BAYESIAN UNCERTAINTY QUANTIFICATION FOR THE CALCULATION OF ALLOY PROPERTIES WITH THE CLUSTER EXPANSION METHOD

Nicholas Zabaras
School of Engineering
The University of Warwick
Coventry CV4 7AL
http://www.zabaras.com/

Joint Work with: Jesper Kristensen, Cornell University

Motivation
Motivation

- **Interest**
  - Thermodynamic computations with quantum mechanical energies:
    - Compute phase diagrams
  - Optimization of properties:
    - Minimize thermal conductivity
  - Etc.

- **Requirement**
  - Traverse phase space with, e.g., Monte Carlo methods
  - Million+ property (e.g., energy) evaluations

- **Problem**
  - Accurate (e.g., ab initio) method not feasible
  - Too expensive (takes too long!)

- **Solution**
  - Sacrifice accuracy for speed:
    - Surrogate model based on few ab initio simulations
Motivation

- **Example of expensive computer codes**
  - Quantum mechanical energies obtained via VASP [1]
  - Heat conductivity from LAMMPS [2]

- **Example of building a surrogate model**
  - Approximate computer code with parametrized surface
  - E.g., expand response surface in a basis

- **Predicting property of interest**
  - Predictions (such as phase transitions) are made based on surrogates

---

**Motivation**

- **Important problem not addressed so far:**
  - What is the induced uncertainty in predicted quantities from using particular parametrizations of surrogates?
  - Can help answer whether surrogate can *even be used* to capture response surface!

- **In context of designing binary alloys with surrogates we ask:**
  - Why does the surrogates predict so well?
  - Can the surrogates be used for to design new materials?
  - *Nobody has answered these central questions!*

- We use Bayesian probability theory to answer these questions

---

J. Kristensen and N. Zabaras, Computer Physics Communications 185 (2014), pp. 2885-2892
Quantum Mechanical Alloy Energies
Focus of this Work

- Quantum mechanical computer code responds with $f(x)$ on seeing input structure $x$

- Computes energy of $x$ from many-electron Schrodinger equation:

$$\hat{H}\Psi = \left[\hat{T} + \hat{V} + \hat{U}\right]\Psi = \left[\sum_{i}^{N} \left(-\frac{\hbar^2}{2m_i} \nabla_i^2\right) + \sum_{i}^{N} V(\vec{r}_i) + \sum_{i<j}^{N} U(\vec{r}_i, \vec{r}_j)\right]\Psi = E\Psi$$

- Time consuming (expensive), so:
  - Parametrize computer code:
    - $f(x; \theta)$ (surrogate model)

- How does the possible choices of $\theta$ (surrogates), consistent with observed data, affect our predictions?
It is surprising these surrogates work well in the context of alloy modeling:
  ➢ Very limited data is observed
  ➢ Very sparse surrogates are (typically) used

What is the affect of the surrogate and the limited data on predictions?
  ➢ In alloy modeling: Nobody knows!
  ➢ Despite extensive use of surrogates [1-5]

Here we answer this important question!

Surrogate Models: Bayesian Probability Theory
Isolate uncertainty arising from parametrization:
The true response surface belongs to the parametrization class

- \( f(.,\theta) \) being periodic, satisfying certain symmetry conditions, differentiability, etc.
- This is the case in alloy modeling

Uncertainty in \( \theta \) propagates through the response surface to predicted quantities

- Model this uncertainty in a Bayesian way
- Prior belief on parametrization: \( p(\theta) \)

- Belief on prediction then becomes*:

\[
p(I) = \int d\theta \delta(I[f(\cdot; \theta)] - I)p(\theta)
\]

E.g., phase transition temperature

Run the computer code and obtain the data set:

\[ D = \{ x_i, f(x_i) \} \]

The likelihood contains all information about \( \theta \) contained in the data \( D \)

\[ \mathcal{L}(D|\theta, \cdot) \]

