Efficient reverse engineering of one-qubit filter functions with dynamical invariants

R. K. L. Colmenar* and J. P. Kestner
Department of Physics, University of Maryland Baltimore County, Baltimore, MD 21250, USA

We derive an integral expression for the filter-transfer function of an arbitrary one-qubit gate through the use of dynamical invariant theory and Hamiltonian reverse engineering. We use this result to define a cost functional which can be efficiently optimized to produce one-qubit control pulses that are robust against specified frequency bands of the noise power spectral density. We demonstrate the utility of our result by generating optimal control pulses that are designed to suppress broadband detuning and pulse amplitude noise. We report an order of magnitude improvement in gate fidelity in comparison with known composite pulse sequences. More broadly, we also use the same theoretical framework to prove the robustness of nonadiabatic geometric quantum gates under specific error models and control constraints.

I. INTRODUCTION

Accurate manipulation of noisy quantum systems is an important problem in optimal control theory with potential applications in the field of chemical reaction control [1–3], quantum sensing [4, 5], and quantum information processing (QIP) [6] to name a few. In QIP, a typical strategy for suppressing errors due to noise is to use dynamical decoupling [7–12] and composite pulse sequences [13–18]. These techniques are designed to perturbatively suppress noise with correlation time scales that are much longer than the target evolution time (quasistatic noise). In many instances, however, quantum devices also suffer from non-static noise that fluctuates on the order of the evolution time or faster [12, 19–21]. Composite pulses have limited efficacy in such cases [22] and can even be detrimental to the quality of the generated quantum gate [23].

An alternative solution to these control problems is to use pulse shaping techniques [24–31]. The main idea of this approach is to find, either analytically or numerically, an appropriate set of time-dependent control Hamiltonian parameters that produces a desired evolution. Since the time-dependent Schrödinger equation (TDSE) is generally not analytically tractable, analytical solutions are typically limited to simple pulse shapes [32] or in restricted settings (e.g., for static error [27, 29] or state transfer protocols [25]). Numerical solutions offer much more flexibility in the control landscape. When combined with the formalism of filter functions [33], which characterizes the sensitivity of a control protocol to the power spectral density of the noise, it is possible to generate quantum gates that are robust against a specified spectral region of noise. Specifically, robust quantum gates are obtained by minimizing the overlap between the control's filter function and the noise power spectral density (PSD) in frequency space. This may be used, along with any control field constraints, to define a cost functional to be minimized using, for example, gradient-based methods. Optimization algorithms that are designed for deep learning and are implemented in platforms such as TensorFlow [34] or Julia's Flux package [35] are especially well-suited for these tasks owing to their built-in automatic differentiation capability. The power and flexibility offered by deep neural networks for solving quantum control problems has been demonstrated in a variety of recent works [36–41]. However, filter function engineering typically involves solving the TDSE for the time evolution operator. It is possible to circumvent this, for example, using Hamiltonian reverse engineering based on the theory of dynamical invariants [31]. Thus, it is possible to further reduce the computational workload of the optimization framework by reparameterizing the cost functional in terms of dynamical invariant parameters.

In this work, we use dynamical invariant theory and Hamiltonian reverse engineering to derive an integral expression for the filter function of an arbitrary one-qubit gate and explore its theoretical and practical applications. Our work is structured as follows. We begin Sec. II by reviewing the theory of dynamical invariants. We follow this up with a derivation of the one-qubit filter function for an arbitrary noise model in terms of the dynamical invariant parameters. We explore the practical applications of our results in Sec. III by numerically searching for optimal control solutions using deep neural networks. Specifically, we consider noise models with a $1/f$ noise spectrum [42] which is prevalent in solid-state qubits [20, 21, 43–46]. In addition, we discuss in Sec. IV some theoretical implications of our result by proving the robustness of geometric quantum gates against certain noise models under a strict only two-axis driving constraint. We then conclude and summarize our findings in Sec. V.

II. DYNAMICAL INVARIANTS

We consider as our starting point a general one-qubit control Hamiltonian with three-axis driving,

$$H_c(t) = \frac{1}{2}\begin{bmatrix} \Delta(t) & \Omega(t)\mathrm{e}^{-i\varphi(t)} \\ \Omega(t)\mathrm{e}^{i\varphi(t)} & -\Delta(t) \end{bmatrix}. \qquad (1)$$

* ralphkc1@umbc.edu

This particular form is relevant in systems such as superconducting qubits [47], quantum dot spin qubits [48], and NMR qubits [49] to name a few, corresponding to the rotating wave approximation for a two-level systm that is driven by an oscillating field with amplitude Ω at a carrier frequency detuned from resonance by Δ, and with phase φ. The solution to the time-dependent Schrödinger equation with this Hamiltonian is not analytically tractable in general. It is possible, however, to use the theory of dynamical invariants to reformulate this problem so as to specify a resulting unitary evolution and then analytically calculate a time-dependent Hamiltonian that would produce it [31]. A dynamical invariant $I(t)$ is a solution to the Liouville-von Neumann equation [50]

$$i\frac{\partial I(t)}{\partial t} - [H_c(t), I(t)] = 0. \tag{2}$$

The eigenvectors $|\phi_n(t)\rangle$ of $I(t)$ are related to the solutions of the Schrödinger equation by a global phase factor: $|\psi_n(t)\rangle = \mathrm{e}^{i\alpha_n(t)}|\phi_n(t)\rangle$, where $\alpha_n(t)$ are the Lewis-Riesenfeld phases given by [51]

$$\alpha_n(t) = \int_0^t \left\langle \phi_n(s) \middle| i\frac{\partial}{\partial s} - H_c(s) \middle| \phi_n(s) \right\rangle \mathrm{d}s. \tag{3}$$

Within this framework, the time evolution operator $U_c(t)$ can be expressed as

$$U_c(t) = \sum_{n=\pm} \mathrm{e}^{i\alpha_n(t)} |\phi_n(t)\rangle \langle \phi_n(0)| \,. \tag{4}$$

Thus, the theory of dynamical invariants effectively transforms the problem of solving the time-dependent Schrödinger equation to finding an appropriate $I(t)$ that satisfies Equation (2). As a consequence, we are free to choose a parametrization for $U_c(t)$ by choosing the $|\phi_n(t)\rangle$ appropriately. Suppose that we choose

$$|\phi_+(t)\rangle = \cos\left[\frac{\gamma(t)}{2}\right]\mathrm{e}^{-i\beta(t)}|0\rangle + \sin\left[\frac{\gamma(t)}{2}\right]|1\rangle\,, \tag{5}$$

$$|\phi_-(t)\rangle = \sin\left[\frac{\gamma(t)}{2}\right]|0\rangle - \cos\left[\frac{\gamma(t)}{2}\right]\mathrm{e}^{i\beta(t)}|1\rangle\,, \tag{6}$$

where $I(t)|\phi_n(t)\rangle = \pm\Omega_0/2\,|\phi_n(t)\rangle$ and Ω_0 is an arbitrary constant with units of frequency. This allows us to express $I(t)$ in a form similar to Equation (1)

$$I(t) = \frac{\Omega_0}{2}\begin{bmatrix}\cos(\gamma) & \sin(\gamma)\mathrm{e}^{-i\beta} \\ \sin(\gamma)\mathrm{e}^{i\beta} & -\cos(\gamma)\end{bmatrix}. \tag{7}$$

If we require Equations (1) and (7) to satisfy Equation (2), we are left with two coupled auxiliary equations [31]

$$\dot{\gamma} = -\Omega\sin(\beta - \varphi) \tag{8}$$

$$\Delta - \dot{\beta} = \Omega\cot(\gamma)\cos(\beta-\varphi), \tag{9}$$

which, along with the appropriate boundary conditions, can be used to determine the control parameters $\Omega(t)$, $\Delta(t)$, and $\varphi(t)$ that targets a desired $U_c(t)$. This choice of parametrization allows us to write $U_c(t)$ strictly in terms of the dynamical invariant parameters and the Lewis-Riesenfeld phase:

$$U_c(t) = \mathrm{e}^{-i\frac{\beta(t)}{2}\sigma_Z}\mathrm{e}^{-i\frac{\gamma(t)}{2}\sigma_Y}\mathrm{e}^{i\frac{\zeta(t)-\zeta(0)}{2}\sigma_Z} \times \mathrm{e}^{i\frac{\gamma(0)}{2}\sigma_Y}\mathrm{e}^{i\frac{\beta(0)}{2}\sigma_Z}, \tag{10}$$

where $\alpha = \alpha_+ = -\alpha_-$ and we introduce a new parameter

$$\zeta(t) = 2\alpha(t) - \beta(t) = -\beta(0) + \int_0^t \frac{\dot{\gamma}\cot(\beta-\varphi)}{\sin\gamma}\,\mathrm{d}t' \tag{11}$$

