Nonparametric Bayesian approach to extrapolation problems in configuration interaction methods

Sota Yoshida*

Department of Physics, the University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
and Liberal and General Education Center, Institute for Promotion of Higher Academic Education, Utsunomiya University, Mine, Utsunomiya, 321-8505, Japan

(Received 12 July 2019; revised 13 March 2020; accepted 24 June 2020; published 5 August 2020)

The configuration interaction methods are powerful tools for exploring various properties of nuclei. However, in practice, it is often necessary to truncate the model space and then to extrapolate the results to very large model space to obtain the best estimations of the exact eigenvalues under a given nuclear interaction. In this study, a nonparametric extrapolation method based on constrained Gaussian processes for configuration interaction methods is presented. The proposed method has many advantages: (i) applicability to small data sets such as results of *ab initio* methods, (ii) flexibility to incorporate constraints, which are guided by physics, into the extrapolation model, (iii) providing predictions with quantified extrapolation uncertainty, etc. An application to the extrapolation problem of ground-state energies in the full configuration interaction method is discussed as an example.

DOI: 10.1103/PhysRevC.102.024305

I. INTRODUCTION

Recent developments in nuclear potentials based on chiral effective field theory (chiral EFT) [1,2] and *ab initio* methods [3–7] have provided deep insights into nuclear many-body problems starting from the fundamental interaction between nucleons, see very recent works [8–11] and references therein.

The full configuration interaction (FCI) method, which is also known as no-core full configuration (NCFC)/no-core shell model (NCSM) [12,13], is one of the successful *ab initio* methods. In FCI, wave functions are represented in a truncated subspace, and the truncation is typically specified by the parameter $N_{\max}$, which defines the maximum number of harmonic oscillator quanta allowed in the many-body states above the lowest configuration for a target nucleus.

Despite enormous efforts for developing efficient codes [14–18] and advances in computing power, the currently available $N_{\max}$ for upper p-shell nuclei is around 10 (see, e.g., Ref. [19]) and this is still far from $N_{\max} = \infty$ corresponding to exact calculations. One usually extrapolates the sequence of results with different $N_{\max}$ to $N_{\max} = \infty$ to estimate the exact value. In previous studies, several extrapolation methods were proposed and the dependence on them was analyzed [18–24]. The most intuitive example is one based on an exponential function [19].

In addition to FCI, such extrapolation techniques are also required in CI calculations for a valence space, which is the so-called shell model, using additional truncations. Representative examples of the truncations are importance truncation scheme [25,26] and Monte Carlo shell model [27,28]. In those calculations, the rapid growth of many-body basis for a valence space is alleviated by selecting a small subset of the many-body basis states, which is physically more relevant. These truncation schemes have been successfully applied to valence CI and also FCI calculations in previous works such as Refs. [29–34].

In these studies, the extrapolation is performed with some specific functions such as an exponential or polynomials, and the coefficients are determined so as to minimize the χ^2 deviation from the given calculated data points. While any of these offers intuitively reasonable extrapolated results, there is a risk of overfitting, i.e., lack of predictive power for true exact values. This overfitting is because χ^2 minimization of a parametric model and point estimation of the parameters leads to too deterministic predictions due to the limited expression power of the model. This is a problematic situation if one intends to discuss quantitative issues such as a level ordering of states with small energy differences, the positions of proton and neutron drip lines, and so on.

In the present study, a novel nonparametric extrapolation method for CI calculations using constrained Gaussian processes (GPs) is proposed. The method gives extrapolated results with quantified uncertainty in a systematic manner.

Evaluating extrapolation uncertainties are helpful for breaking down possible origins of discrepancy between FCI results and experimental data, though the major source of uncertainty is, at the moment, from the input potentials: the low-energy constants and the truncated expansion in chiral EFT. Although GPs are also not free from overfitting, GPs allow the consideration of a wider class of functions, and then it is expected to alleviate underestimating extrapolation uncertainties due to the specific choice of the extrapolation function.

It should also be noted that extrapolation techniques using an artificial neural network (ANN) are proposed recently [35,36]. To train networks, one usually requires large data sets. However, it is still tough to achieve an enormous number of

*s.yoshida@nt.phys.s.u-tokyo.ac.jp

ab initio calculations while varying their inputs such as $N_{\max}$ and harmonic oscillator parameter $\hbar\Omega$. In future applications of full CI and also valence CI methods with importance truncation to heavier systems, it is strongly desired to develop an extrapolation technique applicable even to sparse data. The proposed method is applicable to small data sets too.

The validity of the model is demonstrated by taking the extrapolation problem for ground-state energies in FCI calculations as an example. The code is available on the author's GitHub page [37].

II. FORMULATION OF CONSTRAINED GAUSSIAN PROCESS

Gaussian process (GP) is a popular statistical method as a nonparametric regression model [38]. It is also becoming popular in physics due to its flexibility (see, e.g., recent applications published in APS journals [39–45]). The GP regression can be interpreted as a method to describe distribution over a function space and to perform inference of the probability for each function. This is just what is needed because this enables us to consider an ensemble of many possible functions for the extrapolation, infer probability of each sample function, and then quantify uncertainties in the extrapolated value in a statistical manner.

Interestingly, GPs are mathematically equivalent or related to many other models such as ANN, support vector machines, spline models, and so on. The interested reader is referred to, e.g., Refs. [38,46,47].

Here let us introduce some notations. As in statistics literature, $P(a|b)$ denotes the probability distribution of a under the condition b, and $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multivariate Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. In what follows, two variable sets, data and prediction, are considered. The terminology data, which is distinguished from experimental data, is used to express a set of $X = \{x_i | i = 1, \ldots, D\}$ and $Y = \{y_i | i = 1, \ldots, D\}$. Here it is assumed that D input points are available. The prediction represents positions $X^* = \{x_i^* | i = 1, \ldots, P\}$ and values $Y^* = \{y_i^* | i = 1, \ldots, P\}$ for P points where the target values are not known. Especially in the following application, X denotes currently computable $N_{\max}$, and X^* is a set of $N_{\max}$ at which FCI calculations have not done (e.g., larger $N_{\max}$).

In Gaussian processes, it is assumed that the two target values at the two arbitrary points in the vicinity must be similar, and the so-called kernel functions express the similarities. Then the data values Y and prediction values Y^* are assumed to be generated from the multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ whose covariance matrix $\boldsymbol{\Sigma}$ is given as

$$\Sigma = \begin{bmatrix} K_{XX} & K_{XX^*} \\ K_{XX^*}^T & K_{X^*X^*} \end{bmatrix}. \tag{1}$$

Here K_{XX}, K_{XX^*}, and $K_{X^*X^*}$ are, respectively, $D \times D$, $D \times P$, and $P \times P$ matrices, and these elements are evaluated with a kernel function. It is a common practice to choose this kernel function as the radial basis function (RBF) or the Matérn kernel with $\nu = 3/2$ (Mat32) or $\nu = 5/2$ (Mat52) (see Appendix A).

In this work, the logMat52 kernel function is used for the reasons described in Appendix A. The logMat52 kernel is defined for, e.g., two data points x_i and x_j, as follows:

$$k_{\text{logMat52}}(x_i, x_j) = \tau \left(1 + \frac{\sqrt{5}\eta}{\ell} + \frac{5\eta^2}{3\ell^2}\right) \exp\left(-\frac{\sqrt{5}\eta}{\ell}\right), \tag{2}$$

where $\eta \equiv |\ln x_i - \ln x_j|$, and the global strength τ and the correlation length ℓ are the hyperparameters. Let $\boldsymbol{\theta}$ denote the vector of hyperparameters. I will revisit the issue of hyperparameters later.

Once the kernel function and its hyperparameters are fixed, one can define the joint covariance matrix $\boldsymbol{\Sigma}$ in Eq. (1) for data/prediction as a function of $\boldsymbol{\theta}$. Then, the joint distribution of data $\boldsymbol{y}$ and predictions $\boldsymbol{y}^*$ under the hyperparameters is given as

$$P(\boldsymbol{y}, \boldsymbol{y}^* | \boldsymbol{\theta}) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}^* \end{bmatrix}, \Sigma(\boldsymbol{\theta})\right). \tag{3}$$

It is a common practice for mean vectors to be normalized, i.e., $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^*$ are fixed as zero vectors and the data is scaled to have zero mean and unit variance. The dependence on the choice of mean vectors is also discussed later.

