Title: TBA

Date: Aug 11, 2016 11:00 AM

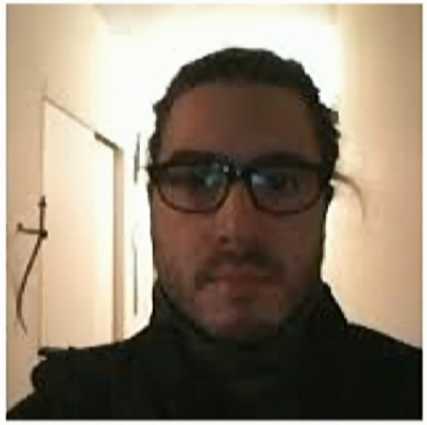
URL: http://pirsa.org/16080015

Abstract:

Collaborators and support

A. von Lilienfeld
Uni Basel

L. F. Arsenault
CU Physics



Richard Neuberg
CU stats

Lauren Hannah
CU Stats & Data Science

Phys. Rev. B90 155136 (2014); arXiv:1506.08858 and to appear

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

$$O(D)=M(D, D_l)$$

Relation determined by intelligent interpolation from outputs O_l produced by learning set D_l

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

Many body physics: why bother with machine learning

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

Many body physics: why bother with machine learning

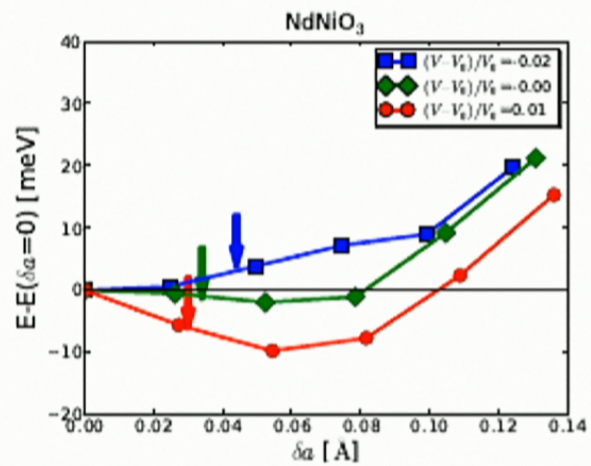
- **Solving a many body problem is VERY expensive=>**
  - **Leverage existing results to obtain an inexpensive, approximate solution**

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**



Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

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.

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

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.

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

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?

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

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?**

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

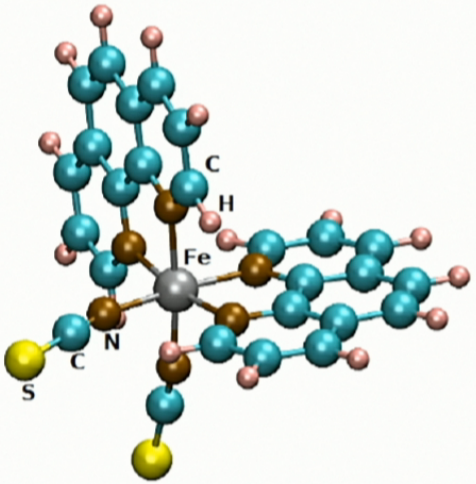
Copyright A. J. Millis 2016

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

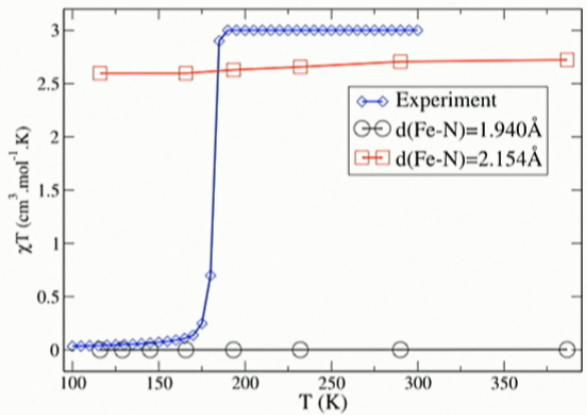
Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

The Quantum Many Body Problem:

Most common version: electrons in a fixed atomic environment

T times magnetic susceptibility vs Temperature



Spin crossover molecule

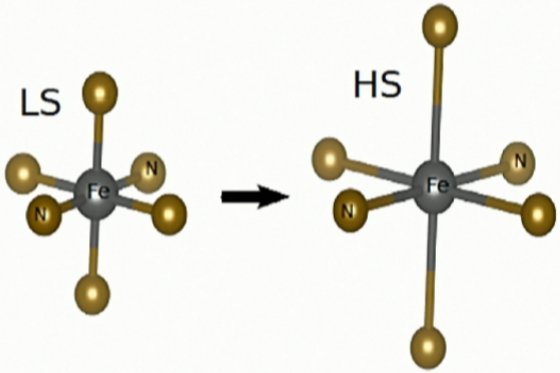
Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

The Quantum Many Body Problem:

Most common version: electrons in a fixed atomic environment

Change of Structure



Spin crossover molecule

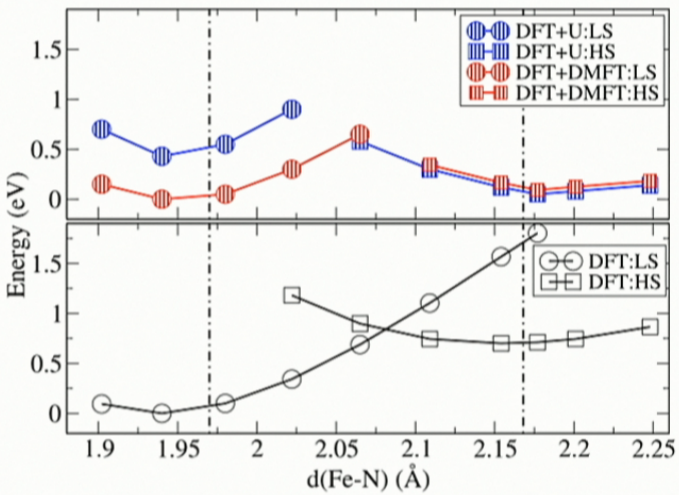
Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

