Bayesian optimization for inverse problems in time-dependent quantum dynamics

Z. Deng\textsuperscript{a,*}, I. Tutunnikov\textsuperscript{b,*}, I. Sh. Averbukh\textsuperscript{b}, M. Thachuk\textsuperscript{a}, and R. V. Krems\textsuperscript{a,c}

\textsuperscript{a}Department of Chemistry, University of British Columbia, Vancouver, B.C. V6T 1Z1, Canada
\textsuperscript{b}AMOS and Department of Chemical and Biological Physics, The Weizmann Institute of Science, Rehovot, 7610001, Israel
\textsuperscript{c}Quantum Matter Institute, University of British Columbia, Vancouver, BC, Canada V6T 1Z4

(Dated: August 17, 2020)

* These authors have contributed equally.
Abstract

We demonstrate an efficient algorithm for inverse problems in time-dependent quantum dynamics based on feedback loops between Hamiltonian parameters and the solutions of the Schrödinger equation. Our approach formulates the inverse problem as a target vector estimation problem and uses Bayesian surrogate models of the Schrödinger equation solutions to direct the optimization of feedback loops. For the surrogate models, we use Gaussian processes with vector outputs and composite kernels built by an iterative algorithm with Bayesian information criterion (BIC) as a kernel selection metric. The outputs of the Gaussian processes are designed to model an observable simultaneously at different time instances. We show that the use of Gaussian processes with vector outputs and the BIC-directed kernel construction reduce the number of iterations in the feedback loops by, at least, a factor of 3. We also demonstrate an application of Bayesian optimization for inverse problems with noisy data. To demonstrate the algorithm, we consider the orientation and alignment of polyatomic molecules SO$_2$ and propylene oxide (PPO) induced by strong laser pulses. We use simulated time evolutions of the orientation or alignment signals to determine the relevant components of the molecular polarizability tensors to within 1% accuracy. We show that, for the five independent components of the polarizability tensor of PPO, this can be achieved with as few as 30 quantum dynamics calculations.
I. INTRODUCTION

The inverse problem in quantum dynamics aims to determine Hamiltonian parameters from experimental observables [1]. This goal underlies spectroscopy applications [2] aiming to determine molecular properties, such as vibrational and rotational constants. Kraitchman analysis [3] can be considered an example of an inverse problem that has been used to determine heavy atom structures of molecules (for a recent example involving a complex chiral molecule see [4]). Electron and x-ray diffraction measurements, which inherently rely on data inversion, have been applied in molecular physics to probe ultrafast molecular processes and recover structures of small gas-phase molecules (see, e.g., [5] and references therein). The Coulomb explosion technique has been improved to allow retrieval of detailed and extremely precise information on molecular properties from experimental measurements (see e.g. [6, 7] and references therein). For many quantum systems, it is possible to establish the relationship between the Hamiltonian parameters and observables using model Hamiltonians and approximations permitting closed-form expressions [8, 9] or simplified numerical calculations of observables, which make the inverse problem feasible. In the present work, we consider an approach to solving inverse problems in quantum dynamics that does not rely on any assumptions about the Hamiltonian and that can be applied to problems, where calculating dynamical observables is extremely time consuming. This can be used for applications in molecular scattering theory, such as the empirical construction of potential energy surfaces for complex molecular systems, including molecular collision and reaction complexes, calibration of theoretical approaches, and studies of mechanisms of energy transfer and chemical reactions in molecular collisions.

The present approach is based on a feedback loop, which adjusts Hamiltonian parameters iteratively to bring predictions from the solutions of the Schrödinger equation into agreement with experimental observations [11]. This approach is widely exploited in optimal control experiments aiming to design external field parameters that yield desired quantum dynamics. While optimal control has been successfully applied to many quantum systems, the inverse problem in quantum dynamics remains a significant challenge. Solving the inverse quantum problems by iterative feedback loops is challenging because: (1) each iteration requires the numerical solution of the nuclear Schrödinger equation, which is time-consuming; (2) finding optimal Hamiltonian parameters requires global optimization within an unknown range of
values; (3) the number of parameters and the dimensionality of the Hamiltonian increase with the complexity of the quantum system, as does the numerical difficulty of solving the Schrödinger equation.

A major thrust of recent work in molecular dynamics research has been to combine traditional simulation methods with machine learning (ML); see, e.g., [12–23]. ML offers new tools for solving inverse problems. For example, it was recently shown that the inverse scattering problem in quantum reaction dynamics can be solved by Bayesian optimization (BO) based on Gaussian processes (GP) as surrogate models of the solutions of the Schrödinger equation [11]. BO is an efficient, gradient-free optimization approach designed for finding the global extrema of non-convex black-box functions [24, 25]. In this example, the black-box function is the departure of $f(x)$ from a target, with $f(x)$ representing the dependence of a calculated observable on the Hamiltonian parameters $x$:

$$f(x) \triangleq \left\{ \hat{H}(x)\Psi = E\Psi \rightarrow \text{Observable}(x) \right\}. \quad (1)$$

In [11], $x$ represented the potential energy surfaces (PES) determining reactive scattering dynamics. More generally, the vector $x$ can represent a set of any parameters that determine the Hamiltonian (e.g., PES, non-adiabatic couplings, molecular properties, molecule-field interactions).

GPs provide an efficient way to explore the dependence of $f$ on $x$ with a small number of quantum calculations. This was exploited in [11] to design the feedback loop

$$\begin{array}{c}
T \\
\circlearrowleft \\
\uparrow \\
\downarrow \\
x
\end{array} 
\xrightarrow{T} f(x) 
$$

(2)

where $T$ is a scalar quantity related to an observable. For example, the purpose of feedback loop (2) in [11] was to obtain the PES $x$ that, when used in a quantum scattering calculation, yielded the energy dependence of the reaction probabilities in agreement with previously published results. To achieve this, $T$ was set to represent the root mean square deviation of $f(x)$ from the reference energy dependence of the reaction probabilities and feedback loop (2) was guided by the minimization of a single parameter $T$. GPs make feedback loop (2) feasible. This is achieved by building a GP model $F(x)$ of $T(x)$ with a small number of quantum calculations. The model $F(x)$ is then used by BO to determine how $x$ should be modified in order to minimize $T(x)$ [11].

In the present work, we report three significant results. First, we demonstrate an im-
proved version of BO for inverse problems with time-dependent observables. We propose an algorithm that represents a time-dependent observable as a vector with \( m \) components corresponding to different time instances. This vector is modelled by a GP \( \mathcal{F}(\mathbf{x}) \) with \( m \) outputs. We show that BO based on such GPs converges much faster than algorithm (2) based on GPs \( \mathcal{F}(\mathbf{x}) \) with a scalar output. The vector algorithm can be illustrated by the feedback loop

\[
\begin{array}{c}
\mathbf{x} \\
\downarrow \quad f(\mathbf{x}, t) \\
T \quad \mathcal{F}(\mathbf{x})
\end{array}
\]

where the observable \( f \) is calculated at different times \( t \) for a given \( \mathbf{x} \) and the GP model is trained by these calculations to model the observable \textit{simultaneously} at different times. The target quantity \( T \) is then calculated from \( \mathcal{F}(\mathbf{x}) \) inheriting the properties of the GP with \( m \) outputs, which are used to direct BO. Second, we show that BO can be applied for inverse problems with noisy observables. This is important for applications using raw experimental data for determining the Hamiltonian parameters. Third, we explore the effect of GP model complexity on the efficiency of BO for quantum inverse problems and demonstrate that a model-selection method based on Bayesian information criterion (BIC) can be used to reduce the number of iterations in feedback loop (3). This model-selection algorithm was previously used to enhance the accuracy of GP models for pattern recognition [26, 27], generalization [28], interpolation and extrapolation [28, 29]. Here, we demonstrate that the BIC can also be used to enhance the efficiency of BO. We propose a specific algorithm that identifies the proper GP model complexity before BO is initialized and iterates the feedback loops with the model complexity thus determined.