The auxiliary equations provide a family of control solutions that allow us to reverse engineer a desired quantum gate. Since the gate only depends on the boundary values of the dynamical invariant parameters, there are infinitely many ways to generate the gate. It is desirable to use this freedom in the control Hamiltonian such that the resulting evolution is also robust against noise. To this end, filter functions provide a convenient method of quantifying the gate fidelity's susceptibility to noise with respect to its spectral properties [33]. The total one-qubit Hamiltonian in the presence of noise can be written as

$$H(t) = H_c(t) + H_e(t), \tag{12}$$

where $H_c(t)$ is the ideal deterministic control Hamiltonian and $H_e(t)$ is the stochastic error Hamiltonian. More explicitly, $H_e(t)$ can generally be expressed as

$$H_e(t) = \sum_q \sum_{i=1}^{3} \delta_q(t)\chi_{q,i}(t)\sigma_i, \tag{13}$$

where q indexes a set of uncorrelated stochastic variables $\delta_q(t)$, $\chi_{q,i}(t)$ contains the sensitivity of the control parameters (which generally can be a function of the parameters themselves) to $\delta_q(t)$, and σ_i are Pauli operators. For sufficiently weak noise, the average gate infidelity $\langle\mathcal{I}\rangle$ of the noisy evolution $U(t)$, which satisfies $i\dot{U}(t) = H(t)U(t)$ where $U(0) = \mathbb{1}$, can be compactly expressed as (see Appendix A)

$$\langle\mathcal{I}\rangle \approx \frac{1}{2\pi}\sum_q \int_{-\infty}^{\infty} S_q(\omega)F_q(\omega)\,\mathrm{d}\omega, \tag{14}$$

where $S_q(\omega)$ denotes the noise PSD for the stochastic variable $\delta_q(t)$ and $F_q(\omega)$ is the corresponding filter function which can be calculated using the following equations:

$$F_q(\omega) = \sum_k |R_{q,k}(\omega)|^2\,, \tag{15}$$

$$R_{q,k}(\omega) = \sum_i \int_0^T \chi_{q,i}(t)R_{ik}(t)\mathrm{e}^{i\omega t}\,\mathrm{d}t, \tag{16}$$

$$R_{ik}(t) = \frac{1}{2}\operatorname{tr}\left(U_c^\dagger(t)\sigma_i U_c(t)\sigma_k\right), \tag{17}$$

where T is the gate time.

Combining Equations (4) and (17) allows us to express Equation (16) as

$$R_{q,k}(\omega) = \frac{1}{2}\sum_{i,n,n'} \langle\phi_n(0)|\sigma_k|\phi_{n'}(0)\rangle \times \int_0^T \mathrm{e}^{i(\alpha_n(t)-\alpha_{n'}(t)+\omega t)}\chi_{q,i}(t)\,\langle\phi_{n'}(t)|\sigma_i|\phi_n(t)\rangle\ \mathrm{d}t. \tag{18}$$

Thus, the filter function corresponding to $\delta_q(t)$ is given by

$$\begin{aligned}F_q(\omega) &= \sum_k R_{q,k}(\omega)R^*_{q,k}(\omega)\\ &= \frac{1}{4}\int_0^T\int_0^T \sum_{\substack{i,j,k,\\ n,m,n',m'}} \langle\phi_n(0)|\sigma_k|\phi_{n'}(0)\rangle\,\langle\phi_m(0)|\sigma_k|\phi_{m'}(0)\rangle\\ &\mathrm{e}^{i(\alpha_n(t_1)-\alpha_{n'}(t_1)+\omega t_1)}\chi_{q,i}(t_1)\,\langle\phi_{n'}(t_1)|\sigma_i|\phi_n(t_1)\rangle\ \mathrm{d}t_1\\ &\mathrm{e}^{i(\alpha_m(t_2)-\alpha_{m'}(t_2)-\omega t_2)}\chi_{q,j}(t_2)\,\langle\phi_{m'}(t_2)|\sigma_j|\phi_m(t_2)\rangle\ \mathrm{d}t_2.\end{aligned} \tag{19}$$

For a given n, n', m, and m', the k-dependent factors of this sum yields

$$\sum_k \langle\phi_n(0)|\sigma_k|\phi_{n'}(0)\rangle\,\langle\phi_m(0)|\sigma_k|\phi_{m'}(0)\rangle = \begin{cases}1 & \text{if } \{n,n',m,m'\}=\{\pm,\pm,\pm,\pm\}\\ -1 & \text{if } \{n,n',m,m'\}=\{\pm,\pm,\mp,\mp\}\\ 2 & \text{if } \{n,n',m,m'\}=\{\pm,\mp,\mp,\pm\}\\ 0 & \text{otherwise}\end{cases}. \tag{20}$$

We can use Eqs. (5), (6), (20) as well as the fact that $\langle\phi_\pm(t)|\sigma_k|\phi_\pm(t)\rangle = -\langle\phi_\mp(t)|\sigma_k|\phi_\mp(t)\rangle$ to simplify Eq. (19) into

$$\begin{aligned}F_q(\omega) = \sum_{i,j}&\left(\int_0^T \langle\phi_+(t)|\sigma_i|\phi_+(t)\rangle\,\chi_{q,i}(t)e^{\imath\omega t}\,\mathrm{d}t\right)\left(\int_0^T \langle\phi_+(t)|\sigma_j|\phi_+(t)\rangle\,\chi_{q,j}(t)e^{-\imath\omega t}\,\mathrm{d}t\right)\\ &+\frac{1}{2}\left(\int_0^T \langle\phi_-(t)|\sigma_i|\phi_+(t)\rangle\,\chi_{q,i}(t)e^{\imath 2\alpha(t)+\imath\omega t}\,\mathrm{d}t\right)\left(\int_0^T \langle\phi_+(t)|\sigma_j|\phi_-(t)\rangle\,\chi_{q,j}(t)e^{-\imath 2\alpha(t)-\imath\omega t}\mathrm{d}t\right)\\ &+\frac{1}{2}\left(\int_0^T \langle\phi_+(t)|\sigma_i|\phi_-(t)\rangle\,\chi_{q,i}(t)e^{-\imath 2\alpha(t)+\imath\omega t}\,\mathrm{d}t\right)\left(\int_0^T \langle\phi_-(t)|\sigma_j|\phi_+(t)\rangle\,\chi_{q,j}(t)e^{\imath 2\alpha(t)-\imath\omega t}\,\mathrm{d}t\right).\end{aligned} \tag{21}$$

Finally, substituting in Eqs. (5) and (6) allows us to compactly write Eq. (21) in the following vectorized expression:

$$F_q(\omega) = \left\|\int_0^T \Lambda(t)\begin{bmatrix}\chi_{q,X}(t)\\ \chi_{q,Y}(t)\\ \chi_{q,Z}(t)\end{bmatrix} e^{i\omega t}\,\mathrm{d}t\right\|^2, \tag{22}$$

where the entries of the matrix Λ are given by

$$\begin{aligned}\Lambda_{11} &= \cos\beta\sin\gamma,\\ \Lambda_{12} &= \sin\beta\sin\gamma,\\ \Lambda_{13} &= \cos\gamma,\\ \Lambda_{21} &= -\cos\beta\cos\gamma\cos\zeta - \sin\beta\sin\zeta,\\ \Lambda_{22} &= -\sin\beta\cos\gamma\cos\zeta + \cos\beta\sin\zeta,\\ \Lambda_{23} &= \sin\gamma\cos\zeta,\\ \Lambda_{31} &= -\cos\beta\cos\gamma\sin\zeta + \sin\beta\cos\zeta,\\ \Lambda_{32} &= -\sin\beta\cos\gamma\sin\zeta - \cos\beta\cos\zeta,\\ \Lambda_{33} &= \sin\gamma\sin\zeta.\end{aligned} \tag{23}$$