By definition of conditional probabilities, the left-hand side of Eq. (3) can be rewritten as

$$P(\boldsymbol{y}, \boldsymbol{y}^* | \boldsymbol{\theta}) = P(\boldsymbol{y}^* | \boldsymbol{y}, \boldsymbol{\theta}) P(\boldsymbol{y} | \boldsymbol{\theta}). \tag{4}$$

Under given $\boldsymbol{\theta}$, one can write down $P(\boldsymbol{y}^* | \boldsymbol{y}, \boldsymbol{\theta})$ and $P(\boldsymbol{y} | \boldsymbol{\theta})$ in a closed form:

$$P(\boldsymbol{y}^* | \boldsymbol{y}, \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{y}^*|\boldsymbol{y}}, \Sigma_{\boldsymbol{y}^*|\boldsymbol{y}}), \tag{5}$$

$$\boldsymbol{\mu}_{\boldsymbol{y}^*|\boldsymbol{y}}(\boldsymbol{\theta}) = \boldsymbol{\mu}^* + K_{XX^*}^T K_{XX}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}), \tag{6}$$

$$\Sigma_{\boldsymbol{y}^*|\boldsymbol{y}}(\boldsymbol{\theta}) = K_{X^*X^*} - K_{XX^*}^T K_{XX}^{-1} K_{XX^*}, \tag{7}$$

$$P(\boldsymbol{y} | \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}, K_{XX}). \tag{8}$$

Regarding the hyperparameters, the so-called maximum *a posteriori* (MAP), i.e., one to maximize the hyperparameter posterior $P(\boldsymbol{\theta}|\boldsymbol{y})$, is often used. However, I do not use a single value for the hyperparameters to receive benefit of GPs; a broader class of sample functions can be obtained by considering various hyperparameters. Their probability distributions are inferred by a sampling scheme to integrate out the hyperparameter dependence. In this case, the posterior distribution of $\boldsymbol{y}^*$ for unobserved input $\boldsymbol{x}^*$ is written as

$$P(\boldsymbol{y}^* | \boldsymbol{y}) \propto \int P(\boldsymbol{y}^* | \boldsymbol{y}, \boldsymbol{\theta}) P(\boldsymbol{y} | \boldsymbol{\theta}) P(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{9}$$

In addition to this, the formulation above is extended to more general one to incorporate physics information into GPs. In many practical situations, the target function is known to have shape constraints (e.g., monotonicity or convexity) or inequality constraints. That is also the case with problems of interest, i.e., energy eigenvalues in FCI are monotonic and (almost) convex with respect to $N_{\max}$. In general, the accuracy of a statistical model such as GP is improved by including

such physics information. To this end, Eq. (9) is extended to

$$P(\boldsymbol{y}^*|\boldsymbol{y},\alpha,\beta,\ldots) \propto \int P(\boldsymbol{y}^*|\boldsymbol{y},\boldsymbol{\theta})P(\boldsymbol{y}|\boldsymbol{\theta})P(\boldsymbol{\theta}) \times P(\alpha,\beta,\ldots|\boldsymbol{y}^*,\boldsymbol{y})d\boldsymbol{\theta}, \tag{10}$$

where $P(\alpha,\beta,\ldots|\boldsymbol{y}^*,\boldsymbol{y})$ is the probability that the constraints $\alpha,\beta,\ldots$ are satisfied under the given $\boldsymbol{y}^*$ and $\boldsymbol{y}$. The contribution to the integral is determined by the balance among the posterior for the prediction, the likelihood for the hyperparameters, the hyperparameter prior, and the fidelity to the constraints. This expression is justified when the $\boldsymbol{\theta}$ is independent of the constraints (see Appendix B. This is a rather general expression, i.e., constraints can be introduced independently for each problem of interest.

In general, the integration in Eq. (10) cannot be evaluated analytically. Therefore, some approximation or sampling scheme is required. In what follows, the integration in Eq. (10) is evaluated by weighted N_p samples as follows:

$$P(\boldsymbol{y}^*|\boldsymbol{y},\alpha,\beta,\ldots) \simeq \sum_{i=1}^{N_p} w^{(i)} P(\boldsymbol{y}^{*(i)}|\boldsymbol{y},\boldsymbol{\theta}^{(i)}), \tag{11}$$

$$w^{(i)} \equiv \frac{P(\boldsymbol{y},\boldsymbol{\theta}^{(i)})P(\alpha,\beta,\ldots|\boldsymbol{y}^{*(i)},\boldsymbol{y})}{\sum_{j=1}^{N_p} P(\boldsymbol{y},\boldsymbol{\theta}^{(j)})P(\boldsymbol{y}^{*(j)}|\boldsymbol{y},\boldsymbol{\theta}^{(j)})P(\alpha,\beta,\ldots|\boldsymbol{y}^{*(j)},\boldsymbol{y})}. \tag{12}$$

In the present study, the particle filtering method [48] (also known as sequential Monte Carlo) is employed as a sampling scheme to evaluate the summation in Eq. (11). In the particle filtering algorithm, states $\{\boldsymbol{\theta}^{(i)},\boldsymbol{y}^{*(i)}\}$ are assigned to particles labeled by $i = 1, 2, \ldots, N_p$, and those particles are evolved independently according to the Metropolis-Hastings method with the so-called resampling scheme; at a certain step of the algorithm, the particles that do not respect the physics constraints are discarded.

III. PROBLEMS OF INTEREST

A. FCI results

In what follows, the constrained GP model is applied to extrapolation problems in FCI calculations; I analyze published FCI results of the ground-state energy of ^{6}Li using JISP16/NNLO$_{\rm opt}$ interaction with $\hbar\Omega = 17.5$ MeV [49] and N3LO interaction with $\hbar\Omega = 16.0$ MeV, which is softened by similarity renormalization group (SRG) method with a flow parameter $\lambda = 2.02$ fm^{-1} [50]. The results are summarized in Fig. 1 as a function of $N_{\rm max}$.

Let $\{(x_1,y_1),(x_2,y_2),\ldots,(x_D,y_D)|x_1 < x_2 < \cdots < x_D\}$ denote the data, i.e., $(x_1,y_1) = (6,-28.602)$, …, $(x_D,y_D) = (14,-31.977)$ in the case of N3LO results. Unlike least-squares fitting of parametric models in which one should remove outliers from data, there is no reason to reduce data in the GP model and all $N_{\rm max}$ results are used as data unless otherwise mentioned.

The extrapolation problem addressed below is to estimate the ground-state energies at $N_{\rm max} > x_D$, and these are expressed as $\{(x_1^*,y_1^*),(x_2^*,y_2^*),\ldots,(x_P^*,y_P^*)|x_1^* < x_2^* < \cdots < x_P^*\}$; Here $x_1^* = x_D + 2$ and P is a large even integer. For the sake of simplicity, let us consider only the ground state of ^{6}Li with natural parity, i.e., only even $N_{\rm max}$. In practice, one needs to truncate at certain finite P value where predictions are converged with respect to $N_{\rm max}$. A detailed discussion about this P will be given later.

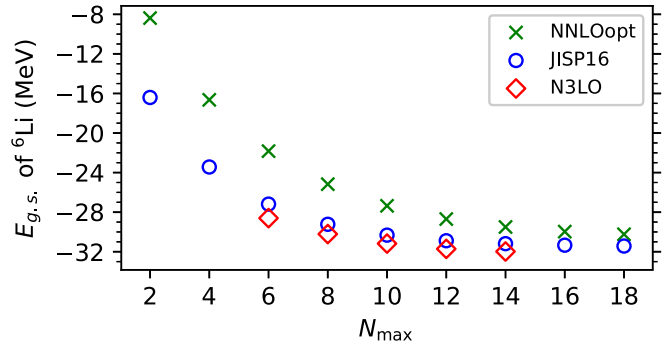

FIG. 1. The FCI results of g.s. energy of ^{6}Li using JISP16/NNLO$_{\rm opt}$ [49] and N3LO [50].

B. Constraints on extrapolation function

As minimal constraints on the extrapolation function to capture the asymptotic behavior of FCI results, the following two constraints, α and β are imposed. The first constraint α is the variational property, i.e., the monotonicity of energy eigenvalues with respect to $N_{\rm max}$:

$$P(\alpha|\boldsymbol{y}^*,\boldsymbol{y}) = \Phi(y_D - y_1^*) \times \Phi(y_1^* - y_2^*) \times \Phi(y_{P-1}^* - y_P^*), \tag{13}$$

where $\Phi(\cdot)$ is the cumulative distribution function

$$\Phi(z) \equiv \int_{-\infty}^{z\kappa} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt, \tag{14}$$

with the controlling parameter of the strictness of constraints κ. This $\Phi(z)$ approaches the step function at $z = 0$ when $\kappa \to \infty$. This κ is set as about 10^6 in the code, which is large enough to impose the constraints with satisfactory accuracy less than 0.01 keV. More precisely, κ is gradually increased in the code to $\approx 10^6$ so as to avoid possible localization at the early steps of the Monte Carlo sampling. I confirmed that the form of this κ as a function of the Monte Carlo step does not affect the extrapolation results other than the sampling efficiency.

The second constraint β is about the convergence pattern. A ratio of the absolute value of energy gains is used as a measure of convergence of ground-state (g.s.) energy in FCI. This ratio r at a certain point x_j is defined as follows:

$$r(x_j) \equiv \left|\frac{y_j - y_{j+1}}{y_{j-1} - y_j}\right|. \tag{15}$$

The $\{r(x_j)\}$ for the given data is shown in Fig. 2. For the energy eigenvalues by the FCI method, the denominator and numerator in Eq. (15) are both positive. If the calculated results of the g.s. energy exactly obey an exponential function, it means r is a constant, which is not the case as shown in Fig. 2.

