Non-parametric Bayesian approach to extrapolation problems in configuration interaction methods

Sota Yoshida$^{1,2,*}$

$^1$Department of Physics, the University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
$^2$Liberal and General Education Center, Institute for Promotion of Higher Academic Education, Utsunomiya University, Mine, Utsunomiya, 321-8505, Japan

(Dated: March 16, 2020)

We propose a non-parametric extrapolation method based on constrained Gaussian processes for configuration interaction methods. Our 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. In the present study, we show an application to the extrapolation needed in full configuration interaction method as an example.

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 sub-space, and the truncation is typically specified by the parameter $N_{\text{max}}$ 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 $N_{\text{max}}$ for upper $p$-shell nuclei is around 10 (see e.g. [19]) and this is still far from $N_{\text{max}} = \infty$ corresponding to exact calculations. One usually extrapolates the sequence of results with different $N_{\text{max}}$ to $N_{\text{max}} = \infty$ 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 [29–34].

In these studies, extrapolation is performed with some specific functions such as exponential or polynomials, and its coefficients are determined so as to minimize the $\chi^2$ 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 $\chi^2$ minimization of a parametric model and point estimation of the parameter 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 like a level ordering of states with small energy differences, the positions of proton and neutron drip lines, and so on.

In the present study, we propose a novel non-parametric extrapolation method for CI calculations using constrained Gaussian processes (GPs), which can give 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 us to consider a wider class of functions, and then it is expected to alleviate underestimating extrapolation uncertainties due to the specific choice of the extrapolation function.

We also note 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 $ab$ initio calculations while varying their inputs such as $N_{\text{max}}$ 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.

We demonstrate the validity of our model by taking the ground state energies obtained by 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 non-parametric 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 we need 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. We refer the interested reader to e.g. [38, 46, 47].

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

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 $N(\mu, \Sigma)$ whose covariance matrix $\Sigma$ is given as

$$
\Sigma = \begin{bmatrix}
K_{XX} & K_{XX^*} \\
K_{X^*X} & K_{X^*X^*}
\end{bmatrix}, \quad (1)
$$

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, we use the logMat52 kernel function 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:

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

where $r \equiv |x_i - x_j|$, and the global strength $\tau$ and correlation length $\ell$ are the hyperparameters. Let $\theta$ denote the vector of hyperparameters. We will revisit the issue of hyperparameters later.

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

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

It is a common practice for mean vectors to be normalized, i.e., $\mu$ and $\mu^*$ 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

$$
P(y, y^*|\theta) = P(y^*|y, \theta)P(y|\theta). \quad (4)
$$

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

$$
\begin{align*}
P(y^*|y, \theta) &= N(\mu_{y^*|y}, \Sigma_{y^*|y}), \quad (5) \\
\mu_{y^*|y}(\theta) &= \mu^* + K^T_{XX^*}K_{XX}^{-1}(y - \mu), \quad (6) \\
\Sigma_{y^*|y}(\theta) &= K_{XX^*} - K^T_{XX^*}K_{XX}^{-1}K_{XX^*}, \quad (7) \\
P(y|\theta) &= N(\mu, K_{XX}). \quad (8)
\end{align*}
$$

Regarding the hyperparameters, the so-called maximum a posteriori (MAP), i.e., one to maximize the hyperparameter posterior $P(\theta|y)$, is often used. However, we do not use a single value for the hyperparameters to receive benefit of GPs; various hyperparameters gives us a much wider class of sample functions. We do inference of their probability distributions by a sampling scheme to integrate out the hyperparameter dependence. In this case, the posterior distribution of $y^*$ for unobserved input $x^*$ is written as

$$
P(y^*|y) \propto \int P(y^*|y, \theta)P(\theta|y)P(\theta)d\theta. \quad (9)
$$

In addition to this, we extend this formulation 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 $N_{\text{max}}$. In general, the accuracy of a statistical model like GP is improved by including such physics information. To this end, we extend Eq. (9) to

$$
P(y^*|\alpha, \beta, \ldots) \propto \int P(y^*|y, \theta)P(\theta)P(\theta) \times \cdots \times P(\alpha, \beta, \ldots | y^*, y)d\theta, \quad (10)
$$

where $P(\alpha, \beta, \ldots | y^*, y)$ is the probability that the constraints $\alpha, \beta, \ldots$ are satisfied under the given $y^*$ and $y$. 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 $\theta$ is independent of the constraints (see Appendix B). This is 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. We evaluate the integration in Eq. (10) by weighted \( N_p \) samples as follows:

\[
P(y^* | y, \alpha, \beta, \ldots) \simeq \sum_{i=1}^{N_p} w(i) P(y^*(i) | y, \theta(i)),
\]

\[
w(i) = \frac{P(y, \theta(i)) P(\alpha, \beta, \ldots | y, \theta(i))}{\sum_{j=1}^{N_p} P(y, \theta(j)) P(\alpha, \beta, \ldots | y, \theta(j))}.
\]