To demonstrate the algorithms proposed here, we consider the orientation and alignment of polyatomic molecules \( \text{SO}_2 \) and propylene oxide \( \text{CH}_3\text{CHCH}_2\text{O} \) (denoted hereafter as PPO) induced by strong laser pulses. We use the time dependence of the orientation and/or alignment signal to determine the relevant components of the polarizability tensor of the molecule. There are several traditional techniques for measuring the average molecular polarizability, including dielectric constant and refractive index measurements, as well as polarizability anisotropy with the help of Rayleigh scattering, separately, or in combination with Kerr effect measurements. In some cases, principal polarizabilities can be recovered by combining the data obtained from two or more independent measurements. For an in-depth
review of these and other modern experimental techniques on electric-dipole polarizabilities, see [30]. The method proposed here may allow measurements of molecular polarizabilities in the gas phase in a single experimental setup. Moreover, in the case of chiral molecules, our approach allows the determination of the off-diagonal polarizability tensor components (when expressed in the frame of principal axes of inertia tensor). The off-diagonal elements determine the relationship between the principal axes of inertia and polarizability, which is important for various applications (e.g., in molecular dynamics simulations). We examine the relative importance of the orientation and alignment signals for the inverse problem and show that certain components of the polarizability tensor can be determined from the signal evolution over a short time interval. Using these examples, we examine the convergence of BO and the accuracy of the Hamiltonian parameters that can be determined with this approach.

II. THEORY

Feedback loops (2) and (3) require three essential ingredients: numerical integration of the Schrödinger equation with a given $x$ to yield $f(x, t)$, construction of GP models of $f(x, t)$, or some quantity related to $f(x, t)$, using the results of the quantum calculations as training data, and Bayesian optimization directed by these GP models to find the optimal $x = x^{\text{opt}}$. The quantum problem must be solved at each iteration of the loops. Quantum dynamics calculations are time-consuming and should be expected to bottleneck the feedback loops. Therefore, it is important to develop an optimization algorithm that converges to $x^{\text{opt}}$ with as few iterations as possible.

A. Quantum dynamics calculations

We consider the asymmetric top molecules treated as rigid tops with three distinct moments of inertia ($I_a < I_b < I_c$) corresponding to three principal axes of inertia $a$, $b$ and $c$ [31, 32]. The interaction with a moderately intense nonresonant short laser pulse is modeled using the potential (in atomic units, a.u.) [33, 34]

$$U = -d_i \tilde{E}_i - \frac{1}{2} \alpha_{ij} \tilde{E}_i \tilde{E}_j,$$  \hspace{1cm} (4)
where \(d_i\) are the components of the molecular dipole moment, \(\alpha_{ij}\) are the components of the molecular polarizability and \(\tilde{E}_i\) are the components of the rapidly oscillating electric field. The central frequency of an optical laser pulse is of the order of \(\propto 10^{15}\) Hz, while the typical rotational frequency of small molecules is of the order of \(\hbar/I \propto 10^{12}\) Hz, where \(\hbar\) is the Planck’s constant and \(I\) is the average moment of inertia. Therefore, we average the interaction energy over the optical cycle, which leads to \(U = -\alpha_{ij}E_iE_j/4\), where \(E_i\) are the components of the slowly varying envelope of the laser pulse. We use the impulsive approximation to treat the interaction of molecules with the laser fields. More details are given in the supplementary information.

We use symmetric top wave functions \(|JKM\rangle\), defined in [31], to express the Hamiltonian of the asymmetric-top molecule including the interaction term. The Hamiltonian is then transformed to the basis of asymmetric-top eigenfunctions that diagonalizes the kinetic part of the Hamiltonian, allowing for an efficient field-free propagation of wave packets. We aim to explore inverse problems with quantum dynamics calculations for both zero and finite temperatures. For zero temperature results, we propagate all relevant quantum states in time and compute the resulting expectation values. At temperatures near 300 K, the number of populated quantum states for molecules such as SO\(_2\) or PPO becomes prohibitively large. Therefore, for all non-zero temperature results, we use a random phase wave functions (RPWF) approach [35]. Instead of propagating each thermally populated state, multiple states \(|\psi_m\rangle\) are prepared as mixtures of all basis states each weighted by the Boltzmann factor and assigned a random phase factor, that is

\[
|\psi_m\rangle = \sum_{n=1}^{n_{\text{max}}} \sqrt{\frac{e^{-\varepsilon_n/k_B T}}{Q}} e^{-i\phi_n} |n\rangle ,
\]

(5)

where \(\varepsilon_n\) is the kinetic energy of the basis state \(|n\rangle\), \(k_B\) is the Boltzmann constant, \(T\) is the temperature, \(Q\) is the partition function, and \(\phi_n\) is a random variable uniformly distributed over \([0, 2\pi]\). The time evolution of wave packets in Eq. (5) is determined by the time-dependence of the stationary states \(|n\rangle\). A total of \(N\) states (5) is used and the expectation value of any observable \(\hat{O}\) is evaluated as

\[
\langle \hat{O}(t) \rangle = \frac{1}{N} \sum_m \langle \psi_m(t) | \hat{O} | \psi_m(t) \rangle .
\]

(6)

This approach was shown [35, 36] to be efficient for dynamics simulations at elevated temperatures where it converges with \(N \lesssim 100\).
In this study, we consider two molecules (SO$_2$ and PPO) and use three observables: the optical birefringence, the degree of alignment and the orientation factor. When anisotropic molecules in the gas phase are excited by a short non-resonant laser pulse, the interaction between the induced dipole and the electric field tends to order molecules along the polarization direction of the laser field. Macroscopically, molecular alignment leads to measurable optical birefringence of the gas [36, 39, 40]. For linear molecules, the degree of birefringence is determined by the degree of alignment $\langle \cos^2 \theta \rangle$, where $\theta$ is the angle between the molecular axis and the polarization direction. For molecules with three distinct principal axes and three distinct polarizabilities, the optical birefringence is given by the weighted average of three alignment terms

$$B(t) \propto \sum_{i=x,y,z} k_i \langle \cos^2(\theta_{iZ}) \rangle,$$

(7)

where $k_z = \alpha^{xx} + \alpha^{yy}$, $k_x = -(2\alpha^{xx} + \alpha^{yy})$, $k_y = \alpha^{xx} - 2\alpha^{yy}$, $\alpha^{zz} = \alpha_{zz} - \alpha_{xx}$, $\alpha^{yz} = \alpha_{zz} - \alpha_{yy}$, and $\theta_{iZ}$ are the angles between the corresponding axes of the rotating molecule fixed frame and the laser pulse polarization (chosen to be along the laboratory $Z$ axis). The birefringence depends explicitly on two differences of polarizabilities, $\alpha^{xx}$ and $\alpha^{yy}$. We use the birefringence (7) for the case of alignment of SO$_2$ by a single linearly polarized laser pulse [33, 37, 38]. In the case of alignment of PPO by a single laser pulse, we use the time dependence of the degree of alignment of the molecular axis $a$ (which is close to the most polarizable axis) with respect to the laboratory $Z$ axis $\langle \cos^2(\theta_{aZ}) \rangle$. The supplementary information provides more details.

In the case of orientation of PPO, the observable is the expectation value of the projection of the molecular dipole vector $\mu$ on the laboratory $Z$ axis, $\mu_Z = \mu \cdot \hat{Z} = \mu \cos(\theta_{\mu Z})$, where $\mu = |\mu|$ is the magnitude of the dipole moment, $\hat{Z}$ is the unit vector along the laboratory $Z$ axis, and $\theta_{\mu Z}$ is the angle between the molecular dipole moment and the $Z$ axis. It was recently shown that chiral molecules can be oriented by a pair of delayed cross-polarized short laser pulses (or more generally by pulses with twisted polarization). The orientation is in the direction perpendicular to the plane defined by the two pulses (see [41] and references therein). We consider two pulses in the $XY$ plane, such that the orientation is along/against the $Z$ axis, and quantify orientation by the dipole orientation factor $\mu \langle \cos(\theta_{\mu Z}) \rangle$ [36, 42]. The orientation effect stems from the off-diagonal elements of the polarizability tensor (in
the frame of the principal axes of inertia tensor) \(\alpha_{ab}, \alpha_{ac}\) and \(\alpha_{bc}\), which is a distinct property of chiral molecules.

B. Gaussian process models

The algorithm described in the next section requires GP models as an intermediate step. A GP can be considered as a limit of a Bayesian neural network with an infinite number of hidden nodes [23, 43]. The inputs to the GP are \(N\) independent variables, collectively denoted by the vector \(\mathbf{x} = [x_1, ..., x_N]^T\). The output of a GP is a scalar function \(y(\mathbf{x})\). The purpose is to model an ensemble of \(n\) data points \(y = [Y_1, ..., Y_n]^T\) located at \(\mathbf{x} = [x_1, ..., x_n]^T\) of the \(N\)-dimensional variable space. It is assumed that these data points can be described as \(f(\mathbf{x}) + \varepsilon\), where \(f(\mathbf{x})\) is some function and \(\varepsilon\) is Gaussian-distributed noise with variance \(\sigma^2\).

