A. von Lilienfeld
Univ Basel

*L. F. Arsenault*
CU Physics

Richard Neuberg
CU stats

Lauren Hannah
CU Stats & Data Science


Support: US DOE
Columbia RISE program

The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016
Machine learning

Relation between input (specified by vector of ‘descriptors’ \( D \)) and set of outputs \( O \)

\[ O(D) = M(D, D_1) \]

Relation determined by intelligent interpolation from outputs \( O_1 \) produced by learning set \( D_1 \)
Many body physics:
why bother with machine learning

The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016
Many body physics: why bother with machine learning

- Solving a many body problem is VERY expensive =>
  - Leverage existing results to obtain an inexpensive, approximate solution

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University
Many body physics: why bother with machine learning

- Often, need (many) small extrapolations from known solutions
  - Structural relaxation (energy as function of many different sets of atomic positions)
  - variations in strain and chemical composition
  - defect energies and ionic mobilities

The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016
Many body physics: why bother with machine learning

- State space of quantum many body problems is exponentially big; parameter space of situations to be considered also often very large.
Many body physics: why bother with machine learning

- State space of quantum many body problems is exponentially big; parameter space of situations to be considered also often very large.
- ‘Big data’ methods find the relevant very low dimensional subspaces of very high dimensional data sets.
Many body physics: why bother with machine learning

- State space of quantum many body problems is exponentially big; parameter space of situations to be considered also often very large.
- 'Big data' methods find the relevant very low dimensional subspaces of very high dimensional data sets.
- Question (no concrete ideas yet): ?Can these methods give us a way through the wilderness of quantum complexity?

The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016
Machine learning

Questions:

- What is right set of input descriptors
- How do you represent the output
- ?How big a learning/training set do you need?
The Quantum Many Body Problem:

Most common version: electrons in a fixed atomic environment

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University
The Quantum Many Body Problem:

Most common version: electrons in a fixed atomic environment

Spin crossover molecule

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University
The Quantum Many Body Problem:

Most common version: electrons in a fixed atomic environment

Key question: magnetic state of Fe ion

Spin crossover molecule

The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016
The Quantum Many Body Problem:

Most common version: electrons in a fixed atomic environment

T times magnetic susceptibility vs Temperature

Spin crossover molecule

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University
The Quantum Many Body Problem:

Most common version: electrons in a fixed atomic environment

Spin crossover molecule

Change of Structure

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University
The Quantum Many Body Problem:

Most common version: electrons in a fixed atomic environment

energy vs structure

Spin crossover molecule

The Simons foundation
Department of Physics, Columbia University
The Quantum Many Body Problem:

Most common version: electrons in a fixed atomic environment

LOTS of (minor) variants:
- Change bond lengths
- Change C-H backbone
- Change immediate Fe environment (N->?)
- Change transition metal ion
- ....

Spin crossover molecule

Opportunity for inference from database

The Simons foundation
Department of Physics, Columbia University
The Quantum Many Body Problem:

High Tc (copper-oxide) superconductivity

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University
The Quantum Many Body Problem:

High Tc (copper-oxide) superconductivity

Complicated atomic arrangement
- subset of `relevant` electronic states
  => minimal theoretical description
  (`Hubbard model`)

The Simons foundation
Department of Physics, Columbia University
The Quantum Many Body Problem:

High Tc (copper-oxide) superconductivity

Complicated atomic arrangement
- subset of `relevant’ electronic states
  => minimal theoretical description
  (`Hubbard model’’)
- many variants (model parameters
depend on details of atomic
arrangements)
The Quantum Many Body Problem:

High Tc (copper-oxide) superconductivity

Complicated atomic arrangement
- subset of `relevant’ electronic states
  => minimal theoretical description
    (`Hubbard model’)
- many variants (model parameters depend on details of atomic arrangements)
- Question: how do we optimize superconductivity—solve many version of model

Opportunity for inference

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University
The Quantum Many Body Problem:

The inputs:
``G_0``—electron propagation in reference noninteracting environment: contains information about atomic positions, background electronic structure

``U``: electron-electron interaction matrix elements between relevant states

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University
The Quantum Many Body Problem:

The output:

```
G
```
describes interacting electron behavior in physical system

```
\Sigma
```
self energy: parametrizes difference in electron behavior between physical system and reference

The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016
The Quantum Many Body Problem: A Functional Relation

Formally: extremizing a functional

\[ \Phi[\{G(p, \omega)\}] = \Phi_{\text{univ}}[\{G\}] + \text{Tr} \ln [G] - \text{Tr} [G_0^{-1} G] \]

Interactions ("U") determine \( \Phi_{\text{univ}} \)

Stationarity: \( \frac{\delta \Phi}{\delta G} = 0 \) determines \( G[G_0] \)

\[ \Rightarrow \text{Need to learn a functional relationship} \]

The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016
The Quantum Many Body Problem: A Functional Relation

Formally: extremizing a functional

\[ \Phi[\{G(p, \omega)\}] = \Phi_{\text{univ}}[\{G\}] + \text{Tr} \ln [G] - \text{Tr} \left[ G_0^{-1} G \right] \]

Interactions ("U") determine \( \Phi_{\text{univ}} \)

Stationarity: \[ \frac{\delta \Phi}{\delta G} = 0 \] determines \( G[G_0] \)

\[ \Rightarrow \text{Need to learn a functional relationship} \]
Learning functional relations

Need:

- representation of function (descriptors)
- way to impose constraints

Would like to know

- Size of needed database
- Accuracy of prediction:
  - global fit
  - particular physically relevant aspects
- Back-inference—can we design desired behavior
Learning functional relations

Need:

- representation of function (descriptors)
- way to impose constraints

Would like to know

- Size of needed database
- Accuracy of prediction:
  - global fit
  - particular physically relevant aspects
- Back-inference—can we design desired behavior
A mathematical digression

$G(z)$ is analytic in complex frequency $(z)$ plane except for branch cut discontinuity across $\text{Im } z = 0$

Spectral function:

$$A(\omega) = \frac{G(\omega - i\delta) - G(\omega + i\delta)}{2i}$$

is non-negative, integrates to $P$, and (typically) has compact support has physical meaning of density of states: prob to add or remove particle

Copyright A. J. Millis 2016
\[
\Rightarrow \quad G(z) = \int \frac{dx}{\pi} \frac{A(x)}{z - x}
\]

Convenient to evaluate \(G\) on imaginary frequency axis at points
\[
z_n = i\omega_n \equiv (2n + 1)\pi T
\]
or to construct
\[
G(\tau) = T \sum_n e^{i\omega_n \tau} G(i\omega_n)
\]

Complex variable theory: A can be reconstructed from \(G(z_n)\)

The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016
Learning Many Body Physics: 
Step 1: the Anderson Model

One (spin-degenerate) interacting level coupled to a bath of noninteracting electrons

\[ H = \sum_{\sigma} \varepsilon_d d_\sigma^\dagger d_\sigma + Ud_\uparrow^\dagger d_\downarrow^\dagger d_\downarrow + \sum_{k, \sigma} \varepsilon_k c_k^\dagger c_{k\sigma} + \sum_{k, \sigma} V_k \left( d_\sigma^\dagger c_{k\sigma} + c_{k\sigma}^\dagger d_\sigma \right). \]


Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University
The Anderson Model

One (spin-degenerate) interacting level coupled to a bath of noninteracting electrons

\[ H = \sum_\sigma \epsilon_d d_\sigma^\dagger d_\sigma + U d_\uparrow^\dagger d_\uparrow^\dagger d_\downarrow d_\downarrow + \sum_{k,\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \sum_{k,\sigma} V_k (d_\sigma^\dagger c_{k\sigma} + c_{k\sigma}^\dagger d_\sigma) \]

Many-body density of states


Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University
The Anderson Model

One (spin-degenerate) interacting level coupled to a bath of noninteracting electrons

\[ H = \sum_{\sigma} \varepsilon_d d_{\sigma}^\dagger d_{\sigma} + U d_{\uparrow}^\dagger d_{\uparrow}^\dagger d_{\downarrow} d_{\downarrow} + \sum_{k,\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \sum_{k,\sigma} V_k \left( d_{\sigma}^\dagger c_{k\sigma} + c_{k\sigma}^\dagger d_{\sigma} \right) \]