We employ the particle filtering method [48] (also known as Sequential Monte Carlo) as a sampling scheme to evaluate the summation in Eq. (11). In our particle filtering algorithm, \( \{\theta(i), y^*(i)\} \) are assigned to particles labeled by \( i = 1, 2, ..., N_p \), 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 which do not respect the physics constraints are discarded.

III. PROBLEMS OF INTEREST

A. FCI results

In what follows, we apply the constrained GP model to extrapolation problems in FCI calculations. We analyze published FCI results of the ground state energy of \(^6\)Li using JISP16/NNLO\textsubscript{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 \( N_{\text{max}} \).

Let \( \{(x_1, y_1), (x_2, y_2), ..., (x_D, y_D)|x_1 < x_2 < \cdots < x_D\} \) denote the data, i.e. \( (x_1, y_1) = (6, \pm 28.602), ..., (x_D, y_D) = (14, \pm 31.977) \) 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 we use all \( N_{\text{max}} \) results as data unless otherwise mentioned.

The extrapolation problem addressed below is to estimate the ground state energies at \( N_{\text{max}} \) larger than \( x_D \) and we express them as \( \{(x_1^*, y_1^*), (x_2^*, y_2^*), ..., (x_P^*, y_P^*) | x_1^* < x_2^* < \cdots < x_P^*\} \); Here \( x_1^* = x_D + 2 \) and \( P \) is large even integer. For the sake of simplicity, we restrict ourselves to consider the ground state of \(^6\)Li with natural parity, i.e. we only consider even \( N_{\text{max}} \). In practice, we truncate at certain finite \( P \) value where predictions are converged with respect to \( N_{\text{max}} \). A detailed discussion about this \( P \) will be given later.

B. Constraints on extrapolation function

As minimal constraints on the extrapolation function to capture the asymptotic behavior of FCI results, we impose the following two constraints, \( \alpha \) and \( \beta \).

![FIG. 1. The FCI results of g.s. energy of \(^6\)Li using JISP16/NNLO\textsubscript{opt} [49] and N3LO [50].](image.png)

![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, NNLO\textsubscript{opt} and N3LO results are slightly shifted to the left and right, respectively. See the text for more details.](image.png)

The constraint \( \alpha \) is variational property, i.e. the monotonicity of energy eigenvalues with respect to \( N_{\text{max}} \):

\[
P(\alpha | y^*, y) = \Phi(y_D - y_1^*) \times \Phi(y_1^* - y_2^*) \times \cdots \times \Phi(y_{P-1}^* - y_P^*).
\]

Here we introduced the Probit function

\[
\Phi(z) \equiv \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{t^2}{2} \right) dt,
\]

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

The second constraint \( \beta \) is about the convergence pattern. We use a ratio of the absolute value of energy gains as a measure of convergence of g.s. energy in FCI. This ratio \( r \) at
where in Eqs. (3)–(8): GP extrapolation model using constraints as the constrained agree with ones with only the constraint on the convergence pattern. When be expected from FIG. 2, this is a rather moderate constraint We use

\[ r(x_j) = \frac{y_j - y_{j+1}}{y_{j-1} - y_j}, \quad (15) \]

We plot \( \{r(x_j)\} \) for the given data 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. This \( r \) can be extended to include predictions \( \{y^*_j\} \):

\[ r(x'_1) = \left| \frac{y^*_1 - y^*_2}{y^*_D - y^*_1} \right|, \quad r(x'_2) = \left| \frac{y^*_2 - y^*_3}{y^*_1 - y^*_2} \right|, \ldots, \]

\[ r(x'_{p-1}) = \left| \frac{y^*_{p-1} - y^*_p}{y^*_{p-2} - y^*_{p-1}} \right|. \quad (16) \]