At any \(\mathbf{x}\), there is a normal distribution \(P(y)\) of values \(y\). When a GP is trained, \(P(y)\) is conditioned by the \(n\) data points \(y\) at \([x_1, ..., x_n]^T\). The mean and variance of this conditional distribution at an arbitrary point \(\mathbf{x}_*\) are given by [23, 44]

\[
\mu_* = k_\mathbf{x}^\top (K + \sigma^2 I)^{-1} y, \quad (8)
\]

\[
\sigma_* = k(\mathbf{x}_*, \mathbf{x}_*) - k_\mathbf{x}^\top (K + \sigma^2 I)^{-1} k_\mathbf{x}, \quad (9)
\]

where \(k_\mathbf{x}\) is a vector with \(n\) entries \(k(\mathbf{x}_*, \mathbf{x}_i)\) and \(K\) is an \(n \times n\) matrix with elements \(k(\mathbf{x}_i, \mathbf{x}_j)\). Eq. (8) is used to predict the value of \(f(\mathbf{x})\) at \(\mathbf{x}_*\). The quantities \(k(\mathbf{x}_i, \mathbf{x}_j)\) are the kernels, which, with a particular choice of the GP prior [23, 44], represent the covariance of the normal distributions of \(y\) at \(\mathbf{x}_i\) and at \(\mathbf{x}_j\).

A particular mathematical form of the kernel function \(k(\mathbf{x}, \mathbf{x}')\) defines a GP model. The choice of this function is not unique. Unless specified otherwise (c.f., Section III.B), we use the functional form [45–47]

\[
k(\mathbf{x}, \mathbf{x}') = \frac{2^{1-v}}{\Gamma(v)} \left(\sqrt{2vr(\mathbf{x}, \mathbf{x}')}\right)^v \mathcal{K}_v \left(\sqrt{2vr(\mathbf{x}, \mathbf{x}')}\right) \quad (10)
\]

where \(r^2(\mathbf{x}, \mathbf{x}') = (\mathbf{x} - \mathbf{x}')^\top M (\mathbf{x} - \mathbf{x}')\) and \(M\) is a diagonal matrix with \(N\) parameters, one parameter for each dimension of \(\mathbf{x}\), \(\mathcal{K}_v\) is the modified Bessel function, \(\Gamma\) is the Gamma function, and \(v = 3/2\). This function is often referred to as the Matérn function. The parameters of the kernel function are found by maximizing the logarithm of the marginal
\[
\log \mathcal{L} = -\frac{1}{2} y^\top K^{-1} y - \frac{1}{2} \log |K| - \frac{n}{2} \log 2\pi.
\] (11)

C. Bayesian optimization with scalar and multiple-output GPs

There are six independent matrix elements that determine the polarizability tensor of a general polyatomic molecule in the frame of the principal axes of inertia tensor $\alpha_{aa}, \alpha_{bb}, \alpha_{cc}, \alpha_{ab}, \alpha_{ac}, \alpha_{bc}$. Depending on the symmetry of the molecule, some of the polarizability tensor matrix elements vanish. In addition, different observables are determined by different parts of the polarizability tensor. We define the vector $x$ as comprising the minimum number of independent parameters of the polarizability tensor determining an observable. This is the maximum amount of information that can be inferred by solving the inverse problem. For example, the laser-field alignment of the planar molecule SO$_2$ is determined by two parameters: $\alpha^{ab} = \alpha_{aa} - \alpha_{bb}$ and $\alpha^{ac} = \alpha_{aa} - \alpha_{cc}$. For this case, we define $x$ as $x = [\alpha^{ab}, \alpha^{ac}]^\top$. The orientation of PPO by laser pulses with twisted polarization is determined by five parameters: $\alpha^{ab}, \alpha^{ac}, \alpha_{ab}, \alpha_{ac}$ and $\alpha_{bc}$. For this case, we define $x$ as the five-dimensional vector $x = [\alpha^{ab}, \alpha^{ac}, \alpha_{ab}, \alpha_{ac}, \alpha_{bc}]^\top$. Given $x$, quantum dynamics calculations described in Section I.A produce the observable (the birefringence signal probing alignment or the orientation factor characterizing orientation) as a function of time $t$. This observable is hereafter denoted $f(x, t)$.

The purpose of BO is to find $x = x^{\text{opt}}$ that leads to $f(x^{\text{opt}}, t)$ in agreement with some reference function $f_{\text{ref}}(t)$ giving the time dependence of the observable. The departure of $f(x, t)$ from $f_{\text{ref}}(t)$ can be quantified by the root mean square error (RMSE)

\[
\mathcal{E}(x) = \frac{1}{m} \left\{ \sum_i^m [f(x, t_i) - f_{\text{ref}}(t_i)]^2 \right\}^{1/2}
\] (12)

with $t_i$ representing discretized points of time along a chosen time interval. The vector $x^{\text{opt}}$ corresponds to the minimum of $\mathcal{E}(x)$. We treat the quantities $f(x, t)$ and $f_{\text{ref}}(t)$ as vectors $f$ and $f_{\text{ref}}$ with $m$ entries corresponding to discretized time values $t_i$. Feedback loops (2) and (3) provide two different algorithms to find the minimum of $\mathcal{E}(x)$.

For loop (2), we define $T(x)$ as

\[
T(x) = -\arccos \left( \frac{f \cdot f_{\text{ref}}}{|f||f_{\text{ref}}|} \right),
\] (13)
which quantifies the similarity between the vectors. Loop (2) uses a scalar, single-output GP model of $T(x)$ thus defined. The algorithm starts by evaluating $T(x)$ at a small number of randomly chosen values $[x_1, ..., x_n]^T$. The results of quantum dynamic calculations for these $x$ are used to build a GP model of $T(x)$. The subsequent calculation of $T(x)$ is performed at the value of $x = \tilde{x}$ that corresponds to the maximum of the acquisition function defined as

$$\alpha(x) = \mu(x) + \kappa \sigma(x), \quad (14)$$

where $\mu$ and $\sigma$ are given by Eqs. (8) and (9), respectively, and $\kappa$ is a fixed parameter. The value of $T(\tilde{x})$ is added to the training points for the GP to generate a new GP with improved values of $\mu$ and $\sigma$ and the process is iterated. The acquisition function defined by Eq. (14) provides a balance between exploration of the entire variable parameter space, driven by the second term, and exploitation of a promising part of the parameter space, driven by the maximization of the first term. As the number of points in the $x$ space increases, $\sigma(x)$ decreases and the maximum of the acquisition function tends to the maximum of $T(x)$ in Eq. (13). Note that $T(x)$ in loop (2) can be chosen as $\mathcal{E}(x)$, as was done in the previous work [11]. In the present work, we found that the choice of the scalar function (13) instead of $\mathcal{E}(x)$ reduces the number of BO iterations.

Loop (3) minimizes the quadratic Euclidean distance, or the square of the 2-norm, between $f$ and $f_{\text{ref}}$, that is $T(x) = |f(x) - f_{\text{ref}}|^2$. Instead of using one GP model to represent $T(x)$, at each iteration, algorithm (3) uses $m$ GP models trained to represent the $m$ components of the vector $f$, as described in [48]. The $m$ GP models are trained simultaneously, assuming the same $(x, x')$ covariance for different $t_i$. This leads to a more complex form of the acquisition function. Whereas in algorithm (2), $T(x)$ is directly represented by a GP, the square of the 2-norm in algorithm (3) follows a non-central Chi-squared distribution with $m$ degrees of freedom. As was shown in [48], this approach is more efficient for minimization of distances than the method using Eq. (14), because it uses the individual vector components for training GPs and leads to an acquisition function based on a non-symmetric distribution better suited for non-negative functions. To obtain the acquisition function, it is necessary to transform the means and uncertainties of the $m$ GPs into quantities characterizing the Chi-squared distribution and then map these quantities onto the parameters of the acquisition function. This was done in [48], where the authors used the results of [49].
to represent the Chi-squared distribution by an approximate Gaussian distribution. This yields the following result for the acquisition function suitable for the minimization of the 2-norm \[48\]:

\[
\alpha(x) = -\sqrt{\delta - \kappa \rho (m + \lambda) \gamma^2},
\]

where

\[
\lambda = \gamma^{-2} \sum_{i=1}^{m} [\mu_i(x) - f_{\text{ref},i}]^2,
\]

\[
\gamma = \sqrt{\frac{1}{m} \sum_{i=1}^{m} \sigma_i^2(x)}
\]

\[
\mu_i \text{ and } \sigma_i \text{ are the mean and the variance of the } i\text{th output of the GP with } m \text{ outputs,}
\]

\[
l = 1 - \frac{r_1 r_3}{3 r_2^2},
\]

with

\[
r_i = 2^{i-1} (i-1)! (m + i \lambda),
\]

\[
\rho = \frac{l r_2^2}{r_1} \left(1 - \frac{(1-l)(1-3l)}{4 r_1^2} r_2^2\right),
\]

and

\[
\delta = 1 + l(l+1) \left(\frac{r_2}{2 r_1^2} - (2-l)(1-3l) \frac{r_2^2}{8 r_1^4}\right).
\]

The parameter \(\kappa\) is used to determine the balance between exploration and exploitation in the optimization algorithm. In the present work, \(\kappa\) was set to 0.8 in both Eq. (14) and (15).