The Quantum Many Body Problem:

Most common version: electrons in a fixed atomic environment

energy vs structure



Spin crossover molecule

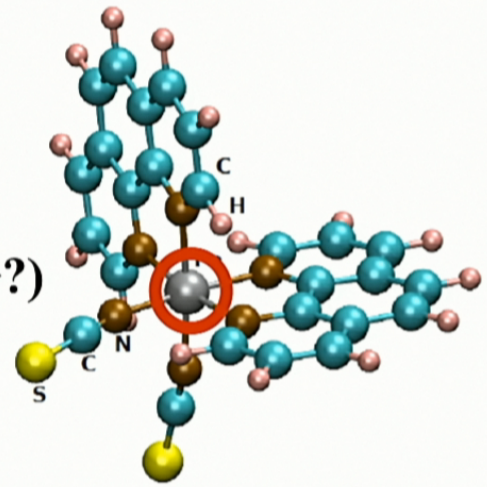
Copyright A. J. Millis 2016

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



LOTS of (minor) variants:

- Change bond lengths
- Change C-H backbone
- Change immediate Fe environment (N->?)
- Change transition metal ion
- ….

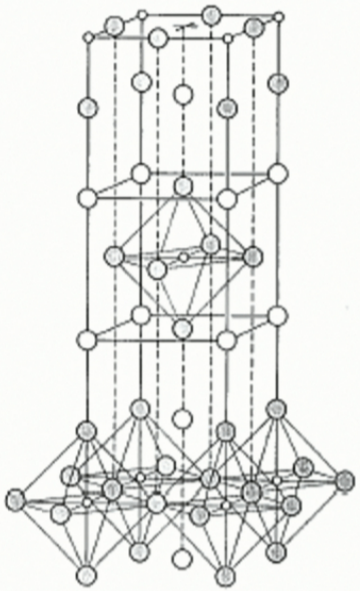
Opportunity for inference from database

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

The Quantum Many Body Problem:

High Tc (copper-oxide) superconductivity



Copyright A. J. Millis 2016

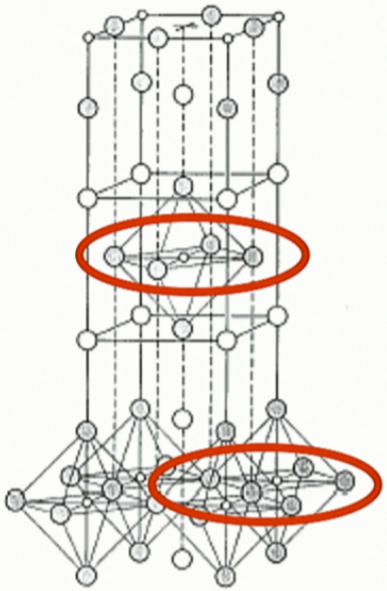
The Simons foundation
Department of Physics, Columbia University

The Quantum Many Body Problem:

High Tc (copper-oxide) superconductivity

Complicated atomic arrangement

- **subset of `relevant' electronic states => minimal theoretical description (``Hubbard model")**



Copyright A. J. Millis 2016

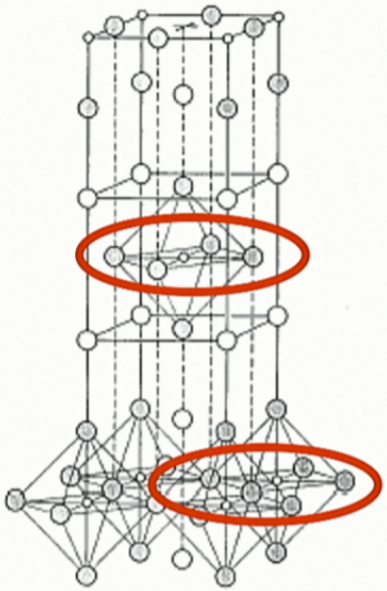
The Simons foundation
Department of Physics, Columbia University

The Quantum Many Body Problem:

High Tc (copper-oxide) superconductivity

Complicated atomic arrangement

- **subset of `relevant' electronic states => minimal theoretical description (``Hubbard model")**
- **many variants (model parameters depend on details of atomic arrangements)**



Copyright A. J. Millis 2016

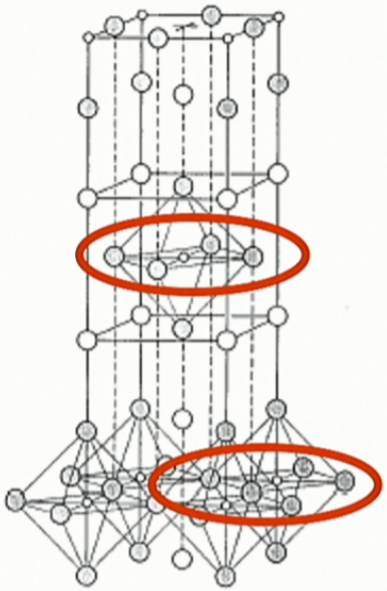
The Simons foundation
Department of Physics, Columbia University

The Quantum Many Body Problem:

High Tc (copper-oxide) superconductivity

Complicated atomic arrangement

- subset of `relevant' electronic states => minimal theoretical description (``Hubbard model")
- 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

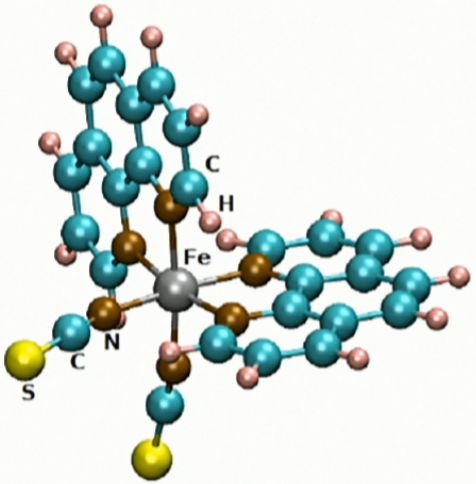
The Simons foundation
Department of Physics, Columbia University

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

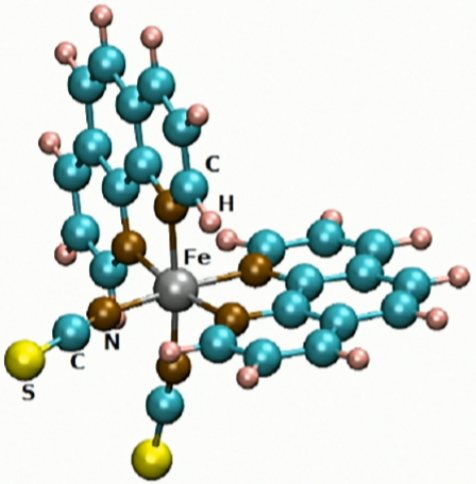
The Simons foundation
Department of Physics, Columbia University

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



Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

The Quantum Many Body Problem: A Functional Relation

Formally: extremizing a functional

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

Interactions ("U") determine Φ_{univ}

Stationarity: $\dfrac{\delta\Phi}{\delta G} = 0$ **determines $G[G_0]$**

=>Need to learn a functional relationship

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

The Quantum Many Body Problem: A Functional Relation

Formally: extremizing a functional

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

Interactions ("U") determine Φ_{univ}

Stationarity: $\dfrac{\delta\Phi}{\delta G} = 0$ **determines $G[G_0]$**

=>Need to learn a functional relationship

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

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**

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

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**

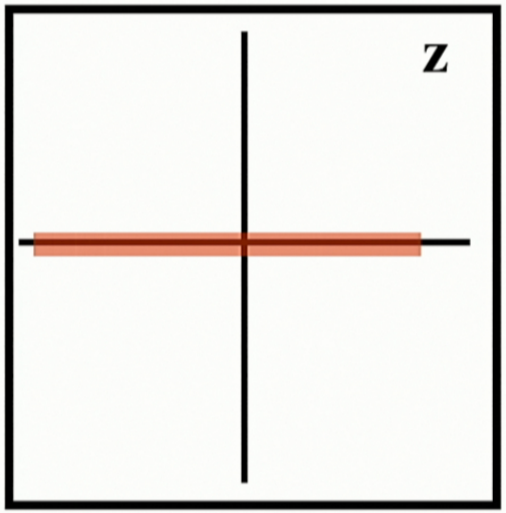
Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

A mathematical digression

G(z) is analytic in complex frequency (z) plane

except for branch cut discontinuity across Im z=0



Spectral function:

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

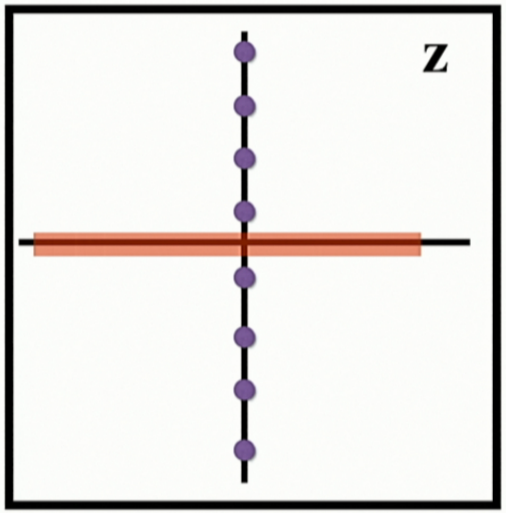
is non-negative, integrates to Pi, and (typically) has compact support

has physical meaning of density of states: prob to add or remove particle

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

$$\Rightarrow \boxed{G(z) = \int \frac{dx}{\pi} \frac{A(x)}{z - x}}$$



Convenient to evaluate G on imaginary frequency axis at points

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

or to construct

$$G(\tau) = T \sum_n e^{i\omega_n \tau} G(i\omega_n)$$

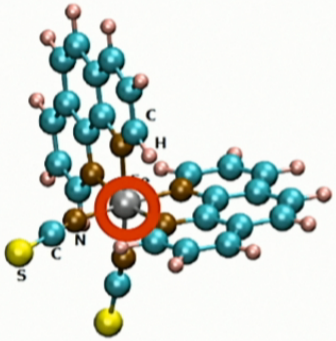
Complex variable theory: A can be reconstructed from $G(z_n)$

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

Learning Many Body Physics: Step 1: the Anderson Model

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



$$H = \sum_{\sigma} \varepsilon_d d_\sigma^\dagger d_\sigma + U d_\uparrow^\dagger d_\uparrow d_\downarrow^\dagger d_\downarrow$$

$$+ \sum_{k,\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \sum_{k,\sigma} V_k \left(d_\sigma^\dagger c_{k\sigma} + c_{k\sigma}^\dagger d_\sigma \right).$$

Louis-Francois Arsenault, Alejandro Lopez-Bezanilla, O. Anatole von Lilienfeld, A. J.M., Phys. Rev. B90 155136 (2014).