This is our main result and we show in the following sections some examples of its utility. Before we proceed, we comment on the form of Eq. (22). First, although the similarity between Eqs. (15)-(16) and (22) might seem to suggest that $R(t)$ and $\Lambda(t)$ are identical and we have not really simplified anything, in fact what we have done is to note that the dependence of $R(t)$ on the value of the dynamical invariant parameters evaluated at $t = 0$ does not affect the filter function value, and $\Lambda(t)$ does not carry that extraneous dependence. Second, certain error models admit an alternative interpretation for Eq. (22). For example, suppose we consider the dephasing and over-rotation noise models. The former can be induced by an additive shift to the qubit detuning, $\Delta(t) \to \Delta(t)+\delta_\Delta(t)$, and the latter can be induced by a multiplicative shift in the pulse amplitude, $\Omega(t) \to \Omega(t)\,(1+\delta_\Omega(t))$. The corresponding error sensitivities are $\boldsymbol{\chi}_\Delta(t) = \frac{1}{2}[0,0,1]^\intercal$ and $\boldsymbol{\chi}_\Omega(t) = \frac{1}{2}[\Omega\cos\varphi, \Omega\sin\varphi, 0]^\intercal$. Substituting these expressions onto Eq. (22) yields the following filter functions

$$F_\Delta(\omega) = \left\| \int_0^T \frac{1}{2} \begin{bmatrix} \cos\gamma \\ \sin\gamma\cos\zeta \\ \sin\gamma\sin\zeta \end{bmatrix} e^{i\omega t}\,\mathrm{d}t \right\|^2, \tag{24}$$

$$F_\Omega(\omega) = \left\| \int_0^T \frac{1}{2} \begin{bmatrix} \dot\zeta\sin^2\gamma \\ \dot\zeta\sin\gamma\cos\gamma\cos\zeta + \dot\gamma\sin\zeta \\ \dot\zeta\sin\gamma\cos\gamma\sin\zeta - \dot\gamma\cos\zeta \end{bmatrix} e^{i\omega t}\,\mathrm{d}t \right\|^2. \tag{25}$$

Up to a scalar factor, the detuning filter function in Eq. (24) can be reinterpreted as a position vector with constant speed $\dot{\vec{r}} = [\cos\gamma, -\sin\gamma\cos\zeta, -\sin\gamma\sin\zeta]$ [52]. If robustness at a certain noise frequency is defined by a vanishing filter function value, robustness against static detuning noise (i.e., at $\omega = 0$) is equivalent to having the position vector trace a closed three-dimensional curve whose curvature is given by $\Omega(t)$. Such a geometric interpretation has been noted previously in the literature [53–59].

A similar observation can be made for the pulse amplitude filter function. Note that the vector in the integrand of Eq. (25) is equivalent to $\dot{\vec{r}} \times \ddot{\vec{r}}$. This can be rewritten as $\Omega\vec{b}$ [60], where $\vec{b}$ is the binormal vector corresponding to $\vec{r}$ and we have used the fact that the curvature $\kappa = \Omega$. Therefore, constructing a quantum gate that is simultaneously robust against static detuning and pulse amplitude noise is mathematically equivalent to finding a closed three-dimensional curve such that $\int_0^T \Omega(t)\vec{b}(t)\,\mathrm{d}t = \vec{0}$. As far as we know, this has not been noted before.

III. BROADBAND NOISE OPTIMIZATION

We demonstrated in Sec. II that it is possible through Hamiltonian reverse engineering to analytically calculate the filter function of an arbitrary one-qubit gate in terms of the dynamical invariant parameters $\beta(t)$, $\gamma(t)$, and $\zeta(t)$ as well as the sensitivity $\chi_{q,i}(t)$. One immediate implication of this result is the possibility of filter function engineering which can be used for error suppression [36–39] or quantum sensing [61]. In the context of error suppression, we can use Equation (22) to define a cost functional which can be minimized in spectral regions where the noise PSD is dominant. This approach allows us to target any robust one-qubit gate provided that we can find an appropriate $\gamma(t)$ and $\beta(t)$. Furthermore, this is different from previous filter function engineering results since calculating the evolution operator is no longer necessary, which helps to reduce the computational workload of the optimization framework.

We consider again as an example the case where our system is subject to detuning and pulse amplitude noise. Note that both Eqs. (24) and (25) depend only on γ and ζ. This means that β is a free parameter up to the boundary conditions imposed by the reverse engineering process. This extra degree of freedom can be used to impose control restrictions such as strict two-axis control. Combining Eqs. (8), (9), and (11) provides us with the reverse engineered Hamiltonian parameters in terms of the dynamical invariant parameters:

$$\Omega = \sqrt{\dot\gamma^2 + \dot\zeta^2\sin^2\gamma} \tag{26}$$

$$\varphi = \beta - \arctan\frac{\dot\gamma}{\dot\zeta\sin\gamma} \tag{27}$$

$$\Delta = \dot\beta - \dot\zeta\cos\gamma. \tag{28}$$

We can set $\Delta = 0$ by solving the differential equation $\dot\beta = \dot\zeta\cos\gamma$ for β with the boundary condition $\beta(0) = -\zeta(0)$. Thus, all properties of the output gate is determined by γ and ζ.

Restricting β in this manner does not necessarily diminish our ability to target arbitrary one-qubit gates. In practice, a finite set of quantum gates are used to target arbitrary operations. Although we can engineer γ and ζ to target gates directly, it is worth pointing out that many qubit qubit implementations have access to virtual Z (VZ) gates [62–65]. These zero-duration gates are essentially perfect and implemented through abrupt changes to the reference phase. It can be shown that any one-qubit gate can be decomposed into the product of Z gates and two $X_{\frac{\pi}{2}}$ [65]: $Z_{\theta_1}X_{\frac{\pi}{2}}Z_{\theta_2}X_{\frac{\pi}{2}}Z_{\theta_3}$. We can rewrite the engineered evolution operator in Eq. (10) as

$$\begin{aligned} U_c(t) &= Z_{\beta(t)}Y_{\gamma(t)}Z_{\zeta(0)-\zeta(t)}Y_{-\gamma(0)}Z_{-\beta(0)} \\ &= Z_{\psi_1}X_\theta Z_{\psi_2}, \end{aligned} \tag{29}$$

where

$$\begin{aligned} \cos(\theta) = {}& \cos(\zeta(0)-\zeta(t))\sin(\gamma(t))\sin(\gamma(0)) \\ &+ \cos(\gamma(t))\cos(\gamma(0)), \end{aligned} \tag{30}$$

and ψ_1 and ψ_2 are angles that depend on the target gate. By setting $\theta = \frac{\pi}{2}$, we can replace all $X_{\frac{\pi}{2}}$ in the gate decomposition with $U_c(T)$ and combining all neighboring Z gates. Since Z gates may be executed virtually, we only need one physical gate, $U_c(T)$ with $\theta = \frac{\pi}{2}$, to produce any one-qubit operation.

Our goal is to minimize the following cost functional:

$$\begin{aligned} \text{cost} = {}& \int_{-\infty}^{\infty} F_\Delta(\omega)S_\Delta(\omega)\,\mathrm{d}\omega + \int_{-\infty}^{\infty} F_\Omega(\omega)S_\Omega(\omega)\,\mathrm{d}\omega \\ &+ |\cos(\zeta(0)-\zeta(T))\sin\gamma(T)\sin\gamma(0) + \cos\gamma(T)\cos\gamma(0)| \\ &\left|\frac{\Omega(0)}{\Omega_{\max}}\right| + \left|\frac{\Omega(T)}{\Omega_{\max}}\right| + \sum_i \max\left(0, \frac{\Omega(t_i)}{\Omega_{\max}} - 1\right). \end{aligned} \tag{31}$$

The first two terms correspond to the infidelity integrals for detuning and amplitude noise with noise PSD S_Δ and S_Ω, respectively. The third term is the constraint that targets $\theta = \frac{\pi}{2}$. The fourth and fifth term sets the boundary value of the pulse amplitude to zero [66]. Finally, the sixth term imposes a maximum value $\Omega_{\max}$ on Ω by discretizing the interval $[0, T]$ and evaluating Ω at each time