Posterior = Likelihood x Prior (up to a constant)

\[ p(\theta|D) \propto \mathcal{L}(D|\theta, \cdot)p(\theta) \]

This is how our belief about the parametrizations changes on seeing \( D \)
Quantify Uncertainty: Algorithm

- Update our belief about the prediction to:

\[ p(I|\mathcal{D}) = \int d\theta \delta(I[f(\cdot; \theta)] - I)p(\theta|\mathcal{D}) \]

- Quantify uncertainty in predictions:
  - I.e., draw samples from \( p(I|D) \)

1) Collect (expensive) data set \( \mathcal{D} = \{x_j\}_{j=1}^N \).

2) Collect parametrizations of the surrogate \( \mathcal{T} = \{\theta_i\}_{i=1}^M \) from (some) Bayesian posterior.

3) for \( i \) in 1...\( M \):
   - Compute the quantity of interest (QI) resulting from the \( i \)th parametrization,
   - call this QI the \( i \)th QI: \( I_i[f(\cdot; \theta_i)] \), or simply \( I_i \).

4) Compute and output the statistics of \( \{I_i\}_{i=1}^M \).
Case Studies: Phase Diagrams of Real Binary Alloys
First, find a parametrization of response surface where:

- True surface is included in parametrization class

Then we employ the uncertainty quantification algorithm:

1) Collect (expensive) data set \( \mathcal{D} = \{x_j\}_{j=1}^N \).

2) Collect parametrizations of the surrogate \( \mathcal{T} = \{\theta_i\}_{i=1}^M \) from (some) Bayesian posterior.

3) **for** \( i \) **in** 1...\( M \):
   
   Compute the quantity of interest (QI) resulting from the \( i \)th parametrization,
   
   call this QI the \( i \)th QI: \( I_i[f(\cdot; \theta_i)] \), or simply \( I_i \).

4) Compute and output the statistics of \( \{I_i\}_{i=1}^M \).
PARAMETRIZE
RESPONSE SURFACE OF
QUANTUM MECHANICAL
ENERGIES
Response surface is Quantum Mechanical Energies from VASP

- Code itself has modeling errors:
  - $K$-space discretization [1]
  - Exchange-correlation term [2]
  - Pseudopotential (e.g., PAW-PBE) [3]
  - Etc.

Ignore computer code errors:
- Assume we observe the true response surface with added Gaussian noise describing the error
- Note: This does not (necessarily) mean the error itself is Gaussian but that we only assume our estimates to depend on the first and second moments of the error distributions

\[ V_s(\vec{r}) = V(\vec{r}) + \int \frac{e^{2n_s(\vec{r}')}}{|\vec{r} - \vec{r}'|} d^3r' + V_{XC}[n_s(\vec{r})] \]

PARAMETRIZE RESPONSE SURFACE

- Replace VASP response surface with surrogate model
  - What should the surrogate model be?

- Quantum Mechanical energy invariant under space group operations of the lattice
  - Surrogate must account for this

- Possible (and popular) surrogate in alloy modeling:
  - The Cluster Expansion*
    - Expansion coefficients (ECI)
    \[ E^{(i)} = \sum_{F} J_F \langle \Pi_F \rangle_{\sigma^{(i)}} \]
    - Correlation functions (Basis)
    - Accounts for symmetries

Corresponds to: Multidimensional discrete Fourier transform*
Generalized Ising model (map atoms to integers)
For binary alloys correlation functions reduce to simple products of atoms \( \sigma \) on atomic sites:

\[
E^{(t)} = J_0 + J_1 \sum_i \sigma_i^{(t)} + J_{2,1} \sum_{i,j} \sigma_i^{(t)} \sigma_j^{(t)} + \cdots + J_{3,1} \sum_{i,j,k} \sigma_i^{(t)} \sigma_j^{(t)} \sigma_k^{(t)} + \cdots
\]

(Not general in this form)

The cluster expansion will be the surrogate of choice

The question is now
- How do we draw samples from it as required by the algorithm?