III. RESULTS

The present work considers two molecules: SO\(_2\) and propylene oxide (PPO). The laser-field alignment of the planar molecule SO\(_2\) is determined by two parameters: \(\alpha^{ab} = \alpha_{aa} - \alpha_{bb}\) and \(\alpha^{ac} = \alpha_{aa} - \alpha_{cc}\). The alignment and orientation of PPO is determined by five parameters [50]: \(\alpha^{ab}, \alpha^{ac}, \alpha_{ab}, \alpha_{ac}\) and \(\alpha_{bc}\). Our goal is to determine the values of these parameters based on a reference signal \(f_{\text{ref}}(t)\).
FIG. 1. The results of BO of the parameters $\alpha^{ab}$, $\alpha^{ac}$, $\alpha_{ab}$, $\alpha_{ac}$ and $\alpha_{bc}$ of PPO using the orientation factor as the reference signal. The vertical dotted line marks the beginning of BO. For illustration purposes, the optimization was repeated 20 times using different initial conditions. The shaded area to the left of the dotted line shows the standard deviation of RMSE (12) resulting from the distribution of these 20 calculations. The shaded area to the right of the vertical dotted line spans the range of RMSE from the minimum to maximum values. The solid curve is the average of 20 calculations. Upper panel: the scalar GP algorithm (2); Lower panel: algorithm (3) based on the multiple-output GP with a complex kernel represented by a linear combination of a Matérn and RQ functions (see Section III.B for details). Insets: the solid curve shows the reference signal $f_{\text{ref}}(t)$ used for BO and the broken curves – the final result computed with the optimized parameters of the polarizability tensor after 90 (upper panel) and 20 (lower panel) iterations.
For the reference signal $f_{\text{ref}}(t)$, we use the birefringence (for SO$_2$) and the alignment and orientation factors (for PPO) computed with the most accurately known theoretical values of the polarizability tensors (see the supplementary information for more details). Unless otherwise specified, the reference curves are the theoretical calculations at zero temperature. For the calculations at zero temperature, the molecules are assumed to be excited by a pulse with the duration 20 fs and the field intensity $5 \times 10^{13}$ W/cm$^2$. We also explore the effect of averaging over quantum states required to simulate the experiments at finite temperature. For the calculations at finite temperature, the pulse duration was chosen to be 100 fs and the field intensity $5 \times 10^{13}$ W/cm$^2$. The initial optimization was performed blindly with one co-author producing the reference results and another performing optimization without information on the reference polarizability tensor parameters.

In the present section, we consider noiseless functions $f_{\text{ref}}(t)$. In the absence of noise, $\sigma$ in Eqs. (8) and (9) must be set to zero. The effect of noise is considered in Section II.A. We refer to algorithm (2) as the ‘scalar’ algorithm and (3) as the ‘vector’ algorithm. For all calculations presented here, we discretize the time variable to represent $f(x, t)$ and $f_{\text{ref}}(t)$ by 400 equally spaced points in the indicated time interval beginning after the laser pulse (alignment) or the second laser pulse (orientation).

Figure 1 illustrates the main features and convergence of BO. The algorithm illustrated in Figure 1 begins with a fixed number (50 in the upper panel and 20 in the lower panel) of quantum dynamics calculations at different values of $x = [\alpha_{ab}, \alpha_{ac}, \alpha_{bc}, \alpha_{bc}, \alpha_{bc}]^\top$ for PPO. The values of these parameters are chosen using the Latin hypercube spacing algorithm to avoid clustering in this five-dimensional space [51, 52]. The reference function $f_{\text{ref}}(t)$ is the expectation value of the $Z$-component of the molecular dipole moment shown in the insets of Figure 1. The deviation of the calculated signal from the reference curve is quantified by the value of RMSE (12). The final time-dependence of $\langle \mu_Z \rangle$ calculated with the optimized parameters of the polarizability signal is shown by the broken curves in the insets of Figure 1.

BO begins at the vertical dotted lines of Figure 1. To illustrate the efficacy and convergence of BO, we repeat BO with 20 different sets of the initial quantum calculations. The shaded area to the left of the dotted line shows the standard deviation of RMSE (12) resulting from the distribution of these 20 sets. Each of the initial conditions leads to a different BO trajectory, bringing RMSE to zero. The shaded area to the right of the vertical
dotted line spans the trajectories from the minimum to maximum values of RMSE. Figure 1 shows that all BO trajectories, regardless of the initial conditions, converge to zero RMSE, and that the convergence is monotonous. The results in the upper panel of Figure 1 were obtained using the scalar approach (2) and the kernel (10). The subsequent discussion explores how BO can be accelerated by employing feedback loop (3) and modifying the kernel complexity, leading to the results in the lower panel of Figure 1.

The numerical results of BO are summarized in Tables I - III. The variable space was set to span \(\alpha_{ab} \in [-30, +30]\) and \(\alpha_{ac} \in [-30, +30]\) for SO\(_2\); and \(\alpha_{ab} \in [-40, 40]\), \(\alpha_{ac} \in [-40, 40]\), \(\alpha_{ab} \in [0, 5]\), \(\alpha_{ac} \in [0, 5]\) and \(\alpha_{bc} \in [0, 5]\) for PPO, all in atomic units. The observation of these results leads to the following conclusions:

- The vector approach (3) converges much faster than the scalar approach (2). Note that the kernels of the GP models in these calculations are represented by the same function (10) so the acceleration of the BO convergence is due to the difference in the algorithms.
- Both approaches can identify the five parameters of the polarizability tensor for PPO with accuracy better than 1%.
- The convergence of the calculations based on the alignment and orientation signals is different.

Figure 2 elucidates the last observation and illustrates the relative convergence of the scalar and vector algorithms. It is clear from the figure that the vector approach reduces the number of BO iterations required at least by a factor of 2. The alignment reference signal appears to constrain the polarizability tensor parameters more effectively, resulting in a further reduction of BO iterations required for full convergence. It is important to note that, in practice, it is not necessary to repeat BO with different initial conditions. However, it may be advisable to perform calculations with, at least two, drastically different initial conditions to quantify convergence. These calculations are completely independent and can be performed in parallel. If quantum dynamics calculations are extremely time consuming, the best approach may be to begin multiple BO calculations with drastically different initial conditions and harvest the results from the BO trajectory approaching the target most steeply.
|          | $\alpha_{ab}$ | $\alpha_{ac}$ |
|----------|--------------|--------------|
| Reference| 10.46        | 12.62        |
| 10 ps (scalar) | 10.47 | 12.47 |
| 100 ps (scalar) | 10.52 | 12.50 |
| 10 ps (vector) | 10.45 | 12.60 |

TABLE I. The polarizability parameters (in a.u.) of SO$_2$ determined by BO with 50 iterations (scalar) and 30 iterations (vector) using the time dependence of birefringence over the indicated time interval. The results shown are averaged over 20 instances of initial conditions for BO. Each initial condition is based on 50 quantum calculations with $\alpha_{ab}$ and $\alpha_{ac}$ chosen randomly using Latin hypercube sampling.

|          | $\alpha_{ab}$ | $\alpha_{ac}$ | $\alpha_{bc}$ |
|----------|--------------|--------------|--------------|
| Reference| 7.67         | 7.76         | 2.56         | 0.85         | 0.65         |