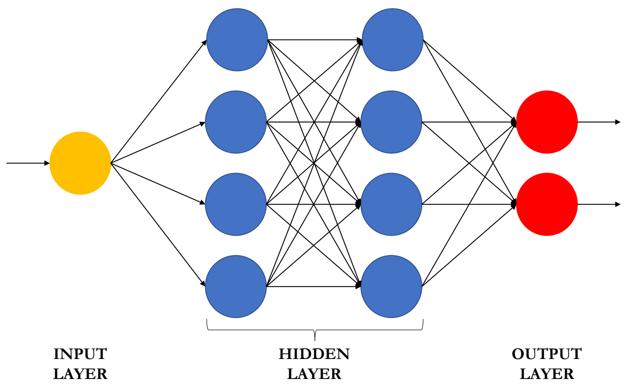

FIG. 1. A schematic diagram of a feedforward deep neural network with one input neuron, two output neurons, and two hidden layers with four neurons each. A network is deep if it has at least two hidden layers. As information flows from the input layer, each subsequent layer nonlinearly transforms incoming information and returns a value. The goal is to train the neural network so that the final output optimizes the cost. In our case, we would like to train a neural network to take time as input and return the optimized dynamical invariant parameters γ and ζ.

value. The cost penalizes any point where $\Omega(t_i) > \Omega_{\max}$ through the function

$$\max(0,x) = \begin{cases} 0 & x \leq 0 \\ x & x > 0 \end{cases}. \tag{32}$$

We demonstrate the flexibility of our approach by considering two examples. We first consider a case where the goal is to produce a gate that acts as a stopband filter against $1/f$ detuning and pulse amplitude noise. We then consider a case where the goal is to produce a gate that is optimal in the presence of $1/f$ pulse amplitude noise and a static detuning noise. To this end, we employ deep neural networks [67, 68] as our optimization framework. The power of neural networks originate from their ability to represent complex ideas as a hierarchy of simpler concepts. This allows them to efficiently identify key abstract properties of a problem, which is highly coveted in tasks such as pattern recognition [69]. It has also been proven that neural networks with sufficient neurons and layers can act as a universal function approximator [70, 71]. This is ideal for our purpose since it eliminates the nontrivial task of finding suitably paramemeterized ansatz function to optimize over that will yield convergent solutions. Furthermore, machine learning frameworks tend to have built-in automatic differentiation capabilities which can be utilized for gradient-based optimization.

In particular, we use a feedforward neural network (sometimes referred to as multilayer perceptron) which is constructed using layers of interconnected computational units called neurons such that information travels only in one direction; starting with an input layer, then a series of hidden layers, and finally onto an output layer. A schematic diagram of a feedforward neural network is shown in Fig. 1. A neural network is deep if it has at least two hidden layers. Each adjacent layers act as a function that takes a vector input and produces a vector output using the following model

$$\boldsymbol{x}_{i+1} = \sigma\left(W_i \boldsymbol{x}_i + \boldsymbol{b}_i\right), \tag{33}$$

where $\boldsymbol{x}_i$ is the input in the i^{th} layer, W_i is a matrix that describes the neural connections between the i^{th} and $(i+1)^{\text{th}}$ layer, $\boldsymbol{b}_i$ is a bias vector, and $\sigma(\cdot)$ is a nonlinear activation function such as $\max(0,\cdot)$ or $\tanh(\cdot)$. Our goal is to train the neural network using machine learning algorithms to return the optimized dynamical invariant parameters γ and ζ on the output layer by feeding in time on the input layer. For our optimization we use a feedforward deep neural network with one input neuron, two hidden layers with 16 neurons each and a tanh activation function, and two output neurons for a total of 338 parameters [72].

A. $1/f$ stopband filter for both detuning and pulse amplitude noise

For our first example, we consider identical noise PSD for detuning and amplitude noise:

$$S_\Delta(\omega) = S_\Omega(\omega) = \begin{cases} \frac{A}{\omega} & \omega_0 \leq |\omega| \leq \omega_c \\ 0 & \text{otherwise} \end{cases}, \tag{34}$$

where $[\omega_0, \omega_c]$ defines the frequency stopband in which we wish to suppress noise. We set $\omega_0 = 10^{-9}\Omega_{\max}$, $\omega_c = 10^{-1}\Omega_{\max}$, and $T = 16\pi/\Omega_{\max}$. We present in Fig. 2 a plot of the optimized control fields and filter functions. The details of our numerical optimization scheme is provided in App. B.

We see from Fig. 2 that the control pulse we produced satisfies the imposed constraints. We compare the total infidelity of our optimized pulse with that of known pulse sequences in the literature that address either detuning noise, pulse amplitude noise, or both. We present in Table I a summary of these comparisons. We find that our broadband optimized pulse yields an infidelity that is at least an order of magnitude lower than than any other pulse sequences. Specifically, the minimum improvement is roughly a factor of 14 which is a comparison with the concatenated CORPSE [73–75] and BB1 [76] pulse sequence (CinBB) [77]. CinBB is able to address static detuning and pulse amplitude noise simultaneously. We attribute the improvement in our result to the fact that pulse sequences are generally designed to suppress static noise. Although they can be used to address noise in the quasitatic regime, their ability to suppress noise that fluctuate on the order of $\Omega_{\max}$ severely limited. However,

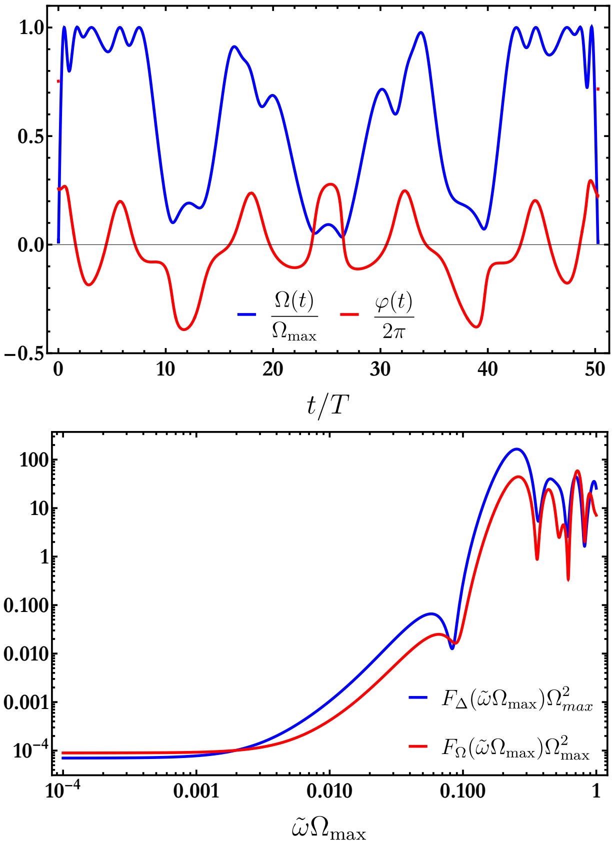

FIG. 2. A plot of the optimized Hamiltonian (TOP) and filter function (BOTTOM) for the case of simultaneous $1/f$ detuning and pulse amplitude noise over a finite frequency range.

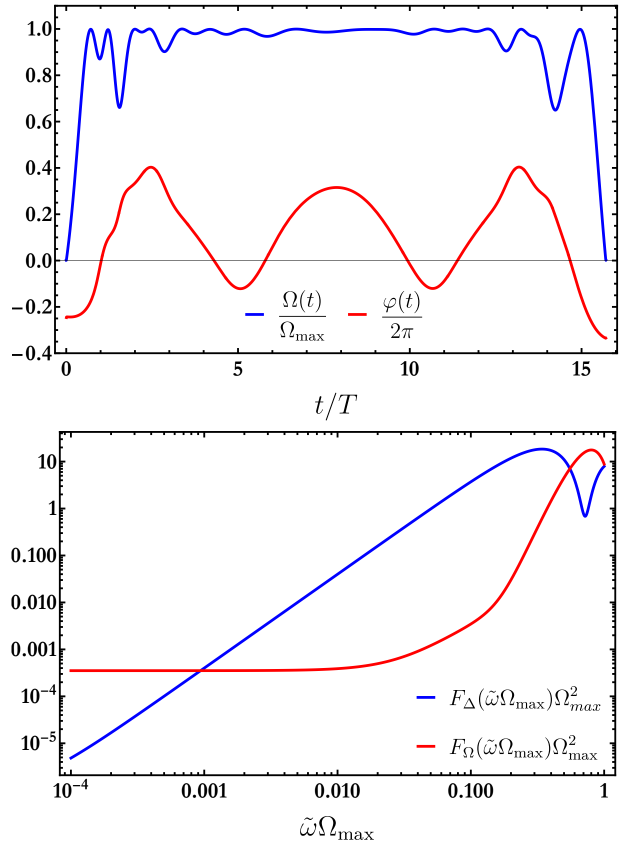

FIG. 3. A plot of the optimized Hamiltonian (TOP) and filter function (BOTTOM) for the case of static detuning noise and $1/f$ pulse amplitude noise. Unlike the previous example, the $1/f$ spectrum here has a $1/f^2$ tail which penalizes large filter function values in the $\omega_c \leq |\omega|$ region.

it is worth noting that our pulse's performance comes at the cost of increased noise sensitivity in frequency regions beyond the indicated stopband.