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

The Anderson Model

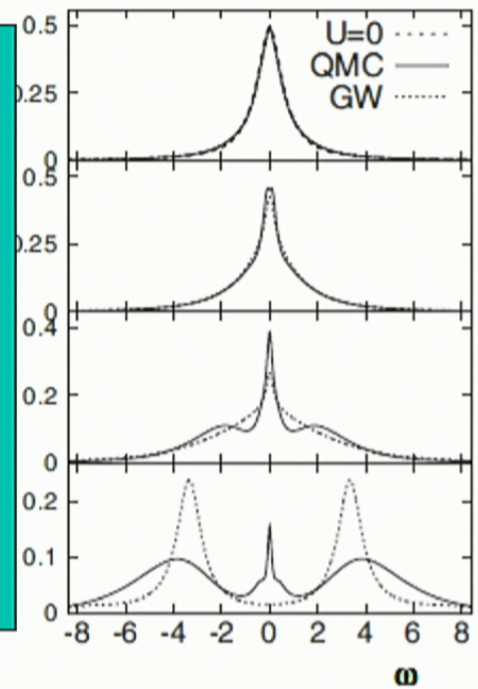
One (spin-degenerate) interacting level coupled to a bath of noninteracting electrons

$$H = \sum_{\sigma} \varepsilon_d d_\sigma^\dagger d_\sigma + U d_\uparrow^\dagger d_\uparrow d_\downarrow^\dagger d_\downarrow$$

$$+ \sum_{k,\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \sum_{k,\sigma} V_k \left(d_\sigma^\dagger c_{k\sigma} + c_{k\sigma}^\dagger d_\sigma \right)$$

Many-body density of states

$$A(\omega) = \mathrm{Im}G(\omega - i\delta)$$



U increases ->

Louis-Francois Arsenault, Alejandro Lopez-Bezanilla, O. Anatole von Lilienfeld, A. J.M., Phys. Rev. B90 155136 (2014).

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

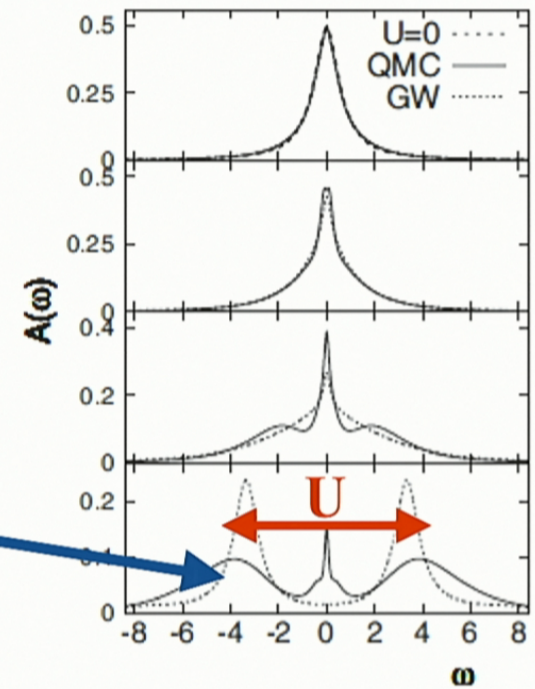
The Anderson Model

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

$$H = \sum_{\sigma} \varepsilon_d d_\sigma^\dagger d_\sigma + U d_\uparrow^\dagger d_\uparrow d_\downarrow^\dagger d_\downarrow$$

$$+ \sum_{k,\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \sum_{k,\sigma} V_k \left(d_\sigma^\dagger c_{k\sigma} + c_{k\sigma}^\dagger d_\sigma \right).$$

Many-body density of states



U increases ->

Side-bands

Louis-Francois Arsenault, Alejandro Lopez-Bezanilla, O. Anatole von Lilienfeld, A. J.M., Phys. Rev. B90 155136 (2014).

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

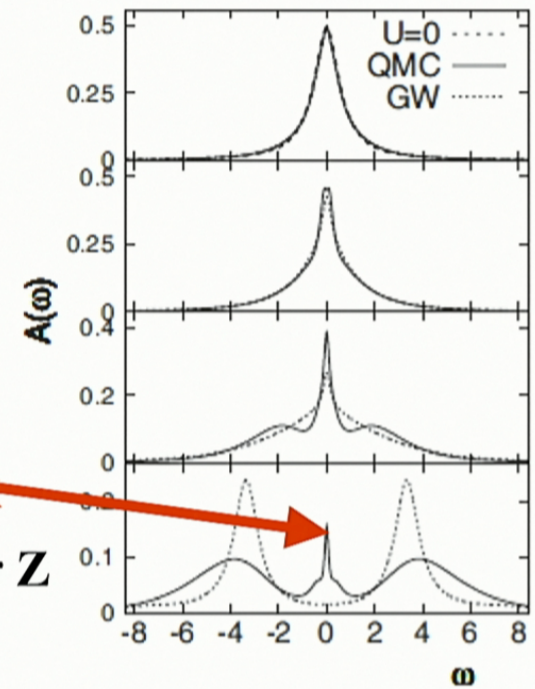
The Anderson Model

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

$$H = \sum_{\sigma} \varepsilon_d d_\sigma^\dagger d_\sigma + U d_\uparrow^\dagger d_\uparrow d_\downarrow^\dagger d_\downarrow$$

$$+ \sum_{k,\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \sum_{k,\sigma} V_k \left(d_\sigma^\dagger c_{k\sigma} + c_{k\sigma}^\dagger d_\sigma \right).$$

Many-body density of states



Central (Kondo) peak

Width=>renormalization factor Z

U increases ->

Louis-Francois Arsenault, Alejandro Lopez-Bezanilla, O. Anatole von Lilienfeld, A. J.M., Phys. Rev. B90 155136 (2014).