TABLE II. The polarizability parameters (in a.u.) of PPO determined by BO with 100 iterations (scalar) and 50 iterations (vector) using the alignment (for alignment) and orientation (for orientation) factors over the indicated time interval. The results shown are averaged over 20 instances of initial conditions for BO. Each initial condition is based on 50 quantum calculations with the polarizability parameters chosen randomly using Latin hypercube sampling.
| $N_i$ | PPO | $\alpha^{ab}$ | $\alpha^{ac}$ | $\alpha_{ab}$ | $\alpha_{ac}$ | $\alpha_{bc}$ |
|------|------|----------------|----------------|--------------|--------------|--------------|
| Reference | 7.67 | 7.76 | 2.56 | 0.85 | 0.65 |

### Alignment

|  | Average | $\bar{\alpha}^{ab}$ | $\bar{\alpha}^{ac}$ | $\bar{\alpha}_{ab}$ | $\bar{\alpha}_{ac}$ | $\bar{\alpha}_{bc}$ |
|---|--------|---------------------|---------------------|---------------------|---------------------|---------------------|
| 50 | 7.669  | 7.761               | 2.560               | 0.850               | 0.649               |
| 50 | 0.008  | 0.007               | 0.0002              | 0.0006              | 0.002               |
| 30 | 7.671  | 7.762               | 2.560               | 0.851               | 0.614               |
| 30 | 0.145  | 0.119               | 0.002               | 0.006               | 0.150               |
| 10 | 17.13  | 16.05               | 3.65                | 2.39                | 1.43                |
| 10 | 11.81  | 10.10               | 1.20                | 1.93                | 1.28                |

### Orientation

|  | Average | $\bar{\alpha}^{ab}$ | $\bar{\alpha}^{ac}$ | $\bar{\alpha}_{ab}$ | $\bar{\alpha}_{ac}$ | $\bar{\alpha}_{bc}$ |
|---|--------|---------------------|---------------------|---------------------|---------------------|---------------------|
| 50 | 7.666  | 7.760               | 2.560               | 0.850               | 0.649               |
| 50 | 0.003  | 0.004               | 0.002               | 0.005               | 0.0009              |
| 30 | 7.670  | 7.760               | 2.560               | 0.851               | 0.614               |
| 30 | 0.004  | 0.003               | 0.002               | 0.006               | 0.005               |
| 10 | 7.938  | 4.125               | 2.389               | 1.574               | 1.402               |
| 10 | 17.62  | 10.09               | 1.25                | 1.45                | 1.59                |

TABLE III. The polarizability parameters (in a.u.) of PPO determined by BO after 35 iterations using the same reference signal as in Table II. $N_i$ represents the number of quantum calculations before BO is initiated (vertical dotted line in Figure 1). The averages and standard deviations (SD) are computed using 20 initial conditions.
FIG. 2. Convergence of BO iterations for the 5-dimensional case of PPO: dot-dashed curve (orange) – scalar BO based on the orientation signal; dashed curve (red) – vector BO based on the orientation signal; solid curve (blue) – vector BO based on the alignment signal. The shaded areas span the range from the minimum to maximum values of RMSE in the set of 20 calculations with different initial conditions. BO is initialized by 50 quantum calculations as in Figure 1. The RMSE is dimensionless for the alignment and in units of Debye for the orientation.
FIG. 3. Time dependence of the birefringence of a gas of SO$_2$ at three different temperatures $(T = 15, 100$ and $300$ K): solid curves – calculation with the reference values of the polarizability tensor parameters; symbols – calculations with the polarizability tensor parameters determined by BO (c.f., Table IV).

Figure 3 presents the results of BO using the reference curves computed at finite temperatures. The quantum dynamics calculations are performed using the RPWF approach with $N = 40$, as described in Section II.A. The polarizability tensor parameters determined from the reference signals shown in Figure 3 are given in Table IV. To illustrate the stability of the predictions, we repeated the calculations with 20 initial conditions, leading to the spreads of the polarizability tensor parameters characterized by the standard deviations listed in Table IV. The results illustrate that the averaging over quantum states required for the finite temperature simulations does not spoil the accuracy of BO and does not affect the convergence of BO.

We would like to briefly discuss several complications that may arise in recovering the polarizabilities from experimental signals. Most significantly, in most experiments, the pulse parameters are not known with sufficient precision. In such cases, the polarizabilities cannot
TABLE IV. The polarizability parameters (in a.u.) of SO$_2$ determined by BO (vector approach) with 20 iterations using the birefringence signal over 100 ps at three different temperatures. The results are averaged over 20 instances of initial conditions for BO, leading to the distributions of values with the standard deviations (SD) listed. Each initial condition is based on 20 quantum calculations with the polarizability parameters chosen randomly using Latin hypercube sampling.

| Temperature | $\alpha^{ba}$ | $\alpha^{bc}$ |
|-------------|---------------|---------------|
| Reference   | -10.46        | 2.16          |
| 15 K        | -10.46        | 2.16          |
|             | SD            | 0.092         |
|             |               | 0.085         |
| 100 K       | Average -10.47| 2.15          |
|             | SD            | 0.038         |
|             |               | 0.087         |
| 300 K       | Average -10.47| 2.15          |
|             | SD            | 0.041         |
|             |               | 0.073         |

be directly recovered, because the Hamiltonian involves products of the pulse intensity and polarizability (see Eq. 4). Moreover, experimental data (e.g. birefringence signal) are often systematically scaled and shifted. We suggest that an auxiliary experiment may, in principle, allow one to resolve these issues. In this experiment, the measurement is done on simple molecules with precisely known polarizability. Using the methods described here, the pulse parameters, signal scaling, and signal shift can be optimized by including them in the $\mathbf{x}$ vector, while keeping the polarizability tensor elements fixed to the accurate values. Next, the main experiment with the complex molecule of interest is carried out under the same conditions. At this point, the procedure continues as described here, i.e. BO is performed with the $\mathbf{x}$ vector including the polarizabilities of the complex molecule, while the pulse parameters, and signal scaling/shift are kept fixed at the values found in the auxiliary experiment.

Depending on the experimental conditions, physical models can be extended to better describe both the auxiliary and the main experiments. For example, the models can include averaging over the spatial intensity profile of the laser beam, molecular centrifugal distortion, hyperfine coupling effects (see, e.g., [53]), etc. As in the case of polarizabilities, the additional parameters can be added to the $\mathbf{x}$ vector.
A. Effect of noise in reference signal

Experimental data generally come with noise. GPs are particularly well suited to describe noisy data, as Gaussian noise is naturally built into the construction of the likelihood function \[23, 44\]. The variance of the noise \( \sigma \) can be treated as a variable parameter when training GPs. BO based on GPs can thus readily account for the experimental noise in the reference signal. To demonstrate this, we modulate the reference birefringence curve to include Gaussian noise in the time domain. Specifically, we generate a random distribution

\[ f_{\text{ref+noise}}(t) = f_{\text{ref}}(t) + \mathcal{N}(\mathcal{M}, \gamma \times \Sigma) , \]

where \( \mathcal{M} \) and \( \Sigma \) are the mean and the variance of the 400 points in \( f_{\text{ref}}(t) \), \( \mathcal{N} \) denotes a normal distribution and \( \gamma \) is a variable percentage factor. For the signal with 10 % noise, \( \gamma = 0.1 \). The values \( f_{\text{ref+noise}}(t) \) thus generated are then used as the reference signal for BO.

Table V summarizes the results for PPO based on the reference signal (the orientation factor) with different amounts of noise. As follows from Table V, 5 % noise allows the determination of the diagonal elements of the polarizability tensor to within 3.4 % and the small off-diagonal elements to within 7.5 %.

Figure 4 shows the results for PPO using the degree of alignment curve with 10 % noise as the reference signal. The figure illustrates that the signal calculated with the optimized polarizability parameters is very close to the mean of the reference signal. The diagonal elements of the polarizability tensor as well as \( \alpha_{ac} \) are constrained to better than 7 %, while the off-diagonal element \( \alpha_{bc} \) deviates from the reference by a factor of 2. This suggests that the degree of alignment is much less sensitive to \( \alpha_{bc} \) than the other polarizability tensor parameters. We note that typical experimental data have much lower noise levels.