B. Static detuning and $1/f$ pulse amplitude noise

For our second example, we consider the case where we have a static detuning noise as well as a $1/f$ pulse amplitude noise:

$$S_\Delta(\omega) = 10A\delta(\omega), \tag{35}$$

$$S_\Omega(\omega) = \begin{cases} 0 & 0 \leq |\omega| \leq \omega_0 \\ \frac{A}{\omega} & \omega_0 \leq |\omega| \leq \omega_c \\ \frac{A\omega_c}{\omega^2} & \omega_c \leq |\omega| \end{cases}, \tag{36}$$

where we have assumed an order of magnitude difference in the detuning and pulse amplitude noise strength. Here we set $\omega_0 = 10^{-9}\Omega_{\max}$, $\omega_c = 10^{-1}\Omega_{\max}$, and $T = 5\pi/\Omega_{\max}$. We present in Fig. 3 a plot of the optimized control fields and filter functions. We again compare our optimized pulse with known pulse sequences and the results are summarized in Table I.

Unlike the previous case, we only see a minimum improvement in infidelity by a factor of 7. This is primarily attributed to penalizing the value of the filter function in regions where $\omega_c \leq |\omega|$. In the previous case, the improvement is due to the fact that filter function values outside the stopgap do not contribute to the infidelity. This is no longer true in this case due to the presence of a $1/f^2$ tail which penalizes large filter function values for frequency values greater than ω_c.

IV. ROBUSTNESS OF GEOMETRIC PHASES

We can also apply our result in Sec. II to explore the robustness properties of geometric quantum gates. Geometric gates are quantum gates with a trivial dynamical phase which means they rely on the geometric phase to produce unitary dynamics. Geometric gates are of practical interest since they are conjectured to be ideal for robust quantum computation owing to the global nature of the accumulated phase. The validity of this conjecture is the subject of many studies with many showing support for the conjecture. However, there are also studies

TABLE I. A comparison of infidelities between our deep neural network output and known composite pulse sequences. The ratio $\mathcal{I}_i/\mathcal{I}_{\text{DNN}}$ compares the infidelity of naive and composite pulses targeting a $X_{\frac{\pi}{2}}$ gate against our optimized gate. The subscript A/B indicates whether the case of Sec. III A or that of Sec. III B is in consideration. We also indicate which pulses are robust against static detuning and/or pulse amplitude noise. We report a substantial decrease in infidelity in all cases we considered.

Pulse	$\mathcal{I}_{\text{A}}/\mathcal{I}_{\text{DNN}}$	$\mathcal{I}_{\text{B}}/\mathcal{I}_{\text{DNN}}$	Robust to δ_Δ?	Robust to δ_Ω?
Naive (Square)	223	14	No	No
Short CORPSE	1110	68	Yes	No
BB1	110	9	No	Yes
CinBB	14	7	Yes	Yes
CinSK	57	15	Yes	Yes

that report situations in which geometric gate are not intrinsically more robust than dynamical gates [57, 78–81] and, in certain scenarios, their sensitivity to noise deteriorates [82–86]. Recently, it was shown in Ref. [87] that the noise sensitivity of geometric and dynamical gates are generically equal. However, when control constraints are present (e.g. strict 2-axis or piecewise constant control), it is possible for a particular phase type to become preferable and naturally robust to some noise.

Here we demonstrate preferential phase robustness in nonadiabatic Abelian geometric gates as a consequence of control constraints. To clarify, we say that a quantum gate is robust against the noise process q at a particular frequency ω if $F_q(\omega) = 0$. Our theoretical framework is ideal for this type of analysis because the notion of geometric and dynamical phases are naturally developed in the theory of dynamical invariants. Using the Lewis-Riesenfeld phase in Eq. (3), the eigenvectors in Eqs. (5) and (6), as well as the auxiliary equations in Eqs. (8) and (9), we can define the geometric phase, $\alpha_{n,g}$, and dynamical phase, $\alpha_{n,d}$, accumulated by the eigenvector $|\phi_n\rangle$ during the evolution as

$$\begin{aligned} \alpha_{n,g}(T) &= \int_0^T \left\langle \phi_\pm(t) \middle| i\frac{\partial}{\partial t} \middle| \phi_\pm(t) \right\rangle \mathrm{d}t \\ &= \pm\alpha(T) \mp \int_0^T \frac{\dot\zeta - \dot\beta\cos\gamma}{2}\,\mathrm{d}t, \end{aligned} \tag{37}$$

$$\begin{aligned} \alpha_{n,d}(T) &= -\int_0^T \langle \phi_\pm(t)|H(t)|\phi_\pm(t)\rangle\,\mathrm{d}t \\ &= \pm\int_0^T \frac{\dot\zeta - \dot\beta\cos\gamma}{2}\,\mathrm{d}t. \end{aligned} \tag{38}$$

Suppose we consider the special case of a constant detuning Δ, which is a fairly common constraint in works considering geometric gates [84, 88–90]. We prove the following theorem for that special case by analyzing the filter function expressions that we derived:

Theorem. *Consider the control Hamiltonian in Eq.* (1) *under the constraint that Δ is constant. Any gate that is robust to static multiplicative amplitude noise (δ_Ω) as well as static additive or multiplicative detuning noise (δ_Δ) is necessarily geometric.*

Proof. Using Eq. (28), we can rewrite the dynamical phase integral in Eq. (38) as

$$\alpha_{n,d}(T) = \pm\int_0^T \frac{\dot\zeta - \dot\beta\cos\gamma}{2}\,\mathrm{d}t \tag{39}$$

$$\begin{aligned} &= \pm\int_0^T \frac{-\Delta\cos\gamma + \dot\zeta\sin^2\gamma}{2}\,\mathrm{d}t \\ &= \mp\frac{\Delta}{2}\int_0^T \cos\gamma\,\mathrm{d}t \pm \frac{1}{2}\int_0^T \dot\zeta\sin^2\gamma\,\mathrm{d}t. \end{aligned} \tag{40}$$

We begin by considering the case where there is additive detuning and multiplicative pulse amplitude noise. Imposing simultaneous robustness against these noise sources would require $F_\Delta(0) = F_\Omega(0) = 0$. However, we see in Eqs. (24) and (25) that the filter function is strictly nonnegative and the only way to achieve robustness against static noise is if every integral vanishes. Specifically, robustness against static additive detuning noise requires $\int_0^T \cos\gamma\,\mathrm{d}t = 0$, while robustness against static amplitude noise requires $\int_0^T \dot\zeta\sin^2\gamma\,\mathrm{d}t = 0$. Notice, however, that these are precisely the integral expressions in Eq. (40). Thus, simultaneous robustness against static detuning and pulse amplitude error necessarily requires the dynamical phase to vanish, i.e., the gate must be geometric.