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

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

Bath characterized by

$$\mathbf{N}(\omega) = \pi \sum_k V_k^2 \delta(\omega - \varepsilon_k)$$

Details not important. We choose

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

=>parameter is V

Louis-Francois Arsenault, Alejandro Lopez-Bezanilla, O. Anatole von Lilienfeld, A. J.M., Phys. Rev. B90 155136 (2014).

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

<=> occupancy of d-level, n_d

$$H = \sum_{\sigma} \varepsilon_d d_\sigma^\dagger d_\sigma + U d_\uparrow^\dagger d_\uparrow d_\downarrow^\dagger d_\downarrow$$

$$+ \sum_{k,\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \sum_{k,\sigma} V_k \left(d_\sigma^\dagger c_{k\sigma} + c_{k\sigma}^\dagger d_\sigma \right).$$

Interaction

descriptor:

(V, U, n_d)

Louis-Francois Arsenault, Alejandro Lopez-Bezanilla, O. Anatole von Lilienfeld, A. J.M., Phys. Rev. B90 155136 (2014).

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

<=> occupancy of d-level, n_d

$$H = \sum_{\sigma} \varepsilon_d d_\sigma^\dagger d_\sigma + U d_\uparrow^\dagger d_\uparrow d_\downarrow^\dagger d_\downarrow$$

$$+ \sum_{k,\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \sum_{k,\sigma} V_k \left(d_\sigma^\dagger c_{k\sigma} + c_{k\sigma}^\dagger d_\sigma \right).$$

Interaction

descriptor:

(V, U, n_d)

Louis-Francois Arsenault, Alejandro Lopez-Bezanilla, O. Anatole von Lilienfeld, A. J.M., Phys. Rev. B90 155136 (2014).

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

$$\mathbf{G}(\tau) = -\left\langle \mathbf{d}_\sigma(\tau)\mathbf{d}_\sigma^\dagger(\mathbf{0})\right\rangle$$

$$H = \sum_\sigma \varepsilon_d d_\sigma^\dagger d_\sigma + U d_\uparrow^\dagger d_\uparrow d_\downarrow^\dagger d_\downarrow$$

$$+ \sum_{k,\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \sum_{k,\sigma} V_k \left(d_\sigma^\dagger c_{k\sigma} + c_{k\sigma}^\dagger d_\sigma\right).$$

Louis-Francois Arsenault, Alejandro Lopez-Bezanilla, O. Anatole von Lilienfeld, A. J.M., Phys. Rev. B90 155136 (2014).

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

$$\mathbf{G}(\tau) = -\left\langle \mathbf{d}_\sigma(\tau)\mathbf{d}_\sigma^\dagger(0)\right\rangle$$

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

Represented as:

- **Coefficients of Legendre polynomials**
- **Continued fraction expansion**
- **Values at frequencies z_n**

Louis-Francois Arsenault, Alejandro Lopez-Bezanilla, O. Anatole von Lilienfeld, A. J.M., Phys. Rev. B90 155136 (2014).

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

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

Louis-Francois Arsenault, Alejandro Lopez-Bezanilla, O. Anatole von Lilienfeld, A. J.M., Phys. Rev. B90 155136 (2014).

Copyright A. J. Millis 2016

Kernel Ridge Regression

$$g_m(\mathbf{D}) = \sum_l \alpha_{lm} K_m(\mathbf{D}_l, \mathbf{D})$$

Kernel

$$K(\mathbf{D}_i, \mathbf{D}) = \mathrm{e}^{-\frac{|\boldsymbol{d}_i|}{\sigma}}$$

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

The Simons foundation
Department of Physics, Columbia University

The Anderson Model: method

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

$$H = \sum_{\sigma} \varepsilon_d d_\sigma^\dagger d_\sigma + U d_\uparrow^\dagger d_\uparrow d_\downarrow^\dagger d_\downarrow$$

$$+ \sum_{k,\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \sum_{k,\sigma} V_k \left(d_\sigma^\dagger c_{k\sigma} + c_{k\sigma}^\dagger d_\sigma \right).$$