Side-bands


Copyright A. J. Millis 2016

Many-body density of states

The Simons foundation
Department of Physics, Columbia University
The Anderson Model

One (spin-degenerate) interacting level coupled to a bath of noninteracting electrons

\[ H = \sum_{\sigma} \epsilon_d d_\sigma^\dagger d_\sigma + U d_\uparrow^\dagger d_\uparrow^\dagger d_\downarrow d_\downarrow + \sum_{k,\sigma} \epsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \sum_{k,\sigma} V_k \left( d_\sigma^\dagger c_{k\sigma} + c_{k\sigma}^\dagger d_\sigma \right). \]

Central (Kondo) peak

Width \(\rightarrow\) renormalization factor \(Z\)


Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University
Anderson Model: Inputs

One (spin-degenerate) interacting level coupled to a bath of noninteracting electrons

\[
H = \sum_{\sigma} \varepsilon_d d_{\sigma}^\dagger d_{\sigma} + U d_{\uparrow}^\dagger d_{\uparrow} d_{\downarrow}^\dagger d_{\downarrow} \\
+ \sum_{k,\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \sum_{k,\sigma} V_k (d_{\sigma}^\dagger c_{k\sigma} + c_{k\sigma}^\dagger d_{\sigma})
\]

Bath characterized by

\[
N(\omega) = \pi \sum_{k} V_k^2 \delta(\omega - \varepsilon_k)
\]

Details not important. We choose

\[
N(\omega) = V^2 \sqrt{1 - \omega^2} \quad (\omega^2 < 1)
\]

=> parameter is V


Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University
The Anderson Model: Inputs

One (spin-degenerate) interacting level coupled to a bath of noninteracting electrons

\[ H = \sum_{\sigma} \epsilon_d d_{\sigma}^\dagger d_{\sigma} + U n_d \]  
\[ + \sum_{k,\sigma} \epsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \sum_{k,\sigma} V_k \left( d_{\sigma}^\dagger c_{k\sigma} + c_{k\sigma}^\dagger d_{\sigma} \right). \]

\( n_d \)  

Interaction

\( (V, U, n_d) \)


Copyright A. J. Millis 2016
The Anderson Model: Inputs

One (spin-degenerate) interacting level coupled to a bath of noninteracting electrons

\[ H = \sum_{\sigma} \varepsilon_d d^\dagger_d d_\sigma + U n_{d \uparrow} n_{d \downarrow} \]

\[ + \sum_{k, \sigma} \varepsilon_k c^\dagger_{k\sigma} c_{k\sigma} + \sum_{k, \sigma} V_k \left( d^\dagger_\sigma c_{k\sigma} + c^\dagger_{k\sigma} d_\sigma \right) . \]

Descriptor: \((V, U, n_d)\)


Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University
The Anderson Model: Output

One (spin-degenerate) interacting level coupled to a bath of noninteracting electrons

\[ G(\tau) = -\left\langle d_\sigma(\tau) d_\sigma^\dagger(0) \right\rangle \]

\[
H = \sum_\sigma \varepsilon_d d_\sigma^\dagger d_\sigma + U d_\uparrow^\dagger d_\uparrow d_\downarrow^\dagger d_\downarrow + \sum_{k,\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \sum_{k,\sigma} V_k \left( d_\sigma^\dagger c_{k\sigma} + c_{k\sigma}^\dagger d_\sigma \right).
\]


Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University
The Anderson Model: Output

One (spin-degenerate) interacting level coupled to a bath of noninteracting electrons

\[
H = \sum_{\sigma} \varepsilon_d d_\sigma^\dagger d_\sigma + U d_\uparrow^\dagger d_\uparrow^\dagger d_\downarrow d_\downarrow + \sum_{k,\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \sum_{\sigma} V_k \left( d_\sigma^\dagger c_{k\sigma} + c_{k\sigma}^\dagger d_\sigma \right).
\]

Represented as:
- Coefficients of Legendre polynomials
- Continued fraction expansion
- Values at frequencies \( z_n \)


Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University
The Anderson Model: method

One (spin-degenerate) interacting level coupled to a bath of noninteracting electrons

\[
H = \sum_{\sigma} \varepsilon_d d_\sigma^\dagger d_\sigma + U d_\uparrow^\dagger d_\uparrow^\dagger d_\downarrow d_\downarrow + \sum_{k,\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \sum_{k,\sigma} V_k \left( d_\sigma^\dagger c_{k\sigma} + c_{k\sigma}^\dagger d_\sigma \right).
\]

Kernel Ridge Regression

\[
g_m(D) = \sum_l \alpha_{lm} K_m(D_l, D)
\]

Kernel

\[
K(D_i, D) = e^{-\frac{|d_i|}{\sigma}}
\]

- Simple exponential kernel worked best.
- \(d = \)Manhattan distance.
- Width large (~10)


Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University
The Anderson Model: method

One (spin-degenerate) interacting level coupled to a bath of noninteracting electrons

\[ H = \sum_{\sigma} \varepsilon_d d^\dagger_\sigma d_\sigma + Ud^\dagger_\uparrow d_\uparrow + Ud^\dagger_\downarrow d_\downarrow + \sum_{k,\sigma} \varepsilon_k c^\dagger_{k\sigma} c_{k\sigma} + \sum_{k,\sigma} V_k \left( d^\dagger_\sigma c_{k\sigma} + c^\dagger_{k\sigma} d_\sigma \right). \]

Kernel Ridge Regression

\[ g_m(D) = \sum_l \alpha_{lm} K_m(D_l, D) \]

coefficients fixed by training set

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University
The Anderson Model: Results

Error averaged over some low order coefficients as function of learning set length


Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University
The Anderson Model: Results

Learning Features of $G$ and $A$: continued fraction representation

![Graph showing learning set length: 500](image)

- **Exact**
- **ML result**


Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University
The Anderson Model: Results

Learning Features of G and A: continued fraction representation

![Graph showing complex function values with annotations](image)

- **Exact**
- **ML result**

Learning Set length: 5000


Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

Learning set size ~1000 needed
The Anderson Model: Results
Learning Features of G and A: continued fraction representation

• Exact
  - ML: 500
  - ML: 5000


The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016
The Anderson Model: Results

the quasiparticle weight $Z$

- Continued fraction
  $X \quad Z_n$
- Time
- Legendre

Learning set size
>1000 needed

Legendre rep. preferred because most compact


Copyright A. J. Millis 2016
The Anderson Model: Results

• We can learn functions
• Legendre rep. preferred (most compact representation; minimizes dimension of input space)
• Required dataset is not small


The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016
Learning a lattice many-body problem

\[ H = \sum_{ij\sigma} t_{ij} c_{i\sigma}^\dagger c_{j\sigma} + U \sum_i \hat{n}_{i\uparrow}\hat{n}_{i\downarrow} \]

Model has two phases:
- Mott Insulator: \( n=1 \), large \( U \).
  Characterized by gap
- Metal, \( n \neq 1 \) or \( U \) not large.
  Characterized by mass enhancement

We will learn a particular approximate solution, obtained by the `dynamical mean field’ method
Learning a lattice many-body problem

$$H = \sum_{ij\sigma} t_{ij} c_{i\sigma}^\dagger c_{j\sigma} + U \sum_{i} \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}$$

Descriptors:
- density of states ($t_{ij}$): Legendre coeffs
- $U$
- electron density

Output:
- phase (metal or insulator)
- $G$ (Legendre coeffs)

The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016
Learning a lattice many-body problem

\[ H = \sum_{ij\sigma} t_{ij} c_{i\sigma}^{\dagger} c_{j\sigma} + U \sum_{i} \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} \]

**Technical improvement:**
**cross validation**

Pick ML
hyperparameters to minimize error in
1st 5 Legendre coeffs of 1000 examples

The Simons foundation
Department of Physics, Columbia University
Cross-validation leads to much smaller errors