Next, we consider the case where there is multiplicative detuning and pulse amplitude noise. The multiplicative detuning filter function can be found using Eq. (28) and is given by

$$\begin{aligned} F_{\Delta,\times}(\omega) &= \left\| \int_0^T \frac{\Delta}{2} \begin{bmatrix} \cos\gamma \\ \sin\gamma\cos\zeta \\ \sin\gamma\sin\zeta \end{bmatrix} e^{i\omega t}\,\mathrm{d}t \right\|^2 \\ &= \left\| \int_0^T \frac{\dot\beta - \dot\zeta\cos\gamma}{2} \begin{bmatrix} \cos\gamma \\ \sin\gamma\cos\zeta \\ \sin\gamma\sin\zeta \end{bmatrix} e^{i\omega t}\,\mathrm{d}t \right\|^2. \end{aligned} \tag{41}$$

Robustness to static noise would require $F_{\Delta,\times}(0) = 0$. We focus in particular on the first integral which can be rewritten as

$$\frac{1}{2}\int_0^T \dot\beta\cos\gamma - \dot\zeta + \dot\zeta\sin^2\gamma\,\mathrm{d}t. \tag{42}$$

Just like before, we note that imposing robustness against static pulse amplitude noise requires $\int_0^T \dot\zeta\sin^2\gamma\,\mathrm{d}t = 0$ which eliminates the last term in expression above. Setting the remaining terms to zero is equivalent to setting Eq. (39) to zero. Therefore, imposing simultaneous robustness against static multiplicative detuning and pulse amplitude noise necessitates a geometric gate. □

We make the following observations. First, this theorem is consistent with other results in the literature. It was previously noted in Refs. [77, 91] that composite pulse sequences with detuning fixed to zero that are designed to be robust against multiplicative pulse amplitude noise (and are trivially robust against multiplicative detuning noise since $\Delta = 0$) are indeed geometric quantum gates. Second, we note that in that special case of $\Delta = 0$, the first term in Eq. (40) vanishes regardless of the value of the integral. In other words, if we don't require robustness to pulse amplitude noise, it *is* possible to obtain dynamical gates that are robust to static detuning noise. A well-known example is the CORPSE family of composite pulses which are designed to be robust against additive detuning noise [73–75]. Third, gates that are robust against static multiplicative pulse amplitude noise are necessarily geometric but the converse isn't true. One specific example of this is the orange-slice geometric gate presented in Ref. [89]. It was shown in Ref. [87] that the pulse amplitude filter function in this particular case does not vanish at $\omega = 0$ despite being a geometric gate. Fourth, we note that the parallel transport condition ($\langle\phi_{\pm}(t)|H(t)|\phi_{\pm}(t)\rangle = 0$) is not necessary to achieve a robust geometric gate; the dynamical phase integral simply has to vanish at the gate time. Finally, this theorem is consistent with the results of Ref. [87]. It is argued there that in the absence of control constraints, geometric and dynamical gates are generically equivalent when it comes to noise sensitivity, and preferential phase robustness can only emerge in the presence of control constraints. In this case, the constraint is considering a strictly constant Δ. Removing the constraint on Δ turns β into a free parameter. According to Eqs. (37) and (38), the geometric and dynamical component of the total phase is directly dependent on our choice of β. Thus, in the absence of constraints, we can freely tune the phase type from dynamical to geometric. Moreover, the filter functions in Eqs. (24) and (25) are independent of β. This indicates that noise sensitivity, as quantified by the filter function, is independent of the phase type in the absence of control constraints as was also shown more generally in Ref. [87].

V. CONCLUSIONS

We make use of dynamical invariant theory in order to analytically reverse engineer a qubit's control Hamiltonian and calculate its corresponding filter function. This allows us to define a cost function strictly in terms of the dynamical invariant parameters which can be optimized to create filter functions with desirable properties. The primary limitation of our theory is its currently limited applicability to two-level systems, with no provision for operations on more than one qubit or correction of population leakage to higher energy levels. (The effects of *virtual* transitions to higher energy levels do not pose a problem, since they can be incorporated in an effective one-qubit Hamiltonian [92].) In those cases a generalized approach such as Ref. [36] is preferable. However, for the specific task of constructing local rotations with robustness against high frequency noise bands, our method is a useful and efficient tool.

We demonstrate the utility of our theory by generating control pulses that are optimized to operate in the presence of broadband noise. One example we considered is creating a stopband filter for both detuning and pulse amplitude noise. We report at least an order of magnitude improvement in infidelity when our optimized pulse is compared with known composite pulse sequences that are designed to address one or both noise types. Although filter funciton engineering itself is not a novel concept [36], our approach is efficient since the reverse engineering process circumvents the need to compute the evolution operator during the optimization process. The optimizer only requires that we calculate a simple integral expression with the engineered parameters as its input. Furthermore, the engineered parameters offer adequate flexibility to simultaneously target abitrary qubit gates while considering control parameter constraints. In principle, more complicated constraints, such as using different basis functions (Chebyshev, Walsh, Slepian, etc.), time-symmetric or antisymmetric control [27, 29, 93], or spectral-phase-only optimization [26] to name a few, can also be incorporated into our theory. Our results can also be applied to quantum sensing where instead the goal is to maximize the filter function in a limited noise spectral bandwidth [61, 94].

More broadly, we used our theoretical framework to analyze the robustness of geometric gates to detuning and pulse amplitude errors. We proved a theorem for the special case of a control constraint under which geometric gates are necessarily superior to dynamical gates. We emphasize that the robustness we report is not a generic property of geometric gates but rather a consequence of imposing control constraints.

The authors acknowledge support from the National Science Foundation under Grant No. 1915064.

Appendix A: Estimating gate fidelity using filter functions

We now provide a more detailed derivation of the average gate infidelity provided in Equation (14) which was reported in Ref. [33]. We begin by writing the noisy Hamiltonian as

$$H(t) = H_c(t) + H_e(t), \tag{A1}$$

where $H_c(t)$ is the deterministic control Hamiltonian and $H_e(t)$ is the stochastic error Hamiltonian which can generally expressed as in Equation (13). By moving to the interaction frame, we can write the noisy time evolution $U(t) = U_c(t)U_e(t)$ where each factors are solutions to the following Schrödinger equations:

$$i\dot{U}_c(t) = H_c(t)U_c(t) \tag{A2}$$

$$i\dot{U}_e(t) = \left(U_c^\dagger(t)H_eU_c(t)\right)U_e(t). \tag{A3}$$

For sufficiently weak noise, we can perturbatively expand $U_e(t)$ using the Magnus expansion and write

$$U_e(t) \approx \exp\left[-i\int_0^T U_c^\dagger(t)H_e(t)U_c(t)\,\mathrm{d}t\right]. \tag{A4}$$

The average gate infidelity is given by

$$\begin{aligned}\langle\mathcal{I}\rangle &= \langle 1 - F_{\mathrm{tr}}\rangle = \left\langle 1 - \left|\mathrm{tr}\left(U_c^\dagger U\right)/\,\mathrm{tr}\left(U_c^\dagger U_c\right)\right|^2\right\rangle \\ &= \left\langle 1 - \left|\mathrm{tr}\,U_e/2\right|^2\right\rangle \\ &\approx \left\langle \mathrm{tr}\int_0^T\int_0^T \left[U_c^\dagger(t_1)H_e(t_1)U_c(t_1)\right]\right. \\ &\qquad \left.\times\left[U_c^\dagger(t_2)H_e(t_2)U_c(t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2\right]\right\rangle. \end{aligned} \tag{A5}$$

We can use the adjoint representation of $U_c(t)$ defined through

$$R_{ij}(t) = \frac{1}{2}\,\mathrm{tr}\left(U_c^\dagger(t)\sigma_iU_c(t)\sigma_j\right) \tag{A6}$$

and Equation (13) to rewrite Equation (A5) into

$$\begin{aligned}\langle\mathcal{I}\rangle \approx \sum_{q,i,j,k}\int_0^T\int_0^T \langle\delta_q(t_1)\delta_q(t_2)\rangle\,\chi_{q,i}(t_1)\chi_{q,j}(t_2) \\ \times R_{ik}(t_1)R_{jk}(t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2.\end{aligned} \tag{A7}$$

We may invoke the Wiener-Khinchin theorem for a wide-sense stationary noise process to express the autocorrelation function of $\delta_q(t)$ as the Fourier transform of its PSD: $\langle\delta_q(t_1),\delta_q(t_2)\rangle = \frac{1}{2\pi}\int_{-\infty}^{\infty}S_q(\omega)\mathrm{e}^{i\omega(t_2-t_1)}\,\mathrm{d}\omega$. If we further define

$$R_{q,k}(\omega) \equiv \sum_i\int_0^T \chi_{q,i}(t)R_{ik}(t)\mathrm{e}^{i\omega t}\,\mathrm{d}t, \tag{A8}$$

we can finally compactly write the gate infidelity as

$$\langle\mathcal{I}\rangle \approx \frac{1}{2\pi}\sum_q\int_{-\infty}^{\infty}S_q(\omega)F_q(\omega)\,\mathrm{d}\omega, \tag{A9}$$

where $F_q(\omega) \equiv \sum_k\left|R_{q,k}(\omega)\right|^2$.