B. Effect of kernel complexity

To build a GP model, one assumes an analytical form for the kernel function \( k(x, x') \) with some unknown parameters. The mathematical form of the kernel defines the GP model. The choice of the kernel function is not unique. As covariances are expected to decrease with the distance in the input space, one typically assumes a kernel function that decays with \( |x - x'| \). For example, in addition to the function \[10\], the following functions are often used for GP models \[45-47\]:
TABLE V. The polarizability parameters (in a.u.) of PPO determined by BO (vector approach) with 50 iterations using the orientation factor over 100 ps of the reference signal including Gaussian noise with magnitude 1%, 3%, and 5%. The results shown are averaged over 20 instances of initial conditions for BO. The standard deviation (SD) is in percentage with respect to the average values. Each initial condition is based on 50 quantum calculations with the polarizability parameters chosen randomly using Latin hypercube sampling.

| Orientation | \( \alpha_{ab} \) | \( \alpha_{ac} \) | \( \alpha_{ab} \) | \( \alpha_{ac} \) | \( \alpha_{bc} \) |
|-------------|----------------|----------------|----------------|----------------|----------------|
| Reference   | 7.67           | 7.76           | 2.56           | 0.85           | 0.65           |
| 1% noise    | 7.637          | 7.799          | 2.577          | 0.862          | 0.650          |
| SD/%        | 2.796          | 1.576          | 3.103          | 4.001          | 4.221          |
| 3% noise    | 7.656          | 7.750          | 2.580          | 0.844          | 0.662          |
| SD/%        | 2.267          | 1.765          | 4.212          | 4.903          | 6.475          |
| 5% noise    | 7.678          | 7.790          | 2.592          | 0.857          | 0.656          |
| SD/%        | 3.380          | 3.244          | 5.112          | 4.426          | 7.492          |

where \( r \) is the same as defined after Eq. (10), \( \alpha \) and \( l \) are free parameters, and ‘RBF’ and ‘RQ’ are the abbreviations for ‘radial basis function’ and ‘rational quadratic’, respectively.

GP models with any kernel provide accurate interpolation in the limit of infinite training data and, consequently, BO with any GP model must converge to the target in the limit of an infinite number of iterations. Therefore, BO is typically performed with GP models trained using one of the kernel functions (10) or (22) - (24). However, the choice of the kernel function may affect the convergence of BO. Therefore, if the iterations are expensive,
FIG. 4. Comparison of the degree of alignment (green dashed curve) computed using the optimized polarizability tensor parameters with the reference signal represented by the red solid curve (mean) and with 10% noise (blue symbols). The optimized polarizability tensor parameters are $7.234 (-0.436), 8.028 (+0.268), 2.665 (+0.105), 0.985 (+0.135), 0.329 (-0.321)$. The values in parentheses are the deviations from the reference values (c.f., Table V).

as is the case for quantum dynamics applications, the choice of the kernel function may be critically important.

The predictive power of GP models with the same number of parameters can be quantified by the magnitude of the log-likelihood function [11]. However, in order to compare two GP models with kernel functions of different complexity, it is more suitable to use the Bayesian information criterion [54] defined as

$$\text{BIC} = \log \mathcal{L} - \frac{1}{2} \mathcal{M} \log n,$$

(25)

where $\mathcal{M}$ is the number of parameters in the kernel function and $n$ is the number of training points. The second term in Eq. (25) penalizes kernels with more parameters.

As demonstrated previously [26, 27], the magnitude of the BIC can be used as a model selection criterion to enhance the prediction power of GP models. The algorithm introduced
builds composite kernels from the simple kernels (10) and (22) - (24) as follows. The kernel selection approach begins by training separately four GP models with each of the kernel functions (10) and (22) - (24), denoted hereafter as \( k_i \). The BIC (25) is computed for each of the GP models with different kernel \( k_i \) and the kernel function of the model with the largest BIC is selected as the preferred kernel \( k_0 = k_i \). The kernel \( k_0 \) is then combined with each of the four kernels (10) and (22) - (24) by forming linear combinations \( c_i k_0 + c_j k_j \) and products \( c_j k_0 \times k_j \), thus leading to eight new functions. Eight new GP models are trained with each of these kernel functions by optimizing both the kernel parameters and the coefficients \( c_i \) and \( c_j \). The kernel of the model with the largest BIC is selected as the new preferred kernel \( k_0 \) and the procedure is iterated.

This process increases the complexity of the kernel function, making GP models more accurate. Here, we use this algorithm to increase the complexity of the GP models of the initial distribution of quantum dynamics calculations before BO, that is the model of the calculations to the left of the vertical broken line in Figure 1. The kernel functions selected by the largest magnitude of the BIC are then fixed and used throughout the subsequent BO iterations. Note that this kernel selection process does not require additional quantum dynamics calculations.

Figure 5 illustrates the effect of the kernel complexity on the convergence of BO (vector model) for the case of PPO with five variable polarizability tensor parameters. As can be seen, BO with the kernel functions corresponding to larger values of the BIC requires fewer iterations to converge. We have verified the generality of this result by repeating the calculations with different initial conditions. It is important to note that the numerical effort to train GP models for inverse problems in quantum dynamics is a negligibly small fraction of the computation time required for solving the dynamics problem. Therefore, the algorithm described here substantially decreases the total computation time by decreasing the number of BO iterations required for convergence.

One can envision an extension of the present algorithm that re-optimizes the kernel complexity at each iteration of BO. Because the iterations are bottlenecked by the quantum dynamics calculations, this does not increase the computation time for inverse quantum problems. We have tested this algorithm and found that optimizing kernels to increase the BIC of the GP models at each iteration does not affect significantly the total number
of iterations required for convergence for the problems considered here. However, other problems may benefit from adjusting the kernels to yield Bayesian models with the largest BIC at each iteration.

IV. CONCLUSION

We have demonstrated and compared three algorithms for inverse problems in time-dependent quantum dynamics based on Bayesian optimization of feedback loops. The first algorithm is the straightforward application of BO designed to minimize a scalar function embodying the difference between predictions of the Schrödinger equation and a reference observable. This algorithm can be based on any Bayesian ML model and use one of several scalar functions to quantify the departure of quantum predictions from reference data, including the RMSE \( (12) \) or Eq. \( (13) \). However, the convergence of BO depends on the underlying ML model. Because each iteration of feedback loops requires numerical integration of the nuclear Schrödinger equation, which is time-consuming, it is critically important to develop methods converging feedback loops with as few iterations as possible.

We have illustrated that the convergence of feedback loops is significantly accelerated when the time variable is discretized and the ML model used for BO is trained to reproduce the target observable at different time instances simultaneously. This algorithm relies on Gaussian processes with multiple outputs and takes advantage of the model prediction and prediction uncertainties at different time instances. In the present work, the GP models exploited correlations between inputs into the models, but not between outputs. This algorithm can thus be further improved by also accounting for correlations between the outputs of a multi-output GP. However, this would make the acquisition function used for BO more complex than Eq. \( (15) \).

Finally, we have illustrated that the convergence of feedback loops is further accelerated by increasing the kernel complexity of the GP models used for BO. We proposed an approach that builds up the complexity of the GP models using a few training points before BO is initialized. The kernel complexity is increased using a greedy search algorithm with the Bayesian information criterion for the model selection, as was previously done to enhance pattern-recognition accuracy of GP models \([26, 28]\). BO is then carried out with the complex kernels for the GP models thus obtained. We have shown that a combination of the vector
approach based on multi-output GPs and the method using GP models with complex kernels reduces the number of BO iterations by more than a factor of 3 compared to BO based on scalar GP models with simple kernels.

To illustrate these algorithms, we considered two-dimensional and five-dimensional inverse problems with several different observables. We have shown that the convergence of BO is different for different observables, while the final results are the same. We have also illustrated that BO is readily suited for solving inverse problems using observables with Gaussian noise. We note that the methods demonstrated in this work do not use explicitly any information on the evolution of the Schrödinger equation solutions with the Hamiltonian parameters, yet the five-dimensional Hamiltonian parameter space for problems considered here can be explored with as few as 30 quantum calculations.

It is important to emphasize that these conclusions are based on the empirical evidence obtained here for specific problems. However, the main qualitative conclusions regarding the acceleration of BO are expected to be general. To illustrate this, we note that the convergence of BO depends on the accuracy of the underlying GPs. Using the vector approach as described in this work effectively enhances inference at each iteration of BO by making the acquisition function more efficient compared to those based on a single GP. This is consistent with the conclusions in [48]. Similarly, using composite kernels for GPs as described in this work, enhances the predictive power of GP for each iteration of BO. We thus conclude that formulating inverse problems in quantum dynamics as a vector estimation problem for BO with composite kernels must generally reduce the total number of quantum calculations required to explore the Hamiltonian parameter space.