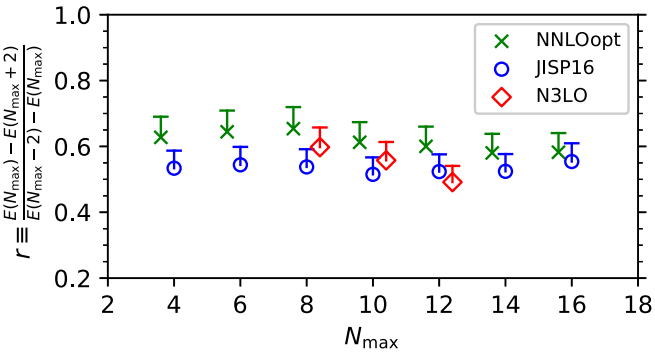

FIG. 2. The ratios of energy gains associated with 10% fluctuation (error bar). The symbols for FCI results are the same as Fig. 1. For the visibility of the figure, $\mathrm{NNLO_{opt}}$ and N3LO results are slightly shifted to the left and right, respectively. See the text for more details.

This r can be extended to include predictions $\{y_j^*\}$:

$$r(x_1^*) = \left|\frac{y_1^* - y_2^*}{y_D - y_1^*}\right|, \quad r(x_2^*) = \left|\frac{y_2^* - y_3^*}{y_1^* - y_2^*}\right|, \ldots,$$
$$r(x_{P-1}^*) = \left|\frac{y_{P-1}^* - y_P^*}{y_{P-2}^* - y_{P-1}^*}\right|. \tag{16}$$

In the present work, the constraint on $\{r\}$ is imposed as follows:

$$P(\beta|\boldsymbol{y}^*, \boldsymbol{y}) = \Phi(R_E - r(x_1^*)) \times \Phi(R_E - r(x_2^*)) \times \cdots \times \Phi(R_E - r(x_{P-1}^*)), \tag{17}$$

where R_E is an upper threshold determined as follows:

$$R_E = r_{\mathrm{mean}} + r_{\mathrm{std}}, \tag{18}$$
$$r_{\mathrm{mean}} \equiv \left|\frac{y_{D-1} - y_D}{y_{D-2} - y_{D-1}}\right|, \tag{19}$$
$$r_{\mathrm{std}} \equiv \sigma_r r_{\mathrm{mean}}. \tag{20}$$

The σ_r is fixed as 0.1 throughout this work for simplicity. As can be expected from Fig. 2, this is a rather moderate constraint on the convergence pattern. When σ_r is large enough, results agree with ones with only the constraint α.

In what follows, the GP extrapolation model using constraints is referred to as constrained Gaussian process (cGP) model.

C. Choice of the mean function

Here let us introduce two different mean functions $\boldsymbol{\mu}^{(*)}$ needed in Eqs. (3)–(8):

(i) (case a) zero mean: $\boldsymbol{\mu} = \mathbf{0}_D$, $\boldsymbol{\mu}^* = \mathbf{0}_P$;
(ii) (case b) B3 fit: mean of data and prediction are both determined by B3 fit [19], i.e., minimizing χ^2 deviation between the largest three $N_{\max}$ data and the exponential function in the form of $E_\infty + c_0 \exp(-c_1 N_{\max})$ with three free parameters (E_∞, c_0, c_1). In this choice, it can be said that preliminary knowledge on the behavior of the quantity is included in terms of the mean function of GPs.

In the following, these are referred to as cGP-a and cGP-b, respectively, and I analyze both cases below.

IV. RESULTS AND DISCUSSIONS

A. Extrapolation of g.s. energies

The cGP predictions for the ground-state energies of ^{6}Li are shown in Fig. 3. Now the extrapolated values at a certain $N_{\max}$ are represented by the histograms. The cGP-a results are shown by the transparent histograms colored in green, and the hatched histograms colored in pink are for the cGP-b results.

It should be noted that the prediction is truncated up to a certain $N_{\max}$ where the mean value is converged within 0.2 keV. This means that possible differences between predictions at the finite $N_{\max}$ and one at $N_{\max} = \infty$ are suppressed below 1 keV because of the constraint β. The $N_{\max}$ giving converged results are 44,38 and 34 for $\mathrm{NNLO_{opt}}$, JISP16, and N3LO, respectively. These numbers are consistent with the intuition that harder interaction requires larger $N_{\max}$ to obtain converged results, while it should be noted that these results are for different $\hbar\Omega$. In addition to the mean values, the 68% and 99% credible intervals are shown and plotted below the histograms. In this work, the 68% interval is determined from the 16th and 84th percentile of the distribution, and the 99%

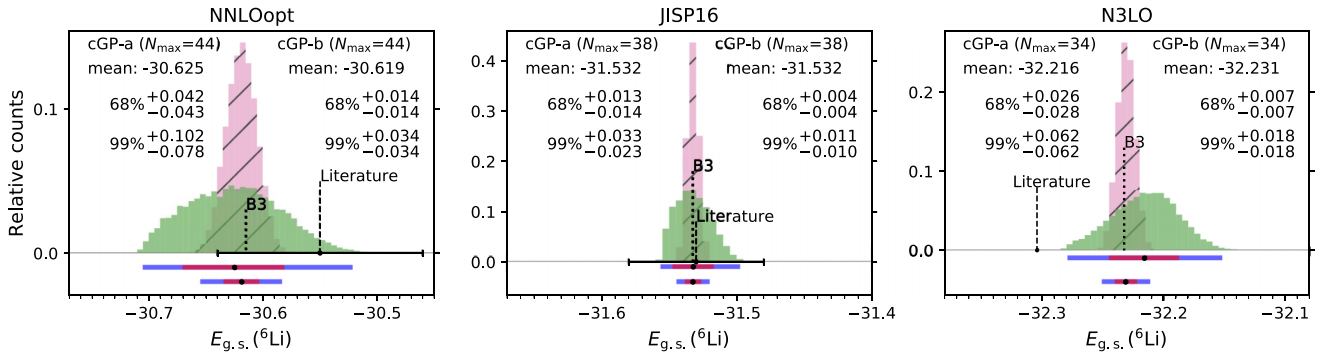

FIG. 3. Posterior distributions of extrapolated value at certain $N_{\max}$ are shown by histograms with 5 keV width. The B3 (dotted lines) denote the extrapolated values with exponential function and dashed lines are associated with extrapolated values in Refs. [49,50] with/without error bar.

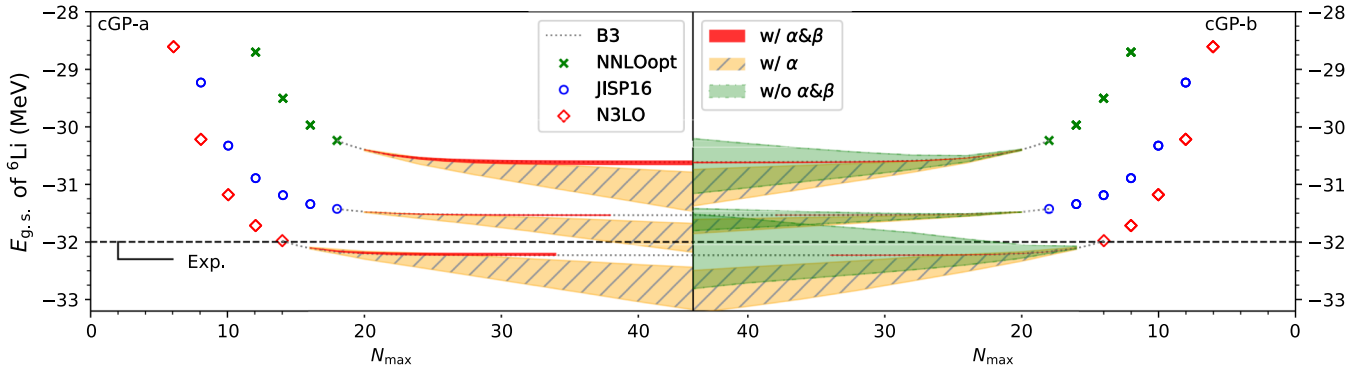

FIG. 4. The plot showing the impact of the constraints α (monotonicity) and β (convexitylike). The bands colored in red, orange (hatched), and green (with dashed dotted line), show the 68% credible intervals of GP predictions with α and β, only α, and without constraints, respectively. The red bands are cut at a certain N_{max}, which gives converged results.

interval is defined in a similar manner. As a whole, the cGP-b gives smaller credible intervals than the cGP-a.