1) Collect (expensive) data set \( \mathcal{D} = \{x_j\}_{j=1}^N \).

2) Collect parametrizations of the surrogate \( \mathcal{T} = \{\theta_i\}_{i=1}^M \) from (some) Bayesian posterior.

3) \textbf{for } i \textbf{ in } 1...M:\n   - Compute the quantity of interest (QI) resulting from the \( i \)th parametrization,
   - call this QI the \( i \)th QI: \( I_i[f(\cdot; \theta_i)] \), or simply \( I_i \).

4) Compute and output the statistics of \( \{I_i\}_{i=1}^M \).
Collect Bayesian Parametrizations of Surrogate
Infinite series: True response surface is part of parametrization class

Truncation required:

- How do we choose truncated parametrization?

\[ E^{(t)} = J_0 + J_1 \sum_i \sigma_i^{(t)} + J_{2,1} \sum_{i,j} \sigma_i^{(t)} \sigma_j^{(t)} + \cdots + J_{3,1} \sum_{i,j,k} \sigma_i^{(t)} \sigma_j^{(t)} \sigma_k^{(t)} + \cdots \]

\[ \gamma = (1, 1, 0, \cdots, 1, \cdots) \]

\[ \beta_{\gamma} = (J_0, J_1, 0, \cdots, J_{3,1}, \cdots) \]

\[ \theta = (\gamma, \beta_{\gamma}) \]

\[ E^{(i)} = \sum_{F} J_F \langle \Pi_F \rangle_{\sigma(i)} \rightarrow y = X \beta_{\gamma} + \epsilon \]
Fully Bayesian treatment of $\theta=(\gamma, \beta)$ based on LASSO:

**Priors:**
\[
p(\beta_{\gamma,j} | \tau, \gamma) = \frac{1}{2\tau} \exp \left( -\frac{||\beta_{\gamma,j}||}{\tau} \right)
\]
\[
p(k | \lambda) = \frac{\exp(-\lambda)\lambda^k}{C k!}
\]

**Hyperpriors:**
\[
p(\lambda) \propto \lambda^{-1}
\]
\[
p(\tau, \delta^2) \propto (\tau\delta)^{-1}
\]

Likelihood of data set with $N$ points (and Gaussian error):
\[
\left( \delta^2 \right)^{-N/2} \exp \left( -\frac{||y - X\beta||^2}{2\delta^2} \right)
\]

---

1. Least Absolute Shrinkage and Selection Operator Method, L$_1$ Constrained Least Squares Method
2. Bayesian Interpretation: Posterior Mode when the parameters to be learned have independent Laplace priors
Posterior = (Prior) x (Likelihood):

\[ p(\gamma, \beta_{\gamma} | \mathcal{D}) \propto (\lambda \tau \delta)^{-1} \exp(-\lambda) \lambda^k \tau^{-k} \exp \left( -\frac{\sum |\beta_{\gamma,j}|}{\tau} \right) \left( \delta^2 \right)^{-N/2} \exp \left( -\frac{\|y - X \beta_{\gamma}\|_2^2}{2\delta^2} \right) \]

- **Prior**
- **Likelihood**

- The posterior encodes uncertainty from the lack of knowledge about the true parametrization.

- A sample from the posterior is one particular parametrization of \( f(\cdot) \).

- *Not in closed form! ---> MCMC methods*
Problem
  - How do we draw samples from this:
    - Posterior not in closed form

A solution
  - Use MCMC method

But
  - Standard MCMC works for a given truncation:
    It cannot tell us an optimal truncation as well as expansion coefficients

Solution
  - Use RJMCMC*
  - RJMCMC = Reversible Jump Markov Chain Monte Carlo
    - MCMC with dimension jumping scheme
      (e.g., birth-death or split-merge)

Example:
Propose incrementing the dimensionality of the sampling space

*Green, Peter J. *Biometrika* 82.4 (1995)
**RJMCMC Chain: Algorithm**

Input: The number of iterations $T$. Random walk step size $\varepsilon$.