Training sets of ~few hundred are acceptable

N_{learn} \sim 1800
Learning a lattice many-body problem

\[ H = \sum_{ij\sigma} t_{ij} c_{i\sigma}^\dagger c_{j\sigma} + U \sum_{i} \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} \]

2 step learning:
- Classification
- Greens function (self energy)

\[ D = [(f_1^0, f_2^0, ..., f_N^0), U, \mu] \]

The Simons foundation
Department of Physics, Columbia University
Classification

- Random Decision Forest
- Use 1st 5 Legendre coeffs of DOS plus U,\mu

Classification: Almost perfectly accurate
Learning a lattice many-body problem

\[ H = \sum_{ij\sigma} t_{ij} c_{i\sigma}^\dagger c_{j\sigma} + U \sum_i \hat{n}_{i\uparrow}\hat{n}_{i\downarrow} \]

G reproduced perfectly

\[ N_{\text{learn}} \approx 1800 \]

- Exact
- ML

The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016
reverse engineering a lattice many-body problem

\[ H = \sum_{ij\sigma} t_{ij} c_{i\sigma}^+ c_{j\sigma} + U \sum_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} \]

**Question:**
What \( t_{ij} \) should I use to get a desired many-body DOS?

**Answer:**
Run the `machine’ backwards

\[ N(\omega) \Rightarrow g(i\omega_n) = G(\tau) \Rightarrow \{G_{\Gamma}\} \Rightarrow \text{D} \]

The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016
reverse engineering a lattice many-body problem

\[ N(\omega) \Rightarrow G(i\omega_n) \Rightarrow G(\tau) \Rightarrow \{G_l\} \Rightarrow \text{ML KRR} \Rightarrow \text{Material} \]

Target: gap of size 2. Method works with database of \( \sim 200 \) examples

\[
\begin{array}{c}
\text{Result} \\
\text{Target}
\end{array}
\]

Copyright A. J. Millis 2016
Summary

- Cross-validation + intelligent representation of functions =>
- many-body problems can be learned+reverse engineering done
- Needed training data sets not outrageously big
- Key point: represent input and output functions in compact orthogonal function basis

?Extension to realistic situation?

The Simons foundation
Department of Physics, Columbia University
Coda: analytical continuation

\[ G(z) = \int \frac{dx}{\pi} \frac{A(x)}{z - x} \]

Math: if you know G on the points (all n)

\[ z_n = i\omega_n = (2n + 1)\pi T \]

you can reconstruct A.
The difficulty:

The operator

\[ K(z, x) = \frac{1}{z - x} \]

has many very small eigenvalues.

\( \Rightarrow \) inversion is ill-posed.

Small errors in \( G(z_n) \) lead to big errors in inferred \( A \)
Coda: analytical continuation

\[ G(z) = \int \frac{dx}{\pi} \frac{A(x)}{z - x} \]

Math: if you know G on the points (all n)

\[ z_n = i\omega_n \equiv (2n + 1)\pi T \]

you can reconstruct A.

The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016
the standard solution: maximum entropy

Minimize

\[ F = \sum_n |G(z_n) - K(z_n, x) \ast A(x)|^2 + \alpha \int dx \ A(x) \ln \frac{A(x)}{M(x)} \]

with respect to \( A(x) \).

*Need `model function’ \( M(x) \) and parameter alpha
*No method for assessing errors.
Instead: ’learn’ the solution

=>

view problem as statistical inference,
not as operator inversion

• Forward problem (G given A) is computationally trivial=>easy to get large database
• Regression =>regularization: eliminates very small eigenvalues (at the expense of removing sharp features from result)
• constraints (positivity, unit area) imposed by projection
• **Error estimates available**

The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016
Instead: 'learn' the solution

=>

view problem as statistical inference, not as operator inversion

- Forward problem (G given A) is computationally trivial => easy to get large database
- Regression => regularization: eliminates very small eigenvalues (at the expense of removing sharp features from result)
- constraints (positivity, unit area) imposed by projection
- **Error estimates available**
Technical point: representation of functions

Represent $A$ by conformal coefficients

$$A(\omega) = \sum_{n=-\infty}^{\infty} a_n \left( \frac{\omega - i\omega_0}{\omega + i\omega_0} \right)^n$$

$$a_n = \oint \frac{dz}{2\pi i} \frac{A(i\omega_0 \left( \frac{1+z}{1-z} \right))}{z^{n+1}}$$

The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016
Learning and Tuning Sets

Many examples, constructed as sums of Gaussians

Figure 1: Two different (a) and (b) possible $A(\omega)$ in the database.