It is instructive to consider the extension of the present approach to problems with more dimensions. While specific applications may be affected by particular details, it is possible to make a few general observations. The extension of the present approach to high-dimensional problems is limited by two factors: the numerical difficulty of the quantum dynamics calculations and the numerical complexity of building GPs for BO. Training a GP by n training points without approximations involves inverting an $n \times n$ matrix, which scales with $n$ as $O(n^3)$. The number of training points $n$ for accurate GP regression of a $D$-dimensional problem can be estimated to be between $10 \times D$ and $20 \times D$ [55]. As demonstrated in the present work, the kernel optimization algorithm used in conjunction with BO reduces this scaling to below the lower end of this estimate. The numerical difficulty of training GP models should,
therefore, become a significant factor only for problems with \( \gtrsim 20 - 30 \) unknown parameters. Kernel optimization becomes more difficult for higher-dimensional problems. However, we have recently implemented the kernel optimization algorithm used here for a 51-dimensional problem \([56]\). For high-dimensional problems, training of GP models can also be accelerated by well-controlled approximations, such as data sparsification \([57, 60]\). The extension of the present approach to problems with up to 30 parameters is thus expected to be limited by the numerical difficulty of the quantum dynamics calculations. Given our present results and the scaling analysis \([55]\), we estimate that a 30-dimensional problem will require between 180 and 600 quantum dynamics calculations, with kernel optimization reducing the problem to the lower end of the estimate.

We also would like to point out that the present approach is not restricted to a particular quantum dynamics simulation method and can be used for open as well as closed systems. Quantum calculations for complex systems must rely on approximations. If quantum dynamics approximations lead to unknown uncertainties, the error of the quantum results will be absorbed into the Hamiltonian parameters determined by BO. This is clearly undesirable. To overcome this problem, one may use Bayesian model calibration \([61]\) to correct the results of quantum dynamics calculations either by isolated rigorous quantum results \([22]\) or by experimental data. As shown in \([62]\), this can be achieved by using GP models with multiple outputs representing approximate and rigorous results. Such GP models are trained to learn correlations both between inputs and between outputs. An output of such GP used for BO as in the present work should yield Hamiltonian parameters accounting for the error of the quantum dynamics approximation.
FIG. 5. Effect of kernel complexity on BO. Upper panel: BO with GPs using the linear kernel \cite{22}; the value of BIC for the initial model with this kernel is 16142; Lower panel (in the order of increasing BIC): BO with GPs using the Matérn kernel (blue dotted curve, BIC = 61269), RBF kernel (green dot-dashed curve, BIC = 65459), rational quadratic kernel (orange dashed, BIC = 75453), a linear combination of RQ, Matérn and RBF (purple spaced dashed curve, BIC = 75465), and a linear combination of RQ and Matérn (red solid curve, BIC = 75470).
DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ACKNOWLEDGMENTS

The authors appreciate valuable insights from Valery Milner, Sharly Fleischer and Dina Rosenberg on experimental aspects of measuring the laser-induced alignment and orientation of SO$_2$ and PPO molecules. This work was supported by NSERC of Canada and by Israel Science Foundation (Grant No. 746/15). I. A. acknowledges the support as the Patricia Elman Bildner Professorial Chair. This research was made possible in part by the historic generosity of the Harold Perlman Family.
[1] B. N. Zakhariev and V. M. Chabanov, New situation in quantum mechanics (wonderful potentials from the inverse problem), *Inverse Problems* **13**, R47 (1997).

[2] I. Levine, *Molecular spectroscopy*, Wiley, New York, 1975.

[3] J. Kraitchman, Determination of molecular structure from microwave spectroscopic data, *Am. J. Phys.* **21**, 17 (1953).

[4] F. E. Marshall, G. Sedo, C. West, B. H. Pate, S. M. Allpress, C. J. Evans, P. D. Godfrey, D. McNaughton, and G.S. Grubbs, The rotational spectrum and complete heavy atom structure of the chiral molecule verbenone, *J. Mol. Spectrosc.* **342**, 109 (2017).

[5] T. Kierspel, A. Morgan, J. Wiese, T. Mullins, A. Aquila, A. Barty, R. Bean, R. Boll, S. Boutet, P. Bucksbaum, H. N. Chapman, L. Christensen, A. Fry, M. Hunter, J. E. Koglin, M. Liang, V. Mariani, A. Natan, J. Robinson, D. Rolles, A. Rudenko, K. Schnorr, H. Stapelfeldt, S. Stern, J. Thøgersen, C. H. Yoon, F. Wang, and J. Küpper, X-ray diffractive imaging of controlled gas-phase molecules: Toward imaging of dynamics in the molecular frame, *J. Chem. Phys.* **152**, 084307 (2020).

[6] M. Pitzer, M. Kunitski, A. S. Johnson, T. Jahnke, H. Sann, F. Sturm, L. Ph. H. Schmidt, H. Schmidt-Böcking, R. Dörner, J. Stohner, J. Kiedrowski, M. Reggelin, S. Marquardt, A. Schießer, R. Berger, and M. S. Schöffler, Direct determination of absolute molecular stereochemistry in gas phase by coulomb explosion imaging, *Science* **341**, 1096 (2013).

[7] E. T. Karamatskos, S. Raabe, T. Mullins, A. Trabattoni, P. Stammer, G. Goldsztejn, R. R. Johansen, K. Dlugolecki, H. Stapelfeldt, M. J. J. Vrakking, S. Trippel, A. Rouzée, and J. Küpper, Molecular movie of ultrafast coherent rotational dynamics of OCS, *Nat. Comm.* **2041**, 3364 (2019).

[8] R. J. LeRoy and R. B. Bernstein, Dissociation Energy and Long-Range Potential of Diatomic Molecules from Vibrational Spacings of Higher Levels, *J. Chem. Phys.* **52**, 3869 (1970).

[9] A. Pashov, W. Jastrzebski, and P. Kowalczyk, Construction of potential curves for diatomic molecular states by the IPA method, *Comp. Phys. Comm.* **128**, 622 (2000).

[10] S. M. Kirschner and J. K.G. Watson, RKR potentials and semiclassical centrifugal constants of diatomic molecules, *J. Mol. Spec.* **47**, 234 (1973).
[11] R. Vargas-Hernandez, Y. Guan, D. H. Zhang, and R. V. Krems, Bayesian optimization for the inverse scattering problem in quantum reaction dynamics, *New J. Phys.* (Fast Track Communication) **21**, 022001 (2019).

[12] F. Caillieza and P. Pernot, Statistical approaches to forcefield calibration and prediction uncertainty in molecular simulation, *J. Chem. Phys.* **134**, 054124 (2011).

[13] G. Reinisch, K. Miki, G. L. Vignoles, B. M. Wong, and C. S. Simmons, An Efficient and Accurate Formalism for the Treatment of Large Amplitude Intramolecular Motion, *J. Chem. Theor. Comp.* **8**, 2713 (2012).

[14] P. Angelikopoulos, C. Papadimitriou, and P. Koumoutsakos, Bayesian uncertainty quantification and propagation in molecular dynamics simulations: A high performance computing framework, *J. Chem. Phys.* **137**, 144103 (2012).

[15] F. Rizzi, H. N. Najm, B. J. Debusschere, K. Sargsyan, M. Salloum, H. Adalsteinsson, and O. M. Knio, Uncertainty Quantification in MD Simulations. Part II: Bayesian Inference of Force-Field Parameters, *Multiscale Model. Simul.* **10**, 1460 (2012).

[16] F. Cailliez, A. Bourasseau, and P. Pernot, Calibration of forcefields for molecular simulation: Sequential design of computer experiments for building cost-efficient kriging metamodels, *J. Comp. Chem.* **35**, 130 (2014).

[17] P. Pernot, The parameter uncertainty inflation fallacy, *J. Chem. Phys.* **147**, 104102 (2017).

[18] S. Wu, P. Angelikopoulos, G. Tauriello, C. Papadimitriou, and P. Koumoutsakos, Fusing heterogeneous data for the calibration of molecular dynamics force fields using hierarchical Bayesian models, *J. Chem. Phys.* **145**, 244112 (2016).

[19] P. L. A. Popelier, QCTFF: On the construction of a novel protein force field, *Int. J. Quant. Chem.* **115**, 1005 (2015).