Other extrapolated values are also shown. The B3 denotes the exponential fit using the largest three N_{max} data. The conventional B3 extrapolation is always included as a special case of two cGP results. The values for literature are from Refs. [49,50]. Here I note that the literature value for N3LO [50] might be obtained by an exponential fit using all five data, while it is not explicitly stated. It must also be noted that it is a highly nontrivial task to fairly compare the results with different extrapolation techniques, because data is truncated in parametric models and some use data with multiple $\hbar\Omega$ as in the A5 extrapolation [49].

In the present case, the sampling scheme with the particle filtering gives converged results within a few keV in case of 20000 particles after 2000 times Metropolis-Hastings updates for each particle, and independent runs reproduce the same results within the Monte Carlo error.

B. Impact of the constraints

Here let us see how extrapolation results are influenced by the two constraints imposed. In Fig. 4, the impact of the constraints is shown. All symbols are the same as in Fig. 1, and the cGP-a and cGP-b results are summarized in the left and right regions, respectively. The bands correspond to the 68% credible intervals of GP predictions with α and β (red), with the only α (hatched orange), and without constraints (green), respectively. Since the 68% errors for the cGP-b prediction is an order of ten keV, the red lines in the right panel are very narrow.

As shown in the textbook [38], predictions of unconstrained zero-mean GP at points far from the data domain converge to zero with a fixed standard error. The case of cGP-a without α and β, which is obviously not appropriate for the current purpose, is omitted for this reason. When one assumes that the wave function is dominated by relatively lower N_{max} configurations, predictions with both constraints α and β are expected to be more reliable than the others.

C. Data dependence

I have used all N_{max} results as data so far. Next, let us explore the dependence of extrapolated values on the used data to test the potential predictive power of the proposed method. In Fig. 5, extrapolated values for both cGP-a and

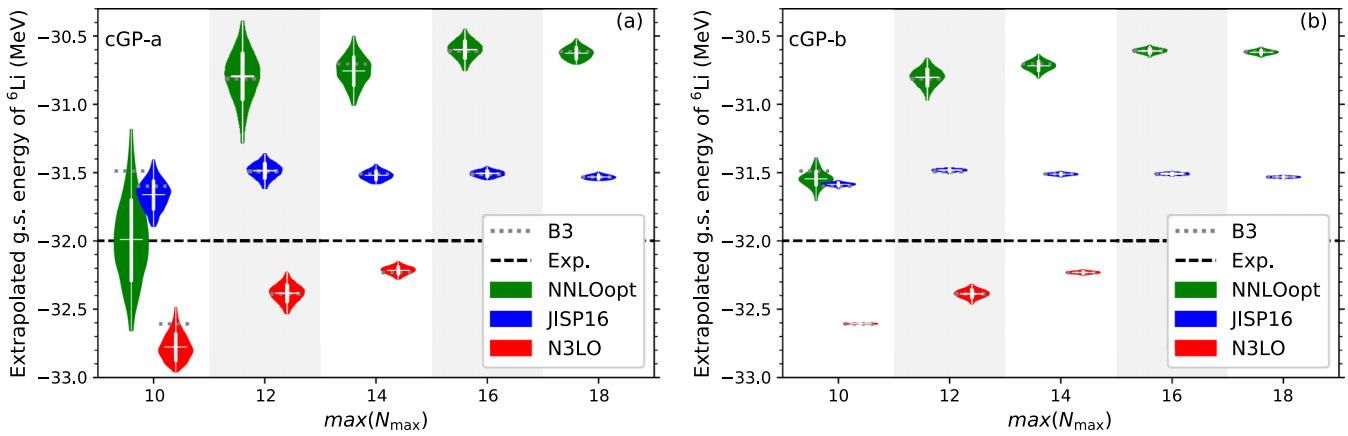

FIG. 5. Extrapolated ground-state energies as a function of maximum N_{max} used as data. The mean values by the cGP models are shown by horizontal line in white. Shaded areas are obtained by a kernel density estimation of the posterior distribution. The two types of vertical solid lines, thick and thin ones, show 68% and 99% credible intervals, respectively. For the visibility, the results for $NNLO_{opt}$ and N3LO are slightly shifted to left and right, respectively.

cGP-b are shown as a function of the maximum $N_{\max}$ used as data, i.e., $x_D = \max(N_{\max})$.

The mean predictions by the cGP models are shown by horizontal line in white, and the shaded areas show the posterior distributions obtained by 20000 particles. The 68% and 99% credible intervals, respectively, are shown by the thick and thin vertical lines, respectively. To improve the readability of the figure, the width of the shaded area is scaled to be the same for each area, and $\mathrm{NNLO_{opt}}$ (N3LO) results are slightly shifted to left (right).

As a whole, the size of credible intervals for the cGP-a is larger than that for the cGP-b, and the credible intervals become smaller as higher $N_{\max}$ data is added with only one exception, i.e., the cGP-b result for N3LO with $\max(N_{\max} = 10)$. This exception can be understood from Eqs. (4)–(8). In this case, the exponential function exactly fits the given three data and then $\boldsymbol{\mu}_{\boldsymbol{y}^*|\boldsymbol{y}}$ in Eq. (6) is identical with $\boldsymbol{\mu}^*$. Any fluctuation of the joint mean value $\boldsymbol{\mu}_{\boldsymbol{y}^*|\boldsymbol{y}}$ is not allowed, and this significantly reduces probability weights for functions other than the B3 fit.

These plots with quantified uncertainties tell us one criterion of where to stop the massive FCI calculations under the given extrapolation model; It is inadvisable to carry out FCI calculations while increasing $N_{\max}$ forever, whereas the point to stop must depend on the desired accuracy. In the rest of this section, let us regard the mean values at rightmost $\max(N_{\max})$ in Fig. 5, i.e., $\max(N_{\max}) = 18$ for $\mathrm{NNLO_{opt}}$ and JISP16, and $\max(N_{\max}) = 14$ for N3LO, as the tentative exact values.

For $\mathrm{NNLO_{opt}}$ and JISP16, an important remark is that the exact values are covered by the cGP-a predictions with a relatively lower $\max(N_{\max})$. The sign of convergence can be seen around $\max(N_{\max}) = 10$–12, i.e., one can choose these as points to stop the calculation. From the behavior of the credible intervals, the cGP-b seems to underestimate the uncertainties than the cGP-a. In other words, the cGP-b takes account of fluctuation of the functional form only around the exponential function, while the cGP-a would include a wider class of functions.

For N3LO, any predictions by cGP-a, cGP-b, and B3 fit with lower $\max(N_{\max})$ have almost no overlap with the tentative exact value. This means that all the models fail to estimate the extrapolation uncertainty, while the cGP-a could be better than the others. The extrapolated values are much more sensitive to the $\max(N_{\max})$ than the results with other interactions; the extrapolation for the N3LO results seems to be a more nontrivial problem than the others. This can be understood from the behavior of the ratio of energy gains r. Especially in the N3LO case, the r is unstable with respect to $N_{\max}$, as seen in Fig. 2. One possibility to cause this nonflat behavior of r is that the calculation have not yet converged, i.e., the additional bindings by increasing $N_{\max}$ cannot be regarded as a simple asymptotic behavior. However, there is not yet enough open data to conclude the origin of the nonflat behavior of r. If one could figure out an additional constraint on the behavior of extrapolated values as a function of $\max(N_{\max})$, such a difficulty in extrapolation, which could be observed in particular nuclei and/or particular interactions, would be alleviated.

V. SUMMARY AND OUTLOOK

In this work, an extrapolation method for CI-type calculations using constrained Gaussian processes is proposed. This method has the following advantages that are required for future generations of *ab initio* studies to achieve more quantitative discussions on observables of interest and on the quality of adopted nuclear interactions.

First, this method does not need to remove outliers and has applicability to sparse data sets, which are strongly needed for future FCI calculations. Second, one can naturally incorporate domain knowledge into the model. It is often the case especially in physics that one knows in advance the behavior of the target quantity at a certain level, which is ranging from empirical laws to physical principles. One can expect that imposing such information into the extrapolation model improves the accuracy of the predictions. This flexibility might be useful to alleviate difficulties in the extrapolation for some particular cases. Third, uncertainty in extrapolation can be quantified in a systematic manner. Although the main source of uncertainties in FCI calculations is the input nuclear potential, evaluating extrapolation uncertainties are helpful for further understandings about the nuclear observables.

Regarding uncertainties from input parameters in nuclear many-body methods, the tremendous efforts to propagate input uncertainty to the observables have been made in the last decade, see, e.g., Refs. [51–65]. In addition to these, it has been shown that eigenvector continuation (EC), which is introduced in Ref. [66], can be used as an efficient emulator of *ab initio* methods, then used for uncertainty quantification and sensitivity analyses on input parameters such as low-energy constants in the chiral EFT potentials [67,68]. It is expected that EC (and some other method) facilitates comprehensive studies of uncertainty propagation in *ab initio* methods combined with an extrapolation method with quantified uncertainties.