Data: $X$ and $Y$.

Output: $\{\theta^t = (\beta^t, \gamma^t) | t \in \{0, \ldots, T\}\}$.

```
begin
  Initialization: set $\theta^0 = (\beta^{(0)}, \gamma^{(0)})$ and $t = 1$.
  repeat
    if $k^{(t-1)} = 1$ then
      $k^{(t)} \leftarrow k^{(t-1)} + U(0,1)$.
    else if $k^{(t-1)} = p$ then
      $k^{(t)} \leftarrow k^{(t-1)} - U(0,1)$.
    else
      $k^{(t)} \leftarrow k^{(t-1)} + U(-1,0,1)$.
  end
  Sample $s \sim N(0, \varepsilon^2)$.
  $K = \gamma^{(t-1)}$ and $K^c \leftarrow \{1, \ldots, p\} \setminus K$.
  if $k^{(t-1)} = k^{(t)}$ then
    Sample $j \sim U(K)$.
    Update $\beta_j^{(t)} \leftarrow \beta_j^{(t-1)} + s$ with an MH step, details in Section 3.2.
  else if $k^{(t)} = k^{(t-1)} + 1$ then
    Sample $j \sim U(K^c)$.
    Perform a "birth" move and update $\beta_j^{(t-1)}$, details in Section 3.3.
  else
    Sample $j \sim U(K)$.
    Perform a "death" move and update $\beta_j^{(t-1)}$, details in Section 3.3.
  end
  $t \leftarrow t + 1$.
until $t = T$.
```


**Standard Metropolis-Hastings**

$$
\min \left\{ \frac{\|\beta\|_1^{-k}}{\|\beta\|_1} \times \left( \frac{\|y - X\beta\|_2^{-(n-1)}}{\|y - X\beta\|_2} \right), 1 \right\}.
$$

**Update/birth/death?**

**Proposal**
Initial guess
- Parameter space modeled is very high dimensional and can be topologically very complex
- For this work we found the response surface to be smooth enough for consistent results
  - Multiple arbitrary starting guesses converged to similar results

Convergence
- Convergence can be slow
- We tested results by running an *extremely* long chain

For the current work
- Method is consistent and reliable
Collecting Bayesian parametrizations of the surrogate:

- Let RJMCMC chain run a long time
- 50% burn-in period
- Collect samples from the chain
- Each sample is a particular parametrization of the response surface agreeing with the observed data

1) Collect (expensive) data set $\mathcal{D} = \{x_j\}_{j=1}^N$.

2) Collect parametrizations of the surrogate $\mathcal{T} = \{\theta_i\}_{i=1}^M$ from (some) Bayesian posterior.

3) for $i$ in 1...$M$:
   - Compute the quantity of interest (QI) resulting from the $i$th parametrization,
   - call this QI the $i$th QI: $I_i[f(\cdot; \theta_i)]$, or simply $I_i$.

4) Compute and output the statistics of $\{I_i\}_{i=1}^M$. 
Data was collected from VASP via the method described in [*]
  ➢ ~1 week of computations

Two alloys:
  ➢ Mg-Li (bcc lattice)
    • ~80 structures
  ➢ Si-Ge (diamond lattice)
    • ~25 structures

Ignore non-configurational degrees of freedom

1) Collect (expensive) data set \( D = \{x_j\}_{j=1}^N \).

2) Collect parametrizations of the surrogate \( T = \{\theta_i\}_{i=1}^M \) from (some) Bayesian posterior.

3) for \( i \) in 1...\( M \):
  Compute the quantity of interest (QI)
  resulting from the \( i \)th parametrization,
  call this QI the \( i \)th QI: \( I_i[f(\cdot; \theta_i)] \), or simply \( I_i \).

4) Compute and output the statistics of \( \{I_i\}_{i=1}^M \).