[20] V. Botu and R. Ramprasad, Adaptive machine learning framework to accelerate ab initio molecular dynamics, *Int. J. Quant. Chem.* **115**, 1074 (2015).

[21] M. Caccin, Z. Li, J. R. Kermode, and A. De Vita, A framework for machine-learning-augmented multiscale atomistic simulations on parallel supercomputers, *Int. J. Quant. Chem.* **115**, 1129 (2015).

[22] J. Cui and R. V. Krems, Gaussian Process Model for Collision Dynamics of Complex Molecules, *Phys. Rev. Lett.* **115**, 073202 (2015).
[23] R. V. Krems, Bayesian Machine Learning for Quantum Molecular Dynamics, *Phys. Chem. Chem. Phys.* **21**, 13392 (2019).

[24] J. Snoek, H. Larochelle, and R. P. Adams, Practical Bayesian optimization of machine learning algorithms, *Adv. Neur. Inf. Process. Sys.* **25**, 2951(2012).

[25] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, Taking the human out of the loop: A review of Bayesian optimization, *Proc. IEEE* **104**, 148 (2016).

[26] D. K. Duvenaud, H. Nickisch, and C. E. Rasmussen, Additive Gaussian Processes, *Adv. Neur. Inf. Proc. Sys.* **24**, 226 (2011).

[27] D. K. Duvenaud, J. Lloyd, R. Grosse, J. B. Tenenbaum, and Z. Ghahramani, Structure Discovery in Nonparametric Regression through Compositional Kernel Search, *Proceedings of the 30th International Conference on Machine Learning Research* **28**, 1166 (2013).

[28] R. Vargas-Hernandez, J. Sous, M. Berciu, and R. V. Krems, Extrapolating quantum observables with machine learning: Inferring multiple phase transitions from properties of a single phase, *Phys. Rev. Lett.* **121**, 255702 (2018).

[29] J. Dai and R. V. Krems, Interpolation and extrapolation of global potential energy surfaces for polyatomic systems by Gaussian processes with composite kernels, *J. Chem. Theory Comput.* **16**, 1386 (2020).

[30] K. D. Bonin and V. V. Kresin, Electric-Dipole Polarizabilities of Atoms, Molecules, and Clusters (World Scientific, 1997).

[31] R. N. Zare, Angular Momentum: Understanding Spatial Aspects in Chemistry and Physics (Wiley-Interscience, 1991).

[32] H. Goldstein, Classical Mechanics (3rd Edition) (Pearson, 2001).

[33] M. Lemeshko, R. V. Krems, J. M. Doyle, and S. Kais, Manipulation of molecules with electromagnetic fields, *Mol. Phys.* **111**, 1648 (2013).

[34] R. Krems, Molecules in Electromagnetic Fields: from ultracold physics to controlled chemistry, John Wiley & Sons (2018).

[35] S. Kallush and S. Fleischer, Orientation dynamics of asymmetric rotors using random phase wave functions, *Phys. Rev. A* **91**, 063420 (2015).

[36] R. Damari, S. Kallush, and S. Fleischer, Rotational Control of Asymmetric Molecules: Dipole-versus Polarizability-Driven Rotational Dynamics, *Phys. Rev. Lett.* **117**, 103001 (2016).
[37] H. Stapelfeldt and T. Seideman, Colloquium: Aligning molecules with strong laser pulses, *Rev. Mod. Phys.* **75**, 543 (2003).

[38] C. P. Koch, M. Lemeshko, and D. Sugny, Quantum control of molecular rotation, *Rev. Mod. Phys.* **91**, 035005 (2019).

[39] O. Faucher, B. Lavorel, E. Hertz, and F. Chaussard, in Progress in Ultrafast Intense Laser Science VII (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011) pp. 79-108.

[40] J. Bert, E. Prost, I. Tutunnikov, P. Béjot, E. Hertz, F. Billard, B. Lavorel, U. Steinitz, I. Sh. Averbukh, and O. Faucher, Optical imaging of coherent molecular rotors, *Las. Phot. Rev.* **14**, 1900344 (2020).

[41] K. Lin, I. Tutunnikov, J. Ma, J. Qiang, L. Zhou, O. Faucher, Y. Prior, I. Sh. Averbukh, and J. Wu, Spatiotemporal rotational dynamics of laser-driven molecules, *Adv. Phot.* **2**, 1 (2020).

[42] P. Babilotte, K. Hamraoui, F. Billard, E. Hertz, B. Lavorel, O. Faucher, and D. Sugny, Observation of the field-free orientation of a symmetric-top molecule by terahertz laser pulses at high temperature, *Phys. Rev. A* **94**, 043403 (2016).

[43] R. M. Neal, Bayesian methods for Machine Learning, NIPS Tutorial (2004): https://www.cs.toronto.edu/~radford/ftp/bayes-tut.pdf

[44] Rasmussen, C. E.; Williams, C. K. I. *Gaussian Processes for Machine Learning*, The MIT Press: Cambridge (2006).

[45] T. Mitchell, M. Morris, and D. Ylvisaker, Existence of Smoothed Stationary Processes on an Interval, *Stoch. Proc. Appl.* **35**, 109 (1990).

[46] N. A. C. Cressie, *Statistics for Spatial Data*, Wiley-Interscience, New York (1993).

[47] M. L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*, Springer Science & Business Media (1999).

[48] A. K. Uhrenholt, B. S. Jensen, Efficient Bayesian Optimization for Target Vector Estimation, *Proc. Mach. Learn. Res.* **89**, 2661 (2019).

[49] M. Sankaran, On the non-central chi-square distribution, *Biometrika* **46**, 235 (1959).

[50] I. Tutunnikov, J. Floss, E. Gershnabel, P. Brumer, and I. Sh. Averbukh, Laser induced persistent orientation of chiral molecules, *Phys. Rev. A* **100**, 043406 (2019).

[51] M. Stein, Large Sample Properties of Simulations Using Latin Hypercube Sampling, *Technometrics* **29**, 143 (1987).
[52] J. Cui and R. V. Krems, Efficient non-parametric fitting of potential energy surfaces for polyatomic molecules with Gaussian processes, *J. Phys. B* **49**, 224001 (2016).

[53] E. F. Thomas, A. A. Søndergaard, B. Shepperson, N. E. Henriksen, and H. Stapelfeldt, Hyperfine-Structure-Induced Depolarization of Impulsively Aligned I₂ Molecules, *Phys. Rev. Lett.* **120**, 163202 (2018).

[54] G. Schwarz, Estimating the dimension of a model, *Annals Stat.* **6**(2), 461 (1978).

[55] J. L. Loeppky, J. Sacks, and W. J. Welch, Choosing the Sample Size of a Computer Experiment: A Practical Guide, *Technometrics* **51**, 366 (2009).

[56] H. Sugisawa, T. Ida, R. V. Krems, Gaussian process model of 51-dimensional potential energy surface for the protonated imidazole dimer. [arXiv:2001.07271](https://arxiv.org/abs/2001.07271).

[57] Y. Cao, M. A. Brubaker, D. J. Fleet, A. Hertzmann, Efficient Optimization for Sparse Gaussian Process Regression, *IEEE Trans. Patt. Anal. Mach. Intell.* **37**, 2415 (2015).

[58] J. Q. Quinonero-Candela and C. E. Rasmussen, A unifying view of sparse approximate Gaussian process regression, *J. Mach. Learn. Res.* **6**, 1939 (2005).

[59] E. Snelson and Z. Ghahramani, in Advances in Neural Information Processing Systems 18, edited by Y. Weiss, B. Schölkopf, and J. Platt (MIT Press, 2006), pp. 1257-1264

[60] J. Schreiter, D. Nguyen-Tuong, and M. Toussaint, Efficient sparsification for Gaussian process regression, *Neurocomputing* **192**, 29 (2016).

[61] M. C. Kennedy and A. O’Hagan, Bayesian calibration of computer models, *J. Royal Stat. Soc. B* **63**, 425 (2001).

[62] A. Jasinski, J. Montaner, R. C. Forrey, B. H. Yang, P. C. Stancil, N. Balakrishnan, J. Dai, R. A. Vargas-Hernández, and R. V. Krems, Machine-learning-corrected quantum dynamics calculations, *Phys. Rev. Research* (Rapid communication), in press (2020); [arXiv:2001.06592](https://arxiv.org/abs/2001.06592).