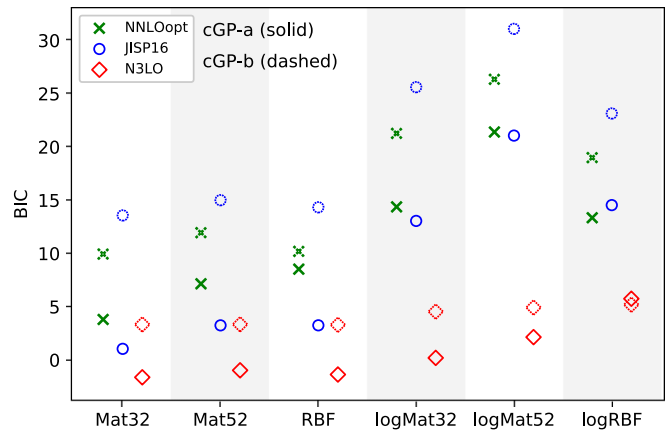

FIG. 6. The BIC for six kernel functions, Mat32 (Matérn kernel with $\nu = 3/2$), Mat52 (Matérn kernel with $\nu = 5/2$), RBF (radial basis function), and their counterparts with logarithm distances. Results for $\mathrm{NNLO_{opt}}$ (N3LO) are slightly shifted to left (right). Symbols drawn by solid and dashed lines are for cGP-a (zero mean) and cGP-b (exponential mean), respectively.

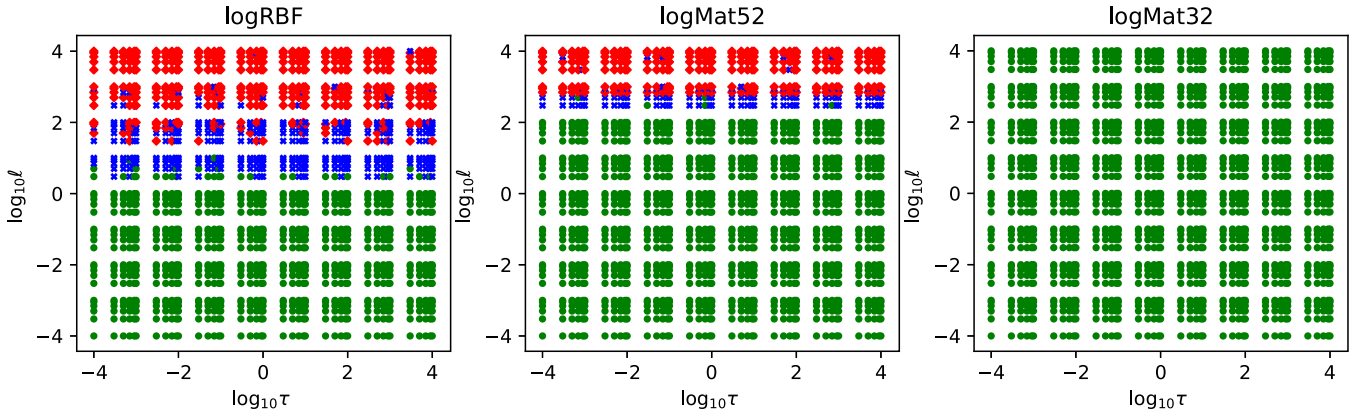

FIG. 7. The summary of PSD properties for the three kernel functions. The covariance matrices having PSD are shown by the filled circle (green). The diamond symbols (red) mean that both K_{XX} and $\Sigma_{y^*|y}$ are non-PSD, and the cross symbols (blue) are assigned if only the $\Sigma_{y^*|y}$ is non-PSD.

The benefits of the uncertainty quantification are not limited to putting error bars in predictions. If one properly propagates uncertainties from the input interaction and also quantifies uncertainties such as that due to the extrapolation, it enables us to visualize nonlinear relation between input and output of many-body calculations and capability of the many-body method. Then it would provide us with footholds to understand some missing contributions, if there were any.

In the present work, the extrapolation method is applied to only the ground-state energies obtained by FCI calculations. When it comes to the extrapolation problem of other quantities or in other systems, the main problem is to find minimal constraints to capture the asymptotic behavior of the quantities well. It is a possible future direction along this line to extend the cGP model to a higher dimension. In case of FCI calculations, for example, one can impose the following additional constraint on GP by extending the formulation to $(N_{\max}, \hbar\Omega)$ space: extrapolated values with different $\hbar\Omega$ should converge to the same value to some extent. The extension of the formulation to a multidimensional space is rather straightforward, while it is expected that one needs more technical analyses in numerical studies such as positive semidefiniteness of covariance matrices. Our model can also be applied to valence CI techniques using an importance truncation in which the extrapolation function is much more nontrivial than the FCI case. It would also be interesting to apply this kind of cGP to finite-size scaling analyses in other systems.

ACKNOWLEDGMENTS

The author thanks Noritaka Shimizu for fruitful discussions on future applications to the extrapolation problem in Monte Carlo shell-model calculations. This work was supported by JSPS KAKENHI (Grant No. 17J06775).

APPENDIX A: KERNEL SELECTION

In this section, I explore the technical details to choose the kernel function. As mentioned in the main text, popular choices are RBF kernel:

$$k_{\mathrm{RBF}}(x_i, x_j) = \tau \exp\left(-\frac{(x_i - x_j)^2}{2\ell^2}\right), \tag{A1}$$

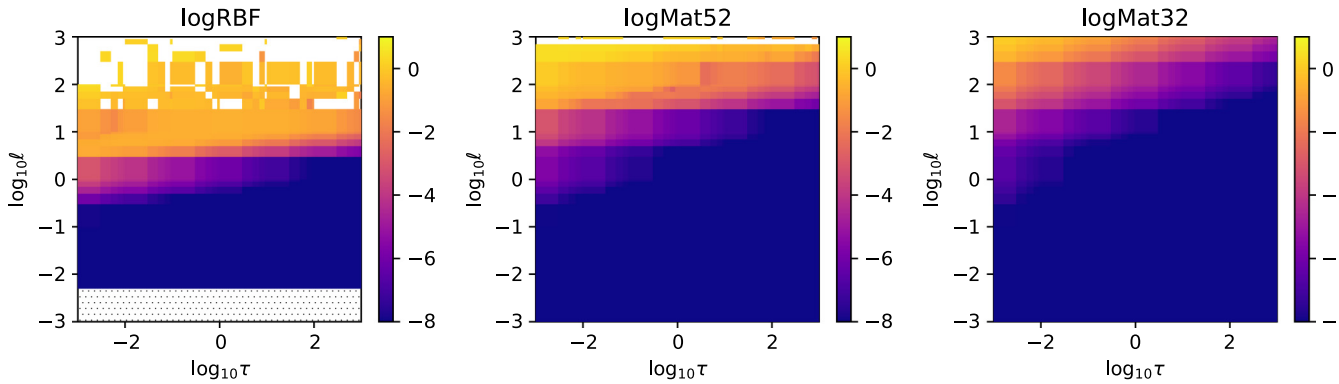

FIG. 8. The plot showing the impact of the ϵ prescription to avoid non-PSD. The color map shows the Δ in Eq. (A9). The hatched region with dots show the points at which the two codes gives the exactly the same result. The region where the code does not give answer is colored in white.

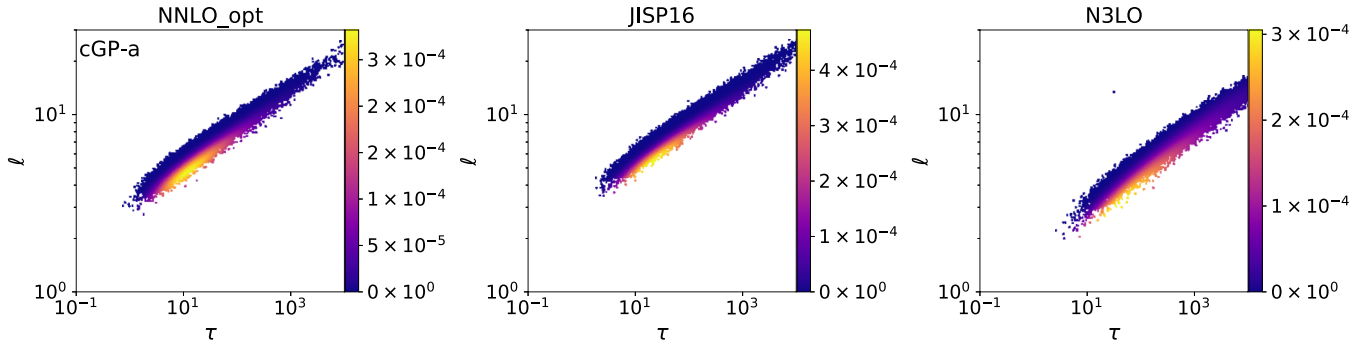

FIG. 9. Projection of the particles to hyperparameter space. The colors show relative contribution of $P(\boldsymbol{\theta}|\boldsymbol{y})$ for cGP-a.

and Matérn kernel:

$$k_{\mathrm{Matrn}}(x_i, x_j; \nu) = \tau \frac{2^{1-\nu}}{\Gamma(\nu)} \xi^{\nu} K_{\nu}(\xi), \tag{A2}$$

$$\xi \equiv \frac{\sqrt{2\nu}|x_i - x_j|}{\ell}, \tag{A3}$$

where Γ is the gamma function and K_ν is the modified Bessel function of the second kind. Especially, Matérn kernels with $\nu = 3/2$ and $\nu = 5/2$ are commonly used. The RBF kernel corresponds to the special case of the Matérn kernel with $\nu = \infty$. For the Matérn kernel, a sample function is k-times mean-square differentiable if and only if $\nu > k$ [38]. For that reason, the Matérn kernels with $\nu > 3/2$ are thought to be appropriate for the purpose, i.e., extrapolation of FCI results, but the analysis below includes the $\nu = 3/2$ case too.