Kernel Ridge Regression

$$g_m(\mathbf{D}) = \sum_l \alpha_{lm} K_m(\mathbf{D}_l, \mathbf{D})$$

coefficients fixed by training set

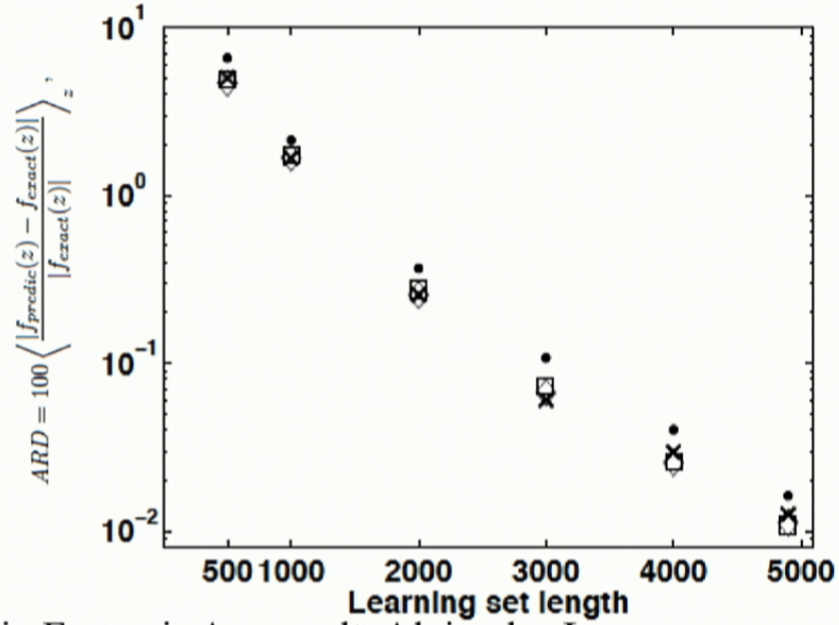
Louis-Francois Arsenault, Alejandro Lopez-Bezanilla, O. Anatole von Lilienfeld, A. J.M., Phys. Rev. B90 155136 (2014).

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



● Continued fraction

x z_n

□ Time

◇ Legendre

Learning set size ~1000 needed

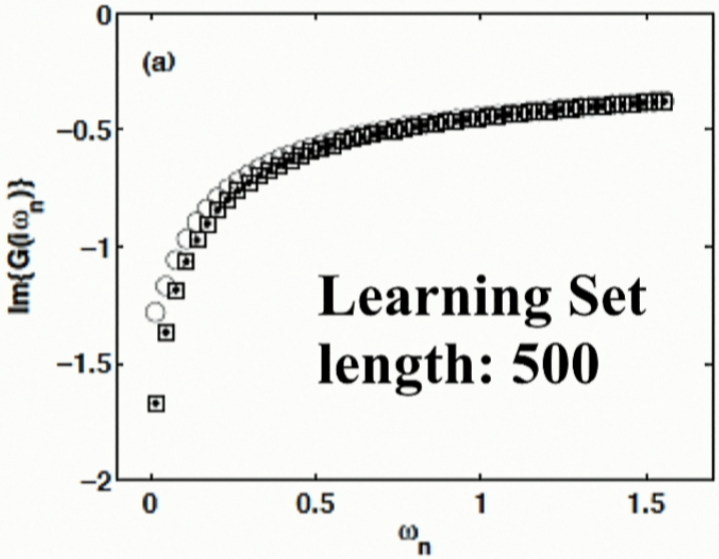
Louis-Francois Arsenault, Alejandro Lopez-Bezanilla, O. Anatole von Lilienfeld, A. J.M., Phys. Rev. B90 155136 (2014).

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



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

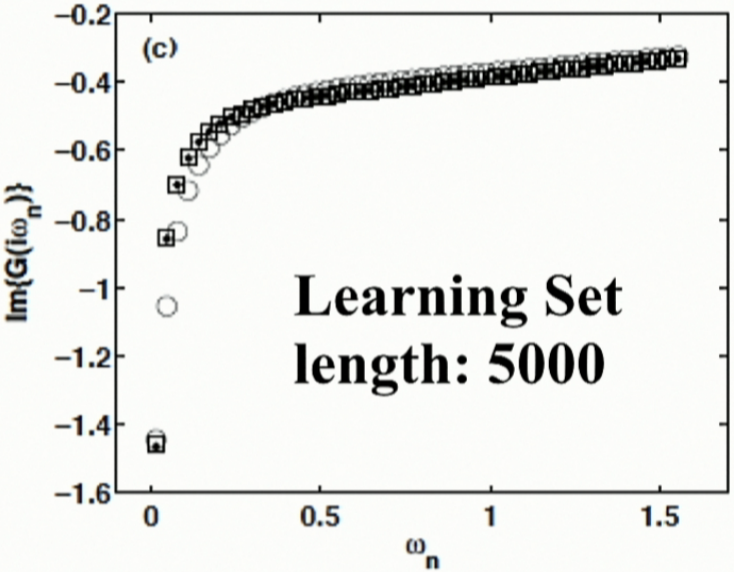
Louis-Francois Arsenault, Alejandro Lopez-Bezanilla, O. Anatole von Lilienfeld, A. J.M., Phys. Rev. B90 155136 (2014).

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



- ● Exact
- ○ ML result

Learning set size ~1000 needed

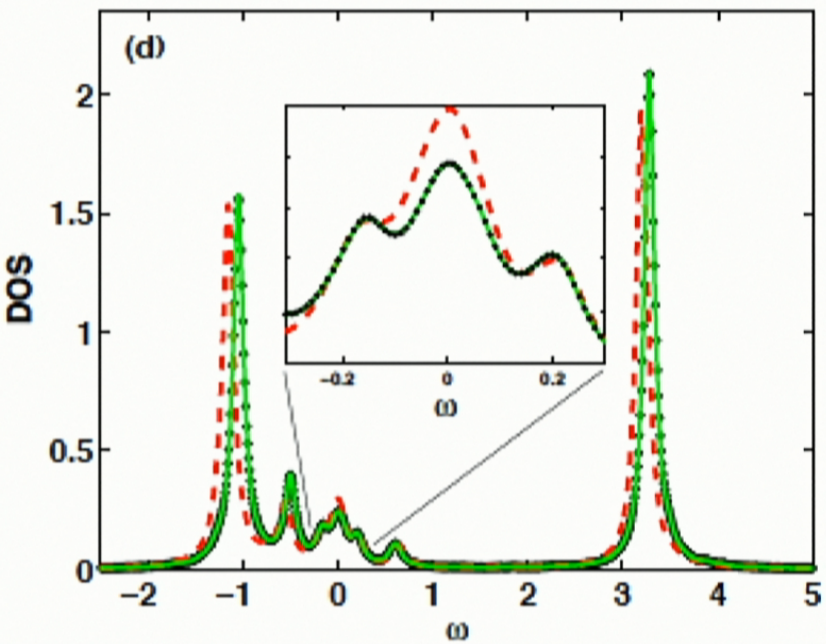
Louis-Francois Arsenault, Alejandro Lopez-Bezanilla, O. Anatole von Lilienfeld, A. J.M., Phys. Rev. B90 155136 (2014).