We impose the constraint on \( \{r\} \) as follows:

\[ P(\beta|y^*, y) = \Phi \left( R_E - r(x'_1) \right) \times \Phi \left( R_E - r(x'_2) \right) \times \ldots \]

\[ \ldots \times \Phi \left( R_E - r(x'_{p-1}) \right), \quad (17) \]

where \( R_E \) is an upper threshold determined as follows:

\[ R_E = r_{\text{mean}} + r_{\text{std}}, \quad (18) \]

\[ r_{\text{mean}} = \frac{y_D - y_1}{y_D - y_{D-1}}, \quad (19) \]

\[ r_{\text{std}} = \sigma_r r_{\text{mean}}. \quad (20) \]

We use \( \sigma_r = 0.1 \) throughout this work for simplicity. As can be expected from FIG. 2, this is a rather moderate constraint on the convergence pattern. When \( \sigma_r \) is large enough, results agree with ones with only the constraint \( \alpha \). We refer to the GP extrapolation model using constraints as the constrained Gaussian process (cGP) model.

### C. Choice of the mean function

Here we introduce two different mean functions \( \mu^{(s)} \) needed in Eqs. (3)–(8):

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

We refer to these as cGP-a and cGP-b, respectively and analyze both cases below.

## IV. RESULTS AND DISCUSSIONS

### A. Extrapolation of g.s. energies

We now present the results of the cGP predictions for the ground-state energies of \(^6\text{Li}\) in FIG. 3. The extrapolated values at a certain \( N_{\text{max}} \) are shown by 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.

We note that the prediction is truncated up to a certain \( N_{\text{max}} \) where the mean value is converged within 0.2 keV. It means that possible differences between predictions at the finite \( N_{\text{max}} \) and one at \( N_{\text{max}} = \infty \) are suppressed below 1 keV because of the constraint \( \beta \). The \( N_{\text{max}} \) giving converged results are 44, 42 and 34 for NNLO\(_\text{opt}\), JISP16, and N3LO, respectively. These numbers are consistent with the intuition that harder interaction requires larger \( N_{\text{max}} \) 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 manuscript, the 68\% interval is determined from the 16th and 84th percentile of the distribution, and the 99\% 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 $N_{\text{max}}$ data. The conventional B3 extrapolation is always included as a special case of two cGP results. The values for Literature are from Ref. [49, 50]. Here we note that the literature value for N3LO [50] might be obtained by an exponential fit using all five data, which is not explicitly stated. It must also be noted that it is a highly non-trivial task to fairly compare the results with different extrapolation techniques, because data is truncated in parametric models and some use data with multiple $h\Omega$ as in the A5 extrapolation [49].

We note that our sampling scheme with the particle filtering gives converged results within a few keV in case of 20,000 particles after 2,000 times Metropolis-Hastings updates for each particle, and that independent runs reproduce the same results within the Monte Carlo error.

B. Impact of the constraints

Here we discuss 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 $\alpha$ and $\beta$ (red), with the only $\alpha$ (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. For this reason, we omit the case of cGP-a without $\alpha$ and $\beta$, which is obviously not appropriate for the current purpose. When one assumes that the wave function is dominated by relatively lower $N_{\text{max}}$ configurations, predictions with both constraints $\alpha$ and $\beta$ are expected to be more reliable than the others.

C. Data dependence

We have used all $N_{\text{max}}$ results as data so far. Here we explore the dependence of extrapolated values on the used data to test the potential predictive power of our extrapolation method. In FIG. 5, extrapolated values for both cGP-a and cGP-b are shown as a function of the maximum $N_{\text{max}}$ used as data, i.e. $x_D = \max(N_{\text{max}})$.

The mean predictions by the cGP models are shown by horizontal line in white, and the shaded areas show the posterior distributions obtained by 20,000 particles. The 68% and 99% credible intervals, respectively, are shown by the thick and thin vertical lines, respectively. We note that the width of the shaded area is scaled to be the same for each area, and NNLOopt (N3LO) results are slightly shifted to left (right) for visibility.

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 $N_{\text{max}}$ data is added with only one exception, which is the cGP-b result for N3LO with $\max(N_{\text{max}}) = 10$. This exception can be understood from Eqs. (4)–(8). In this case, the exponential function exactly fits the given three data and then $\mu_y |y \rangle$ in Eq. (6) is identical with $\mu^*$. Any fluctuation of the joint mean value $\mu_y |y \rangle$ 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 $N_{\text{max}}$ forever, whereas the point to stop must depend on the desired accuracy.