In what follows, let us consider these Mat32 (Matérn kernel with $\nu = 3/2$), Mat52 (Matérn kernel with $\nu = 5/2$), RBF (radial basis function), and their counterparts with logarithm distances:

$$k_{\mathrm{logRBF}} = \tau \exp\left(-\frac{\eta^2}{2\ell^2}\right), \tag{A4}$$

$$k_{\mathrm{logMat52}} = \tau\left(1 + \frac{\sqrt{5}\eta}{\ell} + \frac{5\eta^2}{3\ell^2}\right)\exp\left(-\frac{\sqrt{5}\eta}{\ell}\right), \tag{A5}$$

$$k_{\mathrm{logMat32}} = \tau\left(1 + \frac{\sqrt{3}\eta}{\ell}\right)\exp\left(-\frac{\sqrt{3}\eta}{\ell}\right), \tag{A6}$$

where $\eta \equiv |\ln x_i - \ln x_j|$. Taking the logarithm distance, i.e., replacing $|x_i - x_j|$ in, e.g., Eq. (A1) by $|\ln x_i - \ln x_j|$, makes results independent of the scale of the x axis, and allows the capture of the nonstationary nature of FCI results (results rapidly converge to certain values as functions of $N_{\max}$) [69].

To compare the six kernels, the Bayesian information criterion (BIC) is used:

$$\mathrm{BIC} = \ln P(\boldsymbol{y}|\boldsymbol{\theta}_{\mathrm{ML}}) - \tfrac{1}{2} M \ln n, \tag{A7}$$

where $\boldsymbol{\theta}_{\mathrm{ML}}$ is the maximum likelihood estimation under the given data, M is the dimension of the hyperparameter $\boldsymbol{\theta}$, and n is the number of data. In this case, the values of the latter term are common among the results with the given interaction. In Fig. 6, the BIC for the six kernels is summarized. Colors and symbols for three interactions are the same as Fig. 1. The $\boldsymbol{\theta}_{\mathrm{ML}}$ is obtained through updates of 2000 independent particles by the Metropolis-Hastings method, and the $\boldsymbol{\theta}_{\mathrm{ML}}$ is converged less than 0.1% accuracy. As a whole, taking the logarithm of the distance increases the maximum log-likelihood, which is now equivalent to the maximum BIC.

As stated in the main text, the logMat52 kernel is used throughout this work. This is partly because the logMat52 gives the highest BIC in total. The other reason is its numerical stability over the (log)RBF kernel. The RBF is one of the most popular choices for the kernel. However, its smoothness of the sample functions is often regarded as too high [70], and, in practice, this too smooth nature sometimes breaks down the positive semidefiniteness of covariant matrices in numerical calculations due to rounding errors. That is demonstrated in the following.

In applications of GPs, one needs to obtain conditional mean vectors Eq. (6) and covariance matrices Eq. (7). The covariance matrices such as K_{XX} and $\boldsymbol{\Sigma}$ must be positive semidefinite (PSD) to achieve, e.g., the Cholesky decomposition for K_{XX}^{-1} and to generate samples from the posterior

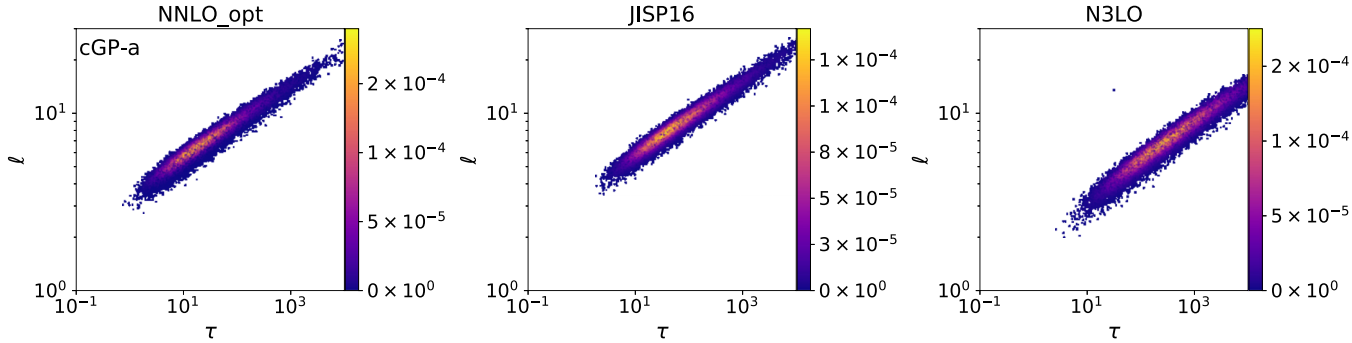

FIG. 10. Projection of the particles to the hyperparameter space. The colors show relative contributions to Eq. (11) in cGP-a.

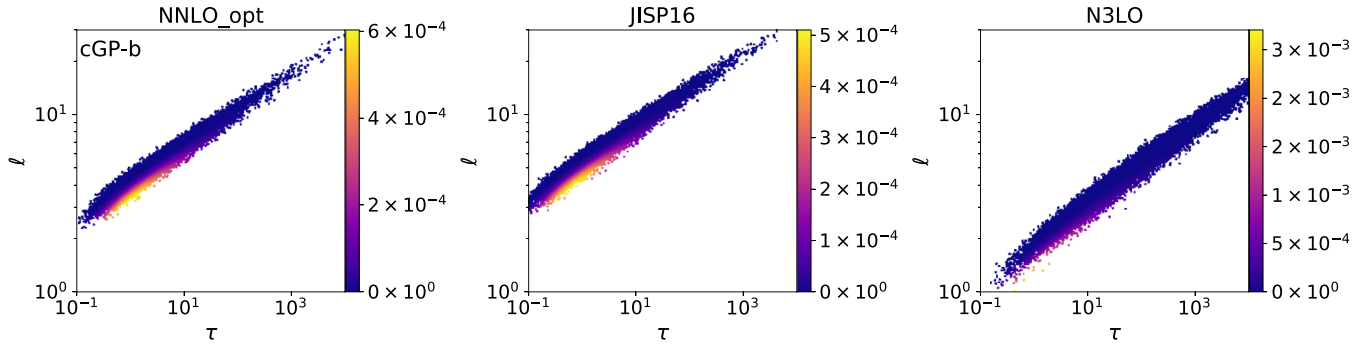

FIG. 11. Projection of the particles to hyperparameter space. The colors show relative contribution of $P(\boldsymbol{\theta}|\boldsymbol{y})$ for cGP-b.

distribution. In some cases (e.g., points are located too close to each other), covariance matrices become non-PSD due to rounding errors, although it must be PSD mathematically. This is also true for the extrapolation method in the present study. The typical prescription to this non-PSD is to add an infinitesimal diagonal matrix to the K_{XX} and/or $\boldsymbol{\Sigma}$. Let us call this the ϵ prescription in the following. This ϵ prescription is mathematically equivalent to assume the observation and/or prediction to have noise. In the current case the target quantity is the calculated energies of light nuclei by *ab initio* full configuration interaction method, i.e., in the order of a few tens of MeV. On the other hand, the typical convergence tolerance of the Lanczos method in shell-model codes on the market is 1.e-5 MeV or better [16,17], i.e., the problems of interest are almost noiseless.

Let A denote a symmetric PSD matrix. If A is perturbed to $A + \delta A$ then A^{-1} becomes $A^{-1} + X$ where

$$||X|| \leqslant ||A^{-1}||^2 ||\delta A|| \tag{A8}$$

for any matrix norm $||\cdot||$ (e.g., Ref. [71]). The upper bound of norm of $||X||$ is not tight when the condition number of A is large, which corresponds to a numerically non-PSD covariance. This error induced by the ϵ prescription is propagated to mean vectors and covariances through Eq. (6) and Eq. (7), and affect generated samples and evaluation of posteriors.

In Fig. 7, the positive semidefiniteness of the covariance matrices are summarized; the conditional mean vector and covariance matrix for the N3LO data are calculated while varying the hyperparameters τ and ℓ in Eq. (A4)–(A6). The diamond symbols (red) correspond to the case that both K_{XX} and $\Sigma_{y^*|y}$ are non-PSD, and the cross symbols (blue) are assigned if only the $\Sigma_{y^*|y}$ is non-PSD. In terms of the length scale ℓ, the logRBF gives non-PSD matrices easier than logMat52 by one or two orders of magnitude.