Appendix B: Numerical optimization method

We describe here the details of our numerical optimization. We used Julia's DiffEqFlux package to create a feedforward deep neural network with one input neuron, two output neurons, and two hidden layers with 16 neurons each. Our goal is to minimize the cost given in Eq. (31). The infidelity integral of a noise process q in the first two terms of Eq. (31) can be expressed as

$$\begin{aligned}\langle\mathcal{I}_q\rangle \approx \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_0^T\int_0^T \left(\Lambda(t_1)\vec{\chi}_q(t_1)\right)^\intercal\Lambda(t_2)\vec{\chi}_q(t_2) \\ \times S_q(\omega)e^{i\omega(t_1-t_2)}\,\mathrm{d}t_1\,\mathrm{d}t_2\,\mathrm{d}\omega,\end{aligned} \tag{B1}$$

where $\vec{\chi} = [\chi_{q,X},\chi_{q,Y},\chi_{q,Z}]^\intercal$ is the error sensitivity vector. In the main text, the noise PSD assumes one of two nontrivial forms: $\frac{A}{\omega}$ and $\frac{A\omega_c}{\omega^2}$. We can evaluate the frequency integrals analytically which are given by

$$\int_{\omega_0}^{\omega_c}\frac{A}{\omega}e^{i\omega t}\,\mathrm{d}t = 2\left(\mathrm{Ci}\left(\omega_ct\right) - \mathrm{Ci}\left(\omega_ot\right)\right), \tag{B2}$$

$$\int_{\omega_c}^{\infty}\frac{A\omega_c}{\omega^2}e^{i\omega t}\,\mathrm{d}t = -\pi\omega_ct + 2\cos\left(\omega_ct\right) + 2\omega_ct\,\mathrm{Si}\left(\omega_ct\right), \tag{B3}$$

where $\mathrm{Ci}(t)$ and $\mathrm{Si}(t)$ are the cosine and sine integral function, respectively. Let us define $g_q(t_1-t_2) = \frac{1}{2\pi}\int_{-\infty}^{\infty}S_q(\omega)e^{i\omega(t_1-t_2)}\,\mathrm{d}\omega$. This allows us to express Eq. (B1) as

$$\int_0^T\int_0^T g_q(t_1-t_2)\left(\Lambda(t_1)\vec{\chi}_q(t_1)\right)^\intercal\Lambda(t_2)\vec{\chi}_q(t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2. \tag{B4}$$

We can approximate the integrals by converting them into a series of matrix multiplications. In particular, we can treat each time integral as an integral operator which has g_q as its kernel and takes in $\boldsymbol{v}_q = \Lambda\vec{\chi}_q$ as input. Therefore, the average infidelity may be rewritten in the following bilinear form

$$\langle\mathcal{I}_q\rangle \approx \boldsymbol{v}_q^\intercal\mathbb{L}\boldsymbol{v}_q, \tag{B5}$$

where $\mathbb{L}$ is a matrix that approximates the double time integral.

In our work, the cost is completely vectorized by evaluating the cost terms in evenly spaced intervals of time. The infidelity integrals are evaluated using Eq. (B5) while derivatives, which are used in evaluating quantities such as Ω in Eq. (26), are implemented using finite differences. Thus, the speed and accuracy of optimization may be controlled by choosing an appropriate level of time discretization.

[1] R. J. Gordon and S. A. Rice, Annual Review of Physical Chemistry **48**, 601 (1997).
[2] A. Assion, T. Baumert, M. Bergt, T. Brixner, B. Kiefer, V. Seyfried, M. Strehle, and G. Gerber, Science **282**, 919 (1998).
[3] S. A. Rice and S. P. Shah, Phys. Chem. Chem. Phys. **4**, 1683 (2002).
[4] F. Poggiali, P. Cappellaro, and N. Fabbri, Phys. Rev. X **8**, 021059 (2018).
[5] M. M. Müller, S. Gherardini, and F. Caruso, Scientific Reports **8**, 14278 (2018).
[6] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2010).
[7] L. Viola, E. Knill, and S. Lloyd, Phys. Rev. Lett. **82**, 2417 (1999).
[8] K. Khodjasteh and D. A. Lidar, Phys. Rev. Lett. **95**, 180501 (2005).
[9] M. J. Biercuk, H. Uys, A. P. VanDevender, N. Shiga, W. M. Itano, and J. J. Bollinger, Nature **458**, 996 (2009).
[10] G. de Lange, Z. H. Wang, D. Ristè, V. V. Dobrovitski, and R. Hanson, Science **330**, 60 (2010).
[11] B. Naydenov, F. Dolde, L. T. Hall, C. Shin, H. Fedder, L. C. L. Hollenberg, F. Jelezko, and J. Wrachtrup, Phys. Rev. B **83**, 081201 (2011).
[12] J. Bylander, S. Gustavsson, F. Yan, F. Yoshihara, K. Harrabi, G. Fitch, D. G. Cory, Y. Nakamura, J.-S. Tsai, and W. D. Oliver, Nature Physics **7**, 565 (2011).
[13] M. H. Levitt, Progress in Nuclear Magnetic Resonance Spectroscopy **18**, 61 (1986).
[14] K. R. Brown, A. W. Harrow, and I. L. Chuang, Phys. Rev. A **70**, 052318 (2004).
[15] L. M. K. Vandersypen and I. L. Chuang, Rev. Mod. Phys. **76**, 1037 (2005).
[16] X. Wang, L. S. Bishop, J. P. Kestner, E. Barnes, K. Sun, and S. Das Sarma, Nature Communications **3**, 997 (2012).
[17] J. P. Kestner, X. Wang, L. S. Bishop, E. Barnes, and S. Das Sarma, Phys. Rev. Lett. **110**, 140502 (2013).
[18] F. A. Calderon-Vargas and J. P. Kestner, Phys. Rev. Lett. **118**, 150502 (2017).
[19] R. W. Simmonds, K. M. Lang, D. A. Hite, S. Nam, D. P. Pappas, and J. M. Martinis, Phys. Rev. Lett. **93**, 077003 (2004).
[20] O. E. Dial, M. D. Shulman, S. P. Harvey, H. Bluhm, V. Umansky, and A. Yacoby, Phys. Rev. Lett. **110**, 146804 (2013).
[21] J. Yoneda, K. Takeda, T. Otsuka, T. Nakajima, M. R. Delbecq, G. Allison, T. Honda, T. Kodera, S. Oda, Y. Hoshi, N. Usami, K. M. Itoh, and S. Tarucha, Nature Nanotechnology **13**, 102 (2018).
[22] C. Kabytayev, T. J. Green, K. Khodjasteh, M. J. Biercuk, L. Viola, and K. R. Brown, Phys. Rev. A **90**, 012316 (2014).
[23] U. Güngördü and J. P. Kestner, Phys. Rev. B **98**, 165301 (2018).
[24] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, Journal of Magnetic Resonance **172**, 296 (2005).
[25] D. Daems, A. Ruschhaupt, D. Sugny, and S. Guérin, Phys. Rev. Lett. **111**, 050404 (2013).
[26] Y. Guo, D. Dong, and C.-C. Shu, Phys. Chem. Chem. Phys. **20**, 9498 (2018).
[27] U. Güngördü and J. P. Kestner, Phys. Rev. A **100**, 062310 (2019).
[28] T. Nöbauer, A. Angerer, B. Bartels, M. Trupke, S. Rotter, J. Schmiedmayer, F. Mintert, and J. Majer, Phys. Rev. Lett. **115**, 190801 (2015).
[29] E. Barnes, X. Wang, and S. Das Sarma, Scientific Reports **5**, 12685 (2015).
[30] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, Phys. Rev. Lett. **103**, 110501 (2009).
[31] X. Chen, E. Torrontegui, and J. G. Muga, Phys. Rev. A **83**, 062116 (2011).
[32] B. T. Torosov and N. V. Vitanov, Phys. Rev. A **83**, 053420 (2011).
[33] T. J. Green, J. Sastrawan, H. Uys, and M. J. Biercuk, New Journal of Physics **15**, 095004 (2013).
[34] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems (2015), software available from tensorflow.org.
[35] M. Innes, E. Saba, K. Fischer, D. Gandhi, M. C. Rudilosso, N. M. Joy, T. Karmali, A. Pal, and V. Shah, Fashionable modelling with flux (2018), arXiv:1811.01457 [cs.PL].
[36] H. Ball, M. J. Biercuk, A. R. R. Carvalho, J. Chen, M. Hush, L. A. D. Castro, L. Li, P. J. Liebermann, H. J. Slatyer, C. Edmunds, V. Frey, C. Hempel, and A. Milne, Quantum Science and Technology **6**, 044011 (2021).
[37] C. D. B. Bentley, H. Ball, M. J. Biercuk, A. R. R. Carvalho, M. R. Hush, and H. J. Slatyer, Advanced Quantum Technologies **3**, 2000044 (2020).
[38] Y. Baum, M. Amico, S. Howell, M. Hush, M. Liuzzi, P. Mundada, T. Merkh, A. R. Carvalho, and M. J. Biercuk, PRX Quantum **2**, 040324 (2021).
[39] A. R. R. Carvalho, H. Ball, M. J. Biercuk, M. R. Hush, and F. Thomsen, Phys. Rev. Applied **15**, 064054 (2021).
[40] U. Güngördü and J. P. Kestner, Robust quantum gates using smooth pulses and physics-informed neural networks (2020), arXiv:2011.02512 [quant-ph].
[41] D. W. Kanaar, S. Wolin, U. Güngördü, and J. P. Kestner, Phys. Rev. B **103**, 235314 (2021).
[42] P. Dutta and P. M. Horn, Rev. Mod. Phys. **53**, 497 (1981).
[43] K. W. Chan, W. Huang, C. H. Yang, J. C. C. Hwang, B. Hensen, T. Tanttu, F. E. Hudson, K. M. Itoh, A. Laucht, A. Morello, and A. S. Dzurak, Phys. Rev. Applied **10**, 044017 (2018).
[44] D. J. Van Harlingen, T. L. Robertson, B. L. T. Plourde, P. A. Reichardt, T. A. Crane, and J. Clarke, Phys. Rev. B **70**, 064517 (2004).
[45] F. Yoshihara, K. Harrabi, A. O. Niskanen, Y. Nakamura, and J. S. Tsai, Phys. Rev. Lett. **97**, 167001 (2006).
[46] M. Kim, H. J. Mamin, M. H. Sherwood, K. Ohno, D. D.