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



- Exact
- ML: 500
- ML: 5000

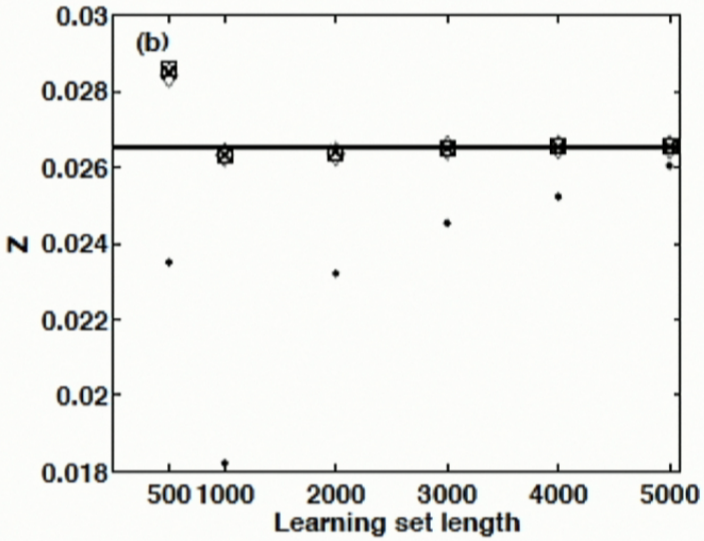
Bezanilla, O. Anatole von Lilienfeld, A. J.M., Phys. Rev. B90 155136 (2014).

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

The Anderson Model: Results

the quasiparticle weight Z



- ● **Continued fraction**
- X **z_n**
- □ **Time**
- ◇ **Legendre**

Learning set size >1000 needed

Legendre rep. preferred because most compact

Louis-Francois Arsenault, Alejandro Lopez-Bezanilla, O. Anatole von Lilienfeld, A. J.M., Phys. Rev. B90 155136 (2014).

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

The Anderson Model: Results

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

Louis-Francois Arsenault, Alejandro Lopez-Bezanilla, O. Anatole von Lilienfeld, A. J.M., Phys. Rev. B90 155136 (2014).

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

Learning a lattice many-body problem

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

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

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

Learning a lattice many-body problem

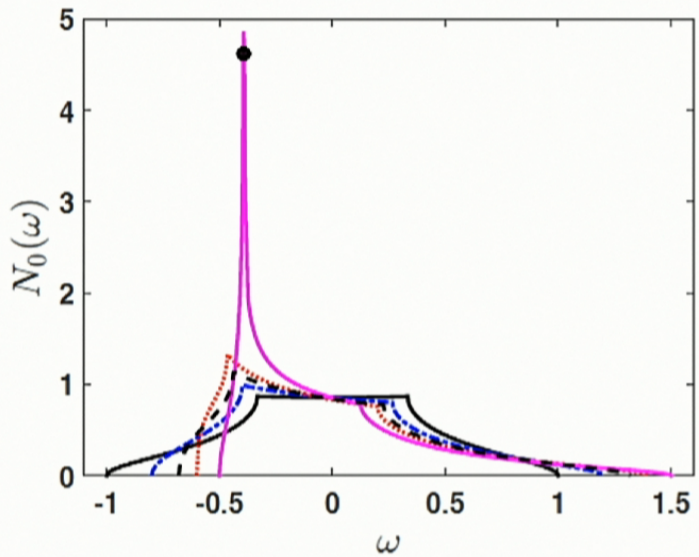
$$\mathbf{H} = \sum_{\mathbf{ij}\sigma} \mathbf{t_{ij}} \mathbf{c}^{\dagger}_{\mathbf{i}\sigma} \mathbf{c}_{\mathbf{j}\sigma} + \mathbf{U} \sum_{\mathbf{i}} \hat{\mathbf{n}}_{\mathbf{i}\uparrow} \hat{\mathbf{n}}_{\mathbf{i}\downarrow}$$

Descriptors:

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

Output:

- phase (metal or insulator)
- G (Legendre coeffs)



Copyright A. J. Millis 2016

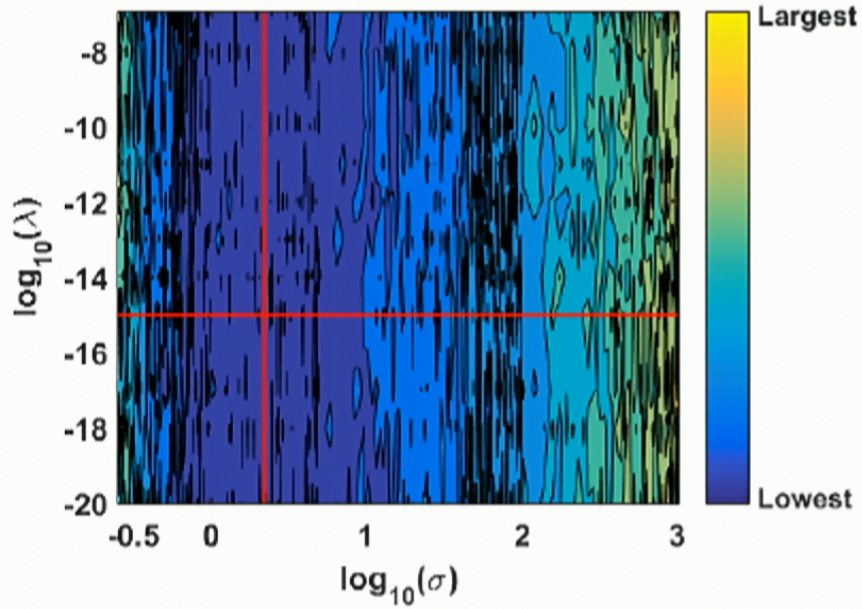
The Simons foundation
Department of Physics, Columbia University