In the rest of this subsection, let us regard the mean values at rightmost $\max(N_{\text{max}})$ in FIG. 5, i.e. $\max(N_{\text{max}}) = 18$ for NNLOopt and JISP16, and $\max(N_{\text{max}}) = 14$ for N3LO, as the (temporary) exact values.

For NNLOopt and JISP16, an important remark is that the exact values are covered by the cGP-a predictions with a relatively lower $\max(N_{\text{max}})$. The sign of convergence can be seen around $\max(N_{\text{max}}) = 10 - 12$, i.e. one can choose these as points to stop the calculation. From the behavior of the credible
V. SUMMARY AND OUTLOOK

We introduced an extrapolation method for CI-type calculations using constrained Gaussian processes. This method has the following advantages that are required for future generations of \textit{ab initio} studies to make more quantitative discussions on observables of interest and on the quality of adopted nuclear interactions.

Firstly, our extrapolation method does not need to remove outliers and has applicability to sparse data sets, which are strongly needed for future FCI calculations. Secondly, one can naturally incorporate domain knowledge into the model. It is often the case especially in physics that one knows in advance 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. Thirdly, 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. [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 \textit{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 \textit{ab initio} methods combined with an extrapolation method with quantified uncertainties.

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

In this work, we discussed 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 a multi-dimensional space is rather straightforward, while it is expected that one needs more technical analyses in numerical studies such as positive semi-definiteness of covariance matrices. Our model can also be applied to valence CI techniques using an importance-truncation in which extrapolation function is much more non-trivial 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 (Grants No. 17J06775).

[1] R. Machleidt and D. Entem, Phys. Rept. 503, 1 (2011).
[2] E. Epelbaum, H.-W. Hammer, and U.-G. Meiβner, 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. Navrátíl, S. Quaglioni, G. Hupin, C. Romero-Redondo, and A. Calci, Physica Scripta 91, 053002 (2016).
[5] G. Hagen, T. Papenbrock, M. Hjorth-Jensen, and D. J. Dean, Reports on Progress in Physics 77, 096302 (2014).
[6] H. Hergert, S. Bogner, T. Morris, A. Schwenk, and K. Tsukiya, Physics Reports 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. Hebler, 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, Nature Physics 15, 428 (2019).
[12] B. R. Barrett, P. Navrátil, and J. P. Vary, Progress in Particle and Nuclear Physics 69, 131 (2013).
[13] P. Navrátil, S. Quaglioni, I. Stetcu, and B. R. Barrett, Journal of Physics G: Nuclear and Particle Physics 36, 083101 (2009).
[14] E. Caurier, G. Martinez-Pinedo, F. Nowacki, A. Poves, and A. P. Zuker, Rev. Mod. Phys. 77, 427 (2005).
[15] J. P. Vary, “The many-fermion dynamics shell-model code,” unpublished (1992); J. P. Vary and D. C. Zheng, “The many-fermion dynamics shell-model code;”, 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, NJ, USA, 2008) pp. 15:1–15:12.
[16] N. Shimizu, “Nuclear shell-model code for massive parallel computation, “kshell”,” (2013), arXiv:1310.5431; N. Shimizu, T. Mizusaki, Y. Utsuno, and Y. Tsunoda, Computer Physics Communications 244, 372 (2019).
[17] C. W. Johnson, W. E. Ormand, K. S. McElvain, and H. Shan, “Bigstick: A flexible configuration-interaction shell-model code,” (2018), 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, 013102 (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 (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, Phys. Rev. Lett. 101, 119201 (2008); R. Roth and P. Navrátil, Phys. Rev. Lett. 101, 119202 (2008).
[26] R. Roth, Phys. Rev. C 79, 064324 (2009).
[27] T. Otsuka, M. Honma, T. Mizusaki, N. Shimizu, and Y. Utsuno, Progress in Particle and Nuclear Physics 47, 319 (2001).
[28] T. Otsuka, Nuclear Physics A 693, 383 (2001), radioactive Nuclear Beams.
[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 (2016).
[31] C. Stumpf, T. Wolfgruber, and R. Roth, “Electromagnetic strength distributions from the ab initio no-core shell model,” (2017), 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. Otsuka, Y. Tsunoda, T. Otsuka, and N. Shimizu, Phys. Rev. C 86, 061301 (2015).
[35] G. A. Negotia, J. P. Vary, G. R. Lukecke, 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/BGP_ECI.
where $\Gamma$ is the gamma function and $K_\nu$ is the modified Bessel function of the second kind. Especially, Matérn kernel 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

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

and Matérn kernel:

$$k_{\text{Matern}}(x_i, x_j; \nu) = \tau \frac{\sqrt{2\nu}}{\Gamma(\nu)} \xi^{\nu} K_\nu(\xi), \quad (A2)$$

where $\Gamma$ is the gamma function and $K_\nu$ is the modified Bessel function of the second kind.
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 our purpose, i.e. extrapolation of FCI results, but we include $\nu = 3/2$ case too.