Database

Learning set
Size 10000

Tuning set I
Size 5000

Tuning set II
Size 5000

Test set
Size 5000

The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016
Learning and Tuning Sets

Figure 1: Two different (a) and (b) possible $A(\omega)$ in the database.

Note: learning involves fitting directly to real frequencies

The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016
Learning and Tuning Sets

![Graphs](image)

Figure 1: Two different (a) and (b) possible $A(\omega)$ in the database.

Note: learning involves fitting directly to real frequencies

Database

- Learning set
  - Size 10000
- Tuning set I
  - Size 5000
- Tuning set II
  - Size 5000
- Test set
  - Size 5000

The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016
Learning and Tuning Sets

Many examples, constructed as sums of Gaussians

Figure 1: Two different (a) and (b) possible $A(\omega)$ in the database.

Tune hyperparameters

Database

The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016
Tuning the hyperparameters
(1st level regression)
Learning and Tuning Sets

Many examples, constructed as sums of Gaussians

Figure 1: Two different (a) and (b) possible $A(\omega)$ in the database.

Tune covariance matrix (projection and error estimates)

Database

- Learning set
  - Size 10000
- Tuning set I
  - Size 5000
- Tuning set II
  - Size 5000
- Test set
  - Size 5000

The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016
Projection

Our functions have constraints (non-negative; low order moments fixed)

Constrained regression very difficult to implement; computationally inefficient.

=> do unconstrained regression, then find function in space that satisfies constraints and is `closest' to unconstrained

\[
\hat{A}_{\text{projected}} = \arg\min_{\{A^*: A^* \text{ satisfies constraints (22) to (25)}\}} ||A^* - \hat{A}||_M.
\]
Subtlety: some regions of frequency space are better determined than others

\[ \| \bullet \|_{\mathcal{M}} = \bullet^T \Sigma^{-1} \bullet \]

\[
\hat{\Sigma} = \sum_{k=1}^{p} \lambda_k v_k v_k^T + \text{diag}(\sum_{k=p+1}^{|\{\omega\}|} \lambda_k v_k v_k^T)
\]

chosen by training

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University
covariance matrix: also allows uncertainty estimates

90% confidence interval

Training of confidence interval done independently of regression. Our statistician collaborators recommend ‘quantile random forests’

The Simons foundation
Department of Physics, Columbia University
Error analysis: two metrics

Kubic-Leibler (minimize entropy)—this is the metric MaxEnt optimizes

\[ KL_j = \int d\omega A^j(\omega) \ln \frac{A^j(\omega)}{\hat{A}^j_{predicted}(\omega)} \]

\[ MAE_j = \frac{1}{N_\omega} \sum_{i=1}^{N_\omega} \left| A^j(\omega_i) - \hat{A}^j_{predicted}(\omega_i) \right|, \]

Mean Absolute Error (minimize difference)
Comparison to MaxEnt: modest noise

Red: statistical regression has lower error
Blue: Maxent has lower error

Figure 5: (Color online) Error for $\epsilon = \mathcal{N}(0, [10^{-3}]^2)$.
Comparison to MaxEnt: modest noise

Red: statistical regression has lower error
Blue: Maxent has lower error

Figure 5: (Color online) Error for $\epsilon = \mathcal{N}(0, [10^{-3}])^2$.

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University
Tiny noise ($10^{-5}$)

Only for smallest noise, and for K-L error is Max-Ent comparably good

The Simons foundation
Department of Physics, Columbia University
Continuation Conclusions

- Ill-posed inverse problem solved by statistical inference
- Method regularizes operator and provides error estimate (indep. of solution)
- A few months work produced a method comparable (better for noisy data) to Max Ent, which has had 2 decades of optimization
- Method could be useful for other inverse problems

The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016