[*] Kristensen, Jesper, Ilias Bilionis, and Nicholas Zabaras.
Physical Review B 87.17 (2013)
Predicting Ground State Line in Mg-Li
PROBLEM SPECIFICATION

- True ground state line
  (Phase diagram at $T=0$)
    - Compute an infinite number of structures
    - Plot formation energy versus composition
    - Form convex hull
      - Lowest energy structures spanning the bottom of the hull are ground states

- Approximation to ground state line
  - Form ground state line from (limited) data set $D$
RJMCMC shows sparse features as expected.

- Model complexity less than a third of maximum (22)
  - ~6

- Model complexities of other methods:
  - LASSO with 10-fold CV:
    - Max complexity: 20
  - Least squares:
    - Maximum complexity: 22

- Mg-Li: 16 2-pt, 4 3-pt + empty and 1-pt. Max spatial extent 20 Å.
- Si-Ge: 9 2-pt, 6 3-pt, Max Spatial extent 8 Å.
Noise estimated by RJMCMC agrees with DFT noise

Agrees with order of magnitude expected from density functional theory

Assumption about model errors seems reasonable

*Xie, Wei, et al. Physical Review B 88.23 (2013)*
Details on particular models considered

Physically more satisfying result than LASSO and Least squares*

Quantify Uncertainty in Ground State Line

Approximate ground state line

Data set $D$

LASSO with 10-fold CV Least squares

99 % Bayesian interval

$p(I|\mathcal{D}) = \int d\theta \delta(I[f(\cdot; \theta)] - I)p(\theta|\mathcal{D})$
The 99 % Bayesian interval is what we call the “statistics” in Step 4:

1) Collect (expensive) data set $\mathcal{D} = \{x_j\}_{j=1}^N$.
2) Collect parametrizations of the surrogate $\mathcal{T} = \{\theta_i\}_{i=1}^M$ from (some) Bayesian posterior.
3) for $i$ in 1...$M$:
   Compute the quantity of interest (QI) resulting from the $i$th parametrization, call this QI the $i$th QI: $I_i[f(\cdot; \theta_i)]$, or simply $I_i$.
4) Compute and output the statistics of $\{I_i\}_{i=1}^M$.

For each sampled posterior (i.e., each cluster expansion consistent with the data $D$), we compute the ground state line and called the $i$th ground state line $I_i$. 
Predictions seem reasonable
- Error bars are much smaller than the order of the formation energies themselves
- How tolerable the error bars are depends on the application of course

But: Cluster Expansion seems to capture well the quantum mechanical energies

Thus, with a rigorous Bayesian treatment of uncertainty propagation
- We now understand why cluster expansions are successful in quantum mechanical energy computations
- The cluster expansion captures variations in the response surface from truncation and estimation of corresponding coefficients
- This enables its use in materials design
PREDICTING PHASE TRANSITION IN Si-Ge
What is a disordered-ordered alloy phase transition?

High temperature
- Atoms are arranged randomly on lattice:
  - Entropic energy wins over configurational energy in the free energy

Low temperature
- Atoms are ordered:
  - Entropic energy very small in magnitude: configurational energy takes over in the free energy and governs atoms

Problem
- When does transition from disordered to ordered state occur?
- We compute this with the surrogate model and an MCMC method
  - Adaptive sequential Monte Carlo (ASMC)*

Goal
- Quantify uncertainty in phase transition (not done before!)

To perform average:
Sample a bunch of states with probability:
\[ \propto \exp(-\beta \hat{E}(\sigma; \gamma)) \]

Traditional solution:
Traditional Markov Chain Monte Carlo (MCMC).