In what follows, we 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 functions.

The covariances due to rounding errors. We briefly demonstrate that in this too smooth nature sometimes breaks down the positive semidefiniteness of covariant matrices in numerical calculations.

The (log)RBF kernel. The RBF is one of the most popular functions is often regarded as too high [70], and, in practice, the $\log$RBF kernel gives non-PSD matrices easier than logMat52 by one or two orders of magnitude.

We calculated the conditional mean vectors and covariance matrices while varying the hyperparameters $\tau$ and $\ell$ in Eq. (A4)–Eq. (A6). In Fig. 7, the positive semi-definiteness of the covariance matrices are summarized. The diamond symbols (red) correspond to the case that both $K_{XX}$ and $\Sigma^\dagger_y \Sigma_y = 0$. The cross symbols (blue) are assigned if only the $\Sigma^\dagger_y \Sigma_y$ is non-PSD. In terms of the length scale $\ell$, the logRBF gives non-PSD matrices easier than logMat52 by one or two orders of magnitude.

Now we show the effects of the $\epsilon$ prescription on the predictions. In Fig. 8, the following quantity is shown:

$$\Delta \equiv \log_{10} \left( \max(\|\mu^\text{own}_y|y\| - \|\mu^\text{lib}_y|y\|) \right),$$

where the superscripts own and lib. mean the conditional mean vector calculated by our own code and one using the Gaussian-Processes.jl library [71], respectively. For the latter, we fixed the size of observation noise as zero and allowed to use the default $\epsilon$ prescription implemented in the library. We note that both $\mu^\text{own}$ and $\mu^\text{lib}$ are now given in a unit of MeV. The hatched regions with dots for smaller $\ell$ means that the results of the two codes are exactly the same. At the points shown by the diamond symbol in Fig. 7, our code without the $\epsilon$ prescription cannot give the mean vectors. Those correspond to the white region appeared in FIG. 8. When we increase the length scale $\ell$ at which the posterior covariance becomes non-PSD, i.e. points shown by the cross symbol in Fig. 7, the $\Delta$ becomes larger because of the immune system in the library. In some cases, the deviations in mean vectors reach a few MeV, which are obviously non-negligible.

One should pay much attention to how the PSD and the observation noise are treated in the codes and whether or not one really can neglect the impact of the $\epsilon$ prescriptions on the predictions, especially when one would like to integrate out the hyperparameter dependence. For more detailed analyses and codes to reproduce the Fig. 7 and FIG. 8, we refer the reader to Ref. [72].

From the analyses above, we concluded that the logMat52 is the most appropriate choice for our extrapolation method.
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.

FIG. 8. The plot showing the impact of the $\epsilon$ prescription to avoid non-PSD. The colormap shows the $\Delta$ in Eq. (A8). The hatched region with dots show the points at which the two codes gives the exactly the same result. The region where our code does not give answer is colored in white.

Appendix B: Extension of the posterior distribution under the constraints

Eq. (10) is derived in the followings. Let the $c$ denote the physics constraints to be imposed.

$$P(y^*, y, c) = \int d\theta P(y^*, y, c, \theta), \quad (B1)$$

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

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

$$\propto \int d\theta P(c|y^*, y)P(y^*|y, \theta)P(y|\theta)P(\theta). \quad (B4)$$

Appendix C: Contributions to the posterior distribution

In Figures 9–12, the hyperparameter distributions are shown. Colors of each dot point show the relative size of contributions to the likelihood $P(\theta|y)$ and to the integral in Eq. (10). In our sampling method, particles are evolved according to the random walk Metropolis-Hastings method. We can see from FIG. 10 and FIG. 12 that particles distribute like an ellipse in hyperparameter space. The particles do not distribute over the non-PSD region discussed in Sec. B.
FIG. 9. Relative contributions to likelihood Eq. (8) for cGP-a.

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

FIG. 11. Relative contributions to likelihood Eq. (8) for cGP-b.

FIG. 12. Projection plot of the posterior to hyperparameter space. The colors show relative contributions to Eq. (10) in cGP-b.