In Fig. 8, the following quantity is shown:

$$\Delta \equiv \log_{10}\left(\max\left(\left|\boldsymbol{\mu}^{\mathrm{w/}}_{y^*|y} - \boldsymbol{\mu}^{\mathrm{w/o}}_{y^*|y}\right|\right)\right), \tag{A9}$$

where the superscripts, w/ and w/o, mean the conditional mean vectors calculated with and without ϵ prescription, respectively. It is noted that the mean vectors are now given in a unit of MeV and that $\epsilon =$ 1.e-12 is used, which is much smaller than that used in common libraries such as scikit-learn [72] and GaussianProcesses.jl [73]. The hatched regions with dots for smaller ℓ means that the results of the two codes are exactly the same. The white regions appearing in Fig. 8 correspond to the diamond symbols in Fig. 7, i.e., the mean vectors cannot be evaluated without the ϵ prescription. As the length scale ℓ increased, the posterior covariance becomes non-PSD (cross symbols in Fig. 7), then the Δ becomes larger. In some cases, the deviations in mean vectors reach a few MeV, which are obviously non-negligible.

One should be careful as to how the PSD and the observation noise are treated in the codes or libraries and the possible impact of the ϵ prescriptions on the predictions, especially when one would like to integrate out the hyperparameter dependence. The codes to reproduce Figs. 7 and 8 are provided in Ref. [74]. From the analyses above, I concluded that the logMat52 is the most appropriate choice for the extrapolation method.

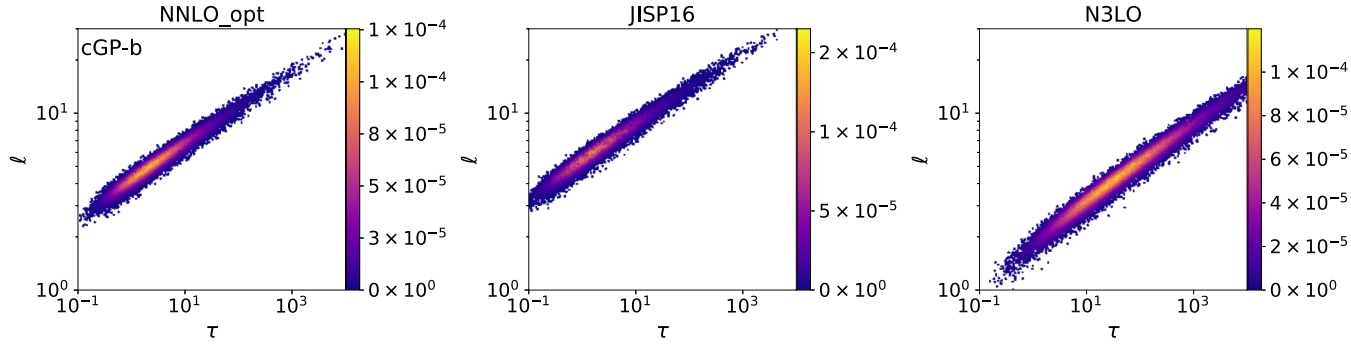

FIG. 12. Projection of the particles to the hyperparameter space. The colors show relative contributions to Eq. (11) in cGP-b.

APPENDIX B: EXTENSION OF THE POSTERIOR DISTRIBUTION UNDER THE CONSTRAINTS

Equation (10) is derived in the following. Let the $\boldsymbol{c}$ denote the physics constraints to be imposed.

$$P(\boldsymbol{y}^*, \boldsymbol{y}, \boldsymbol{c}) = \int P(\boldsymbol{y}^*, \boldsymbol{y}, \boldsymbol{c}, \boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{B1}$$

$$P(\boldsymbol{y}^*|\boldsymbol{y}, \boldsymbol{c}) = \int \frac{P(\boldsymbol{y}^*, \boldsymbol{y}, \boldsymbol{c}, \boldsymbol{\theta})}{P(\boldsymbol{y}, \boldsymbol{c})} d\boldsymbol{\theta}, \tag{B2}$$

$$= \int \frac{P(\boldsymbol{c}|\boldsymbol{y}^*, \boldsymbol{y}, \boldsymbol{\theta}) P(\boldsymbol{y}^*, \boldsymbol{y}, \boldsymbol{\theta})}{P(\boldsymbol{y}, \boldsymbol{c})} d\boldsymbol{\theta}, \tag{B3}$$

$$\propto \int P(\boldsymbol{c}|\boldsymbol{y}^*, \boldsymbol{y}) P(\boldsymbol{y}^*|\boldsymbol{y}, \boldsymbol{\theta}) P(\boldsymbol{y}|\boldsymbol{\theta}) P(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{B4}$$

In the last line, the constraints and the hyperparameters are assumed to be independent of each other.

APPENDIX C: CONTRIBUTIONS TO THE POSTERIOR DISTRIBUTION

In Figs. 9–12, the hyperparameter distributions are shown. Colors of each dot show the relative size of contributions to $P(\boldsymbol{\theta}|\boldsymbol{y})$ and to the right-hand side of Eq. (11). In the sampling method, particles are evolved according to the random walk Metropolis-Hastings method. One can see from Figs. 10 and 12 that particles distribute like an ellipse in hyperparameter space. The particles do not distribute over the non-PSD region discussed in Appendix A, and the relatively longer correlation length is favored so as to make sampled function obeying the monotonicity and convexitylike conditions. This is reflected as asymmetric color distributions of $P(\boldsymbol{\theta}|\boldsymbol{y})$ in Figs. 9 and 11.

[1] R. Machleidt and D. Entem, Phys. Rep. **503**, 1 (2011).
[2] E. Epelbaum, H.-W. Hammer, and U.-G. Meißner, Rev. Mod. Phys. **81**, 1773 (2009).
[3] J. Carlson, S. Gandolfi, F. Pederiva, S. C. Pieper, R. Schiavilla, K. E. Schmidt, and R. B. Wiringa, Rev. Mod. Phys. **87**, 1067 (2015).
[4] P. Navrátil, S. Quaglioni, G. Hupin, C. Romero-Redondo, and A. Calci, Phys. Scr. **91**, 053002 (2016).
[5] G. Hagen, T. Papenbrock, M. Hjorth-Jensen, and D. J. Dean, Rep. Prog. Phys. **77**, 096302 (2014).
[6] H. Hergert, S. Bogner, T. Morris, A. Schwenk, and K. Tsukiyama, Phys. Rep. **621**, 165 (2016), memorial volume in Honor of Gerald E. Brown.
[7] C. Barbieri and A. Carbone, Self-consistent Green's function approaches, in *An Advanced Course in Computational Nuclear Physics: Bridging the Scales from Quarks to Neutron Stars*, edited by M. Hjorth-Jensen, M. P. Lombardo, and U. van Kolck (Springer International Publishing, Cham, 2017), pp. 571–644.
[8] D. Lonardoni, J. Carlson, S. Gandolfi, J. E. Lynn, K. E. Schmidt, A. Schwenk, and X. B. Wang, Phys. Rev. Lett. **120**, 122502 (2018).
[9] T. D. Morris, J. Simonis, S. R. Stroberg, C. Stumpf, G. Hagen, J. D. Holt, G. R. Jansen, T. Papenbrock, R. Roth, and A. Schwenk, Phys. Rev. Lett. **120**, 152503 (2018).
[10] C. Drischler, K. Hebeler, and A. Schwenk, Phys. Rev. Lett. **122**, 042501 (2019).
[11] P. Gysbers, G. Hagen, J. D. Holt, G. R. Jansen, T. D. Morris, P. Navrátil, T. Papenbrock, S. Quaglioni, A. Schwenk, S. R. Stroberg, and K. A. Wendt, Nat. Phys. **15**, 428 (2019).
[12] B. R. Barrett, P. Navrátil, and J. P. Vary, Prog. Part. Nucl. Phys. **69**, 131 (2013).
[13] P. Navrátil, S. Quaglioni, I. Stetcu, and B. R. Barrett, J. Phys. G: Nucl. Part. Phys. **36**, 083101 (2009).
[14] E. Caurier, G. Martínez-Pinedo, F. Nowack, A. Poves, and A. P. Zuker, Rev. Mod. Phys. **77**, 427 (2005).
[15] J. P. Vary, The many-fermion dynamics shell-model code (unpublished); J. P. Vary and D. C. Zheng (unpublished); P. Sternberg, E. G. Ng, C. Yang, P. Maris, J. P. Vary, M. Sosonkina, and H. V. Le, in *Proceedings of the 2008 ACM/IEEE Conference on Supercomputing*, SC '08 (IEEE Press, Piscataway, 2008), pp. 15:1–15:12.
[16] N. Shimizu, arXiv:1310.5431; N. Shimizu, T. Mizusaki, Y. Utsuno, and Y. Tsunoda, Comput. Phys. Commun. **244**, 372 (2019).
[17] C. W. Johnson, W. E. Ormand, K. S. McElvain, and H. Shan, arXiv:1801.08432.
[18] C. Forssén, B. D. Carlsson, H. T. Johansson, D. Sääf, A. Bansal, G. Hagen, and T. Papenbrock, Phys. Rev. C **97**, 034328 (2018).
[19] P. Maris, J. P. Vary, and A. M. Shirokov, Phys. Rev. C **79**, 014308 (2009).
[20] H. Zhan, A. Nogga, B. R. Barrett, J. P. Vary, and P. Navrátil, Phys. Rev. C **69**, 034302 (2004).
[21] S. A. Coon, M. I. Avetian, M. K. G. Kruse, U. van Kolck, P. Maris, and J. P. Vary, Phys. Rev. C **86**, 054002 (2012).
[22] R. J. Furnstahl, G. Hagen, and T. Papenbrock, Phys. Rev. C **86**, 031301(R) (2012).
[23] S. N. More, A. Ekström, R. J. Furnstahl, G. Hagen, and T. Papenbrock, Phys. Rev. C **87**, 044326 (2013).
[24] K. A. Wendt, C. Forssén, T. Papenbrock, and D. Sääf, Phys. Rev. C **91**, 061301(R) (2015).
[25] R. Roth and P. Navrátil, Phys. Rev. Lett. **99**, 092501 (2007); D. J. Dean, G. Hagen, M. Hjorth-Jensen, T. Papenbrock, and A. Schwenk, *ibid.* **101**, 119201 (2008); R. Roth and P. Navrátil, *ibid.* **101**, 119202 (2008).
[26] R. Roth, Phys. Rev. C **79**, 064324 (2009).
[27] T. Otsuka, M. Honma, T. Mizusaki, N. Shimizu, and Y. Utsuno, Prog. Part. Nucl. Phys. **47**, 319 (2001).
[28] T. Otsuka, Nucl. Phys. A **693**, 383 (2001).
[29] R. Roth, J. Langhammer, A. Calci, S. Binder, and P. Navrátil, Phys. Rev. Lett. **107**, 072501 (2011).
[30] C. Stumpf, J. Braun, and R. Roth, Phys. Rev. C **93**, 021301(R) (2016).
[31] C. Stumpf, T. Wolfgruber, and R. Roth, arXiv:1709.06840.
[32] T. Abe, P. Maris, T. Otsuka, N. Shimizu, Y. Utsuno, and J. P. Vary, Phys. Rev. C **86**, 054301 (2012).
[33] L. Liu, T. Otsuka, N. Shimizu, Y. Utsuno, and R. Roth, Phys. Rev. C **86**, 014302 (2012).
[34] T. Togashi, Y. Tsunoda, T. Otsuka, and N. Shimizu, Phys. Rev. Lett. **117**, 172502 (2016).