Go to each \( T \) and thermalize+take \( M \) samples.

\[ \langle Q(T) \rangle = \frac{1}{M} \sum_{j=1}^{M} Q(\sigma^{(j)}) \]

A peak (divergence in the limit of an infinite lattice) in the specific heat signals the phase transition.
Better Alternative: Adaptive Sequential Monte Carlo
\[
p(\sigma | \beta_0) \approx \sum_{i=1}^{N_P} w_0^{(i)} \delta(\sigma - \sigma^{(i)})
\]

- 64 `particles’ starting at 2000 K
- 30x30x30 MC cell with periodic BCs
- Adaptive step size (Max 10 K)
- 100 flips per lattice size – double spin flip dynamics
- Avoid introducing additional uncertainties using the same # of particles with each sample & same seed.

\[
\langle Q(T) \rangle = \sum_{j=1}^{N_P} w_s^{(j)} Q(\sigma_s^{(j)})
\]
Obtain Phase Transition: Adaptive Sequential Monte Carlo

- Particle approximation

\[ p(\sigma | \beta) \approx \sum_{i=1}^{N_p} w_0^{(i)}(\beta) \delta(\sigma - \sigma^{(i)}) , \]

- Introduce path connecting two temperatures (\( \xi \in [0, 1] \)):

\[ p_\gamma(\sigma) \propto \exp(-\beta_\xi E(\sigma)) \]

- With the intermediate temperature defined as:

\[ \beta_\xi \equiv (1 - \xi) \beta_0 + \xi \beta_1 \]

1. Set \( s = 0 \) and \( \xi_s = 0 \). Sample a particle approximation from \( p_{\xi_0}(\sigma) : \{ w_0^{(i)} , \sigma_0^{(i)} \}_{i=1}^{N_p} \).

2. Determine \( \xi_{s+1} \in [\xi_s, 1] \) such that

\[ \text{ESS}(\xi_{s+1}) := \frac{1}{\sum_{i=1}^{N_p} (w_s^{(i)})^2} = \zeta \text{ESS}(\xi_s) . \]

where \( w_{s+1}^{(i)} \) is the normalized version of \( W_{s+1}^{(i)} = w_s^{(i)} \hat{w}_{s+1}^{(i)} \) with:

\[ \hat{w}_{s+1}^{(i)} = \exp\left( - (\beta_{\xi{s+1}} - \beta_{\xi_s}) E_s^{(i)} \right) . \]

3. If \( \text{ESS}(\xi_{s+1}) < \text{ESS}_m \) then resample.

4. For \( i = \{1...N_p\} \) perform \( N_m \) steps using a MCMC transition kernel with initial state given by the previous particle.

5. If \( \xi_s = 1 \), STOP. Otherwise set \( s \leftarrow s + 1 \) and go to step 2.

Not shown here
- Diamond Si-Ge at 50 % composition

- Sparse as expected

- Model complexity less than a third of maximum
  - \( \sim 4 \)

- Compare to other methods:
  - LASSO with 10-fold CV:
    - 10
  - Least squares:
    - Maximum complexity
      - 16
Noise is as expected from density functional theory calculations.

On the order of meV per atom.
Details on particular models considered:

Again: Most physically satisfying of considered models

Interesting observation

- Whenever least squares and LASSO disagree on underlying physics, RJMCMC predicts that no conclusion can be drawn based on observed data!
RESULTS

5 runs

99 % Bayesian interval

19.7 K
Si–Ge Conclusion

- Again: Predictions seem reasonable
  - Error bars are much smaller than the order of the transitions themselves

- Cluster Expansion seems to capture well the quantum mechanical energies

- As expected from numerous applications the cluster expansion does work well for alloy modeling

- Overall Conclusion
  - We have provided a rigorous framework for quantifying the uncertainty in the cluster expansion
  - We introduced a fully Bayesian posterior on the cluster expansion inspired by LASSO and sampled it via RJMCMC
  - On rigorous grounds we can say that the cluster expansion describes quantum mechanical energies well
  - Our framework is useful for other properties as well
CONCLUSION

- Bayesian Probability Theory to quantify uncertainty in cluster expansion

- We introduced a fully Bayesian posterior on the cluster expansion inspired by LASSO and sampled it via RJMCMC

- On rigorous grounds we can say that the cluster expansion describes quantum mechanical energies well

- Our framework is useful for other properties as well
THANK YOU