Cao. "Variational quantum factoring". Preprint (2018).

[7] Carlos Bravo-Prieto, Ryan LaRose, Marco Cerezo, Yigit Subasi, Lukasz Cincio, and Patrick J Coles. "Variational quantum linear solver". Preprint (2019).

[8] Ryan Babbush and Hartmut Neven. "Training quantum evolutions using sublogical controls" (2019). US Patent 10,275,717.

[9] Louis-Paul Henry, Slimane Thabet, Constantin Dalyac, and Loïc Henriet. "Quantum evolution kernel: Machine learning on graphs with programmable arrays of qubits". Physical Review A **104**, 032416 (2021).

[10] Constantin Dalyac, Loïc Henriet, Emmanuel Jeandel, Wolfgang Lechner, Simon Perdrix, Marc Porcheron, and Margarita Veshchezerova. "Qualifying quantum approaches for hard industrial optimization problems. a case study in the field of smart-charging of electric vehicles". EPJ Quantum Technology **8**, 12 (2021).

[11] Ryan Sweke, Frederik Wilde, Johannes Meyer, Maria Schuld, Paul K Fährmann, Barthélémy Meynard-Piganeau, and Jens Eisert. "Stochastic gradient descent for hybrid quantum-classical optimization". Quantum **4**, 314 (2020).

[12] Jun Li, Xiaodong Yang, Xinhua Peng, and Chang-Pu Sun. "Hybrid quantum-classical approach to quantum optimal control". Phys. Rev. Lett. **118**, 150503 (2017).

[13] Leonardo Banchi and Gavin E. Crooks. "Measuring analytic gradients of general quantum evolution with the stochastic parameter shift rule". Quantum **5**, 386 (2021).

[14] Richard P Feynman. "An operator calculus having applications in quantum electrodynamics". Physical Review **84**, 108 (1951).

[15] Ralph M Wilcox. "Exponential operators and parameter differentiation in quantum physics". Journal of Mathematical Physics **8**, 962–982 (1967).

[16] Javier Gil Vidal and Dirk Oliver Theis. "Calculus on parameterized quantum circuits". Preprint (2018).

[17] David Wierichs, Josh Izaac, Cody Wang, and Cedric Yen-Yu Lin. "General parameter-shift rules for quantum gradients". Preprint (2021).

[18] Dirk Oliver Theis. "Optimality of finite-support parameter shift rules for derivatives of variational quantum circuits". Preprint (2021).

[19] Michael Reed and Barry Simon. "Methods of Modern Mathematical Physics II: Fourier Analysis, Self-Adjointness". Volume 2. Academic Press. (1975).

[20] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven. "Barren plateaus in quantum neural network training landscapes". Nature communications **9**, 4812 (2018).

[21] Andrew Arrasmith, Zoë Holmes, Marco Cerezo, and Patrick J Coles. "Equivalence of quantum barren plateaus to cost concentration and narrow gorges". Quantum Science and Technology **7**, 045015 (2022).

[22] Walter Rudin. "Functional Analysis". McGraw-Hill. (1991).

[23] Elias M Stein and Rami Shakarchi. "Fourier Analysis: An Introduction". Volume 1. Princeton University Press. (2011).

[24] Gerald B Folland. "A Course in Abstract Harmonic Analysis". Volume 29. CRC press. (2016).

[25] Don Zagier. "The dilogarithm function". In Frontiers in number theory, physics, and geometry II. Pages 3–65. Springer (2007).

[26] Leonard C Maximon. "The dilogarithm function for complex argument". Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences **459**, 2807–2819 (2003).

[27] Elias M Stein and Rami Shakarchi. "Complex Analysis". Volume 2. Princeton University Press. (2010).

[28] Walter Rudin. "Real and Complex Analysis". McGraw-Hill. (1987).

[29] Heinz Bauer. "Maß- und Integrationstheorie". Walter de Gruyter. (1992). 2nd edition.

[30] Franz Rellich and Joseph Berkowitz. "Perturbation Theory of Eigenvalue Problems". CRC Press. (1969).