Awschalom, and D. Rugar, Phys. Rev. Lett. **115**, 087602 (2015).
[47] J. Koch, T. M. Yu, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. A **76**, 042319 (2007).
[48] A. Laucht, R. Kalra, S. Simmons, J. P. Dehollain, J. T. Muhonen, F. A. Mohiyaddin, S. Freer, F. E. Hudson, K. M. Itoh, D. N. Jamieson, J. C. McCallum, A. S. Dzurak, and A. Morello, Nature Nanotechnology **12**, 61 (2017).
[49] N. A. Gershenfeld and I. L. Chuang, Science **275**, 350 (1997).
[50] H. R. Lewis, Phys. Rev. Lett. **18**, 510 (1967).
[51] H. R. Lewis and W. B. Riesenfeld, Journal of Mathematical Physics **10**, 1458 (1969).
[52] The sign difference in comparison with Eq. (24) is a consequence of our choice of parameterization for the dynamical invariant eigenvectors and is irrelevant since only the magnitude of $\dot{\vec{r}}$ matters.
[53] J. Zeng, X.-H. Deng, A. Russo, and E. Barnes, New Journal of Physics **20**, 033011 (2018).
[54] J. Zeng and E. Barnes, Phys. Rev. A **98**, 012301 (2018).
[55] J. Zeng, C. H. Yang, A. S. Dzurak, and E. Barnes, Phys. Rev. A **99**, 052321 (2019).
[56] D. Buterakos, S. Das Sarma, and E. Barnes, PRX Quantum **2**, 010341 (2021).
[57] W. Dong, F. Zhuang, S. E. Economou, and E. Barnes, PRX Quantum **2**, 030333 (2021).
[58] E. Barnes, F. A. Calderon-Vargas, W. Dong, B. Li, J. Zeng, and F. Zhuang, Dynamically corrected gates from geometric space curves (2021), arXiv:2103.16015.
[59] F. Zhuang, J. Zeng, S. E. Economou, and E. Barnes, Quantum **6**, 639 (2022).
[60] E. Kreyszig, *Differential Geometry*, Dover Books on Mathematics (Dover Publications, 2013).
[61] L. M. Norris, D. Lucarelli, V. M. Frey, S. Mavadia, M. J. Biercuk, and L. Viola, Phys. Rev. A **98**, 032315 (2018).
[62] E. Knill, R. Laflamme, R. Martinez, and C.-H. Tseng, Nature **404**, 368 (2000).
[63] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Phys. Rev. A **77**, 012307 (2008).
[64] B. R. Johnson, M. P. da Silva, C. A. Ryan, S. Kimmel, J. M. Chow, and T. A. Ohki, New Journal of Physics **17**, 113019 (2015).
[65] D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta, Phys. Rev. A **96**, 022330 (2017).
[66] These constraints are not necessary but they help with the overall experimental feasibility of the pulses we produce.
[67] Y. LeCun, Y. Bengio, and G. Hinton, Nature **521**, 436 (2015).
[68] J. Schmidhuber, Neural Networks **61**, 85 (2015).
[69] A. Krizhevsky, I. Sutskever, and G. E. Hinton, Commun. ACM **60**, 84–90 (2017).
[70] G. Cybenko, Mathematics of Control, Signals and Systems **2**, 303 (1989).
[71] K. Hornik, Neural Networks **4**, 251 (1991).
[72] In feedforward neural networks, each neural connection adds one parameter. Furthermore, with the exception of the input neurons, each neuron contains an additional bias parameter. Thus, if we have a 1-3-2 network (one input neuron, one hidden layer with three neurons, and two output neurons), we have $(1*3+3)+(3*2+2) = 14$ free parameters to optimize. In our work, we used a 1-16-16-2 network which has $(1*16+16)+(16*16+16)+(16*2+2) = 338$ free parameters.
[73] H. K. Cummins and J. A. Jones, New Journal of Physics **2**, 006 (2000).
[74] H. Cummins and J. Jones, Journal of Magnetic Resonance **148**, 338 (2001).
[75] H. K. Cummins, G. Llewellyn, and J. A. Jones, Phys. Rev. A **67**, 042308 (2003).
[76] S. Wimperis, Journal of Magnetic Resonance, Series A **109**, 221 (1994).
[77] M. Bando, T. Ichikawa, Y. Kondo, and M. Nakahara, Journal of the Physical Society of Japan **82**, 014004 (2013).
[78] A. Nazir, T. P. Spiller, and W. J. Munro, Phys. Rev. A **65**, 042303 (2002).
[79] A. Blais and A.-M. S. Tremblay, Phys. Rev. A **67**, 012308 (2003).
[80] Y. Ota and Y. Kondo, Phys. Rev. A **80**, 024302 (2009).
[81] S.-B. Zheng, C.-P. Yang, and F. Nori, Phys. Rev. A **93**, 032313 (2016).
[82] P. Solinas, P. Zanardi, and N. Zanghì, Phys. Rev. A **70**, 042316 (2004).
[83] A. Carollo, I. Fuentes-Guridi, M. F. m. c. Santos, and V. Vedral, Phys. Rev. Lett. **92**, 020402 (2004).
[84] S.-L. Zhu and P. Zanardi, Phys. Rev. A **72**, 020301 (2005).
[85] J. Dajka, M. Mierzejewski, and J. Łuczka, Journal of Physics A: Mathematical and Theoretical **41**, 012001 (2007).
[86] M. Johansson, E. Sjöqvist, L. M. Andersson, M. Ericsson, B. Hessmo, K. Singh, and D. M. Tong, Phys. Rev. A **86**, 062322 (2012).
[87] R. K. Colmenar, U. Güngördü, and J. Kestner, Invalidating the robustness conjecture for geometric quantum gates (2021), arXiv:2105.02882.
[88] S.-L. Zhu and Z. D. Wang, Phys. Rev. Lett. **89**, 097902 (2002).
[89] P. Z. Zhao, X.-D. Cui, G. F. Xu, E. Sjöqvist, and D. M. Tong, Phys. Rev. A **96**, 052316 (2017).
[90] J. Xu, S. Li, T. Chen, and Z.-Y. Xue, Frontiers of Physics **15**, 41503 (2020).
[91] T. Ichikawa, M. Bando, Y. Kondo, and M. Nakahara, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **370**, 4671 (2012).
[92] J. R. Schrieffer and P. A. Wolff, Phys. Rev. **149**, 491 (1966).
[93] N. E. Bonesteel, D. Stepanenko, and D. P. DiVincenzo, Phys. Rev. Lett. **87**, 207901 (2001).
[94] V. M. Frey, S. Mavadia, L. M. Norris, W. de Ferranti, D. Lucarelli, L. Viola, and M. J. Biercuk, Nature Communications **8**, 2189 (2017).