[35] G. A. Negoita, J. P. Vary, G. R. Luecke, P. Maris, A. M. Shirokov, I. J. Shin, Y. Kim, E. G. Ng, C. Yang, M. Lockner, and G. M. Prabhu, Phys. Rev. C **99**, 054308 (2019).

[36] W. G. Jiang, G. Hagen, and T. Papenbrock, Phys. Rev. C **100**, 054326 (2019).

[37] The code is available on GitHub https://github.com/SotaYoshida/BCGP_ECI.

[38] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)* (The MIT Press, Cambridge, 2005).

[39] L. Neufcourt, Y. Cao, W. Nazarewicz, and F. Viens, Phys. Rev. C **98**, 034318 (2018).

[40] R. A. Vargas-Hernández, J. Sous, M. Berciu, and R. V. Krems, Phys. Rev. Lett. **121**, 255702 (2018).

[41] S. Das, G. Siopsis, and C. Weedbrook, Phys. Rev. A **97**, 022315 (2018).

[42] D. Wee, J. Kim, S. Bang, G. Samsonidze, and B. Kozinsky, Phys. Rev. Mater. **3**, 033803 (2019).

[43] L. Neufcourt, Y. Cao, W. Nazarewicz, E. Olsen, and F. Viens, Phys. Rev. Lett. **122**, 062502 (2019).

[44] S. Zhou, P. Giuliani, J. Piekarewicz, A. Bhattacharya, and D. Pati, Phys. Rev. C **99**, 055202 (2019).

[45] J. A. Melendez, R. J. Furnstahl, D. R. Phillips, M. T. Pratola, and S. Wesolowski, Phys. Rev. C **100**, 044001 (2019).

[46] R. M. Neal, Bayesian learning for neural networks, Ph.D. thesis, University of Toronto, 1995.

[47] J. Lee, Y. Bahri, R. Novak, S. S. Schoenholz, J. Pennington, and J. Sohl-Dickstein, arXiv:1711.00165.

[48] N. Gordon, D. Salmond, and A. Smith, IEEE Proc. F, Radar and Signal Processing **140**, 107 (1993).

[49] I. J. Shin, Y. Kim, P. Maris, J. P. Vary, C. Forssén, J. Rotureau, and N. Michel, J. Phys. G: Nucl. Part. Phys. **44**, 075103 (2017).

[50] M. K. G. Kruse, E. D. Jurgenson, P. Navrátil, B. R. Barrett, and W. E. Ormand, Phys. Rev. C **87**, 044301 (2013).

[51] P.-G. Reinhard and W. Nazarewicz, Phys. Rev. C **81**, 051303(R) (2010).

[52] F. J. Fattoyev and J. Piekarewicz, Phys. Rev. C **84**, 064302 (2011).

[53] M. Kortelainen, J. Erler, W. Nazarewicz, N. Birge, Y. Gao, and E. Olsen, Phys. Rev. C **88**, 031305(R) (2013).

[54] Y. Gao, J. Dobaczewski, M. Kortelainen, J. Toivanen, and D. Tarpanov, Phys. Rev. C **87**, 034324 (2013).

[55] J. Dobaczewski, W. Nazarewicz, and P.-G. Reinhard, J. Phys. G: Nucl. Part. Phys. **41**, 074001 (2014).

[56] S. Goriely and R. Capote, Phys. Rev. C **89**, 054318 (2014).

[57] J. D. McDonnell, N. Schunck, D. Higdon, J. Sarich, S. M. Wild, and W. Nazarewicz, Phys. Rev. Lett. **114**, 122501 (2015).

[58] R. N. Pérez, J. E. Amaro, and E. R. Arriola, Phys. Rev. C **88**, 064002 (2013); **91**, 029901(E) (2015).

[59] R. Navarro Pérez, J. E. Amaro, and E. Ruiz Arriola, Phys. Rev. C **89**, 064006 (2014).

[60] S. E. Agbemava, A. V. Afanasjev, D. Ray, and P. Ring, Phys. Rev. C **89**, 054320 (2014).

[61] A. Ekström, B. D. Carlsson, K. A. Wendt, C. Forssén, M. H. Jensen, R. Machleidt, and S. M. Wild, J. Phys. G: Nucl. Part. Phys. **42**, 034003 (2015).

[62] B. D. Carlsson, A. Ekström, C. Forssén, D. F. Strömberg, G. R. Jansen, O. Lilja, M. Lindby, B. A. Mattsson, and K. A. Wendt, Phys. Rev. X **6**, 011019 (2016).

[63] A. E. Lovell, F. M. Nunes, J. Sarich, and S. M. Wild, Phys. Rev. C **95**, 024611 (2017).

[64] Y. Jaganathen, R. M. I. Betan, N. Michel, W. Nazarewicz, and M. Płoszajczak, Phys. Rev. C **96**, 054316 (2017).

[65] S. Yoshida, N. Shimizu, T. Togashi, and T. Otsuka, Phys. Rev. C **98**, 061301(R) (2018).

[66] D. Frame, R. He, I. Ipsen, D. Lee, D. Lee, and E. Rrapaj, Phys. Rev. Lett. **121**, 032501 (2018).

[67] S. König, A. Ekström, K. Hebeler, D. Lee, and A. Schwenk, arXiv:1909.08446.

[68] A. Ekström and G. Hagen, Phys. Rev. Lett. **123**, 252501 (2019).

[69] Z. Noumir, P. Honeine, and C. Richard, in *2012 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)* (IEEE, 2012), pp. 1–6.

[70] M. L. Stein, *Interpolation of Spatial Data*, Springer Series in Statistics (Springer-Verlag, New York, 1999); See also Sec. 4.2.1 of Ref. [38].

[71] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd ed. (Society for Industrial and Applied Mathematics, Philadelphia, 2002).

[72] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, J. Mach. Learn. Res. **12**, 2825 (2011).

[73] J. Fairbrother, C. Nemeth, M. Rischard, J. Brea, and T. Pinder, arXiv:1812.09064.

[74] Detailed analyses on positive semi-definiteness of covariance matrices https://github.com/SotaYoshida/IssueOnGPs.