Learning a lattice many-body problem

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

Technical improvement: cross validation

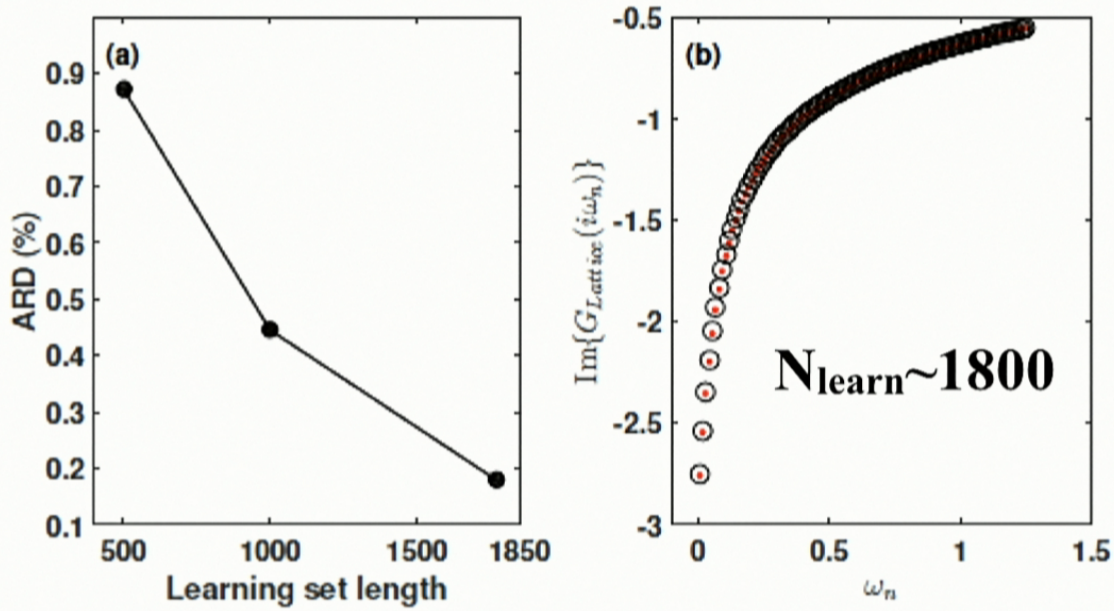
Pick ML hyperparameters to minimize error in 1st 5 Legendre coeffs of 1000 examples



Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

Cross-validation leads to much smaller errors



Training sets of ~few hundred are acceptable

The Simons foundation
Department of Physics, Columbia University

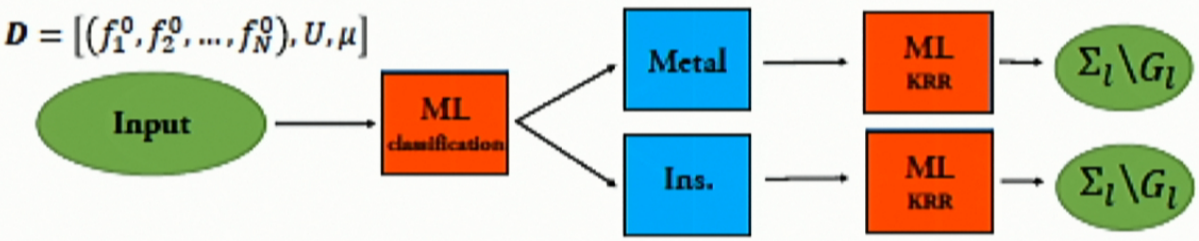
Copyright A. J. Millis 2016

Learning a lattice many-body problem

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

2 step learning:

- **Classification**
- **Greens function (self energy)**



Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

Classification

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

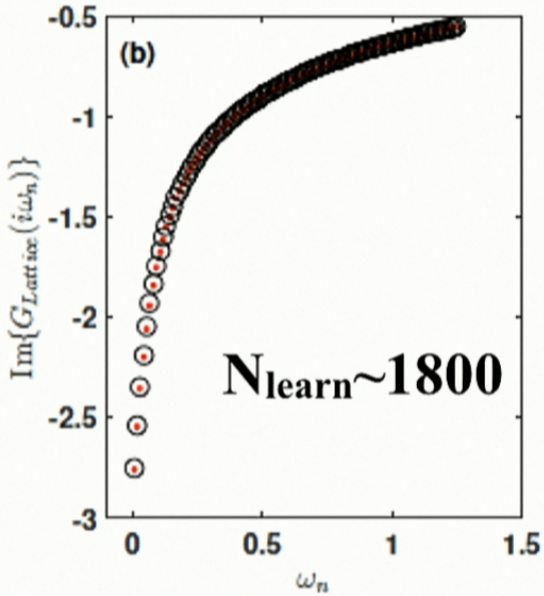
Classification: Almost perfectly accurate

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

Learning a lattice many-body problem

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



G reproduced perfectly

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

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

reverse engineering a lattice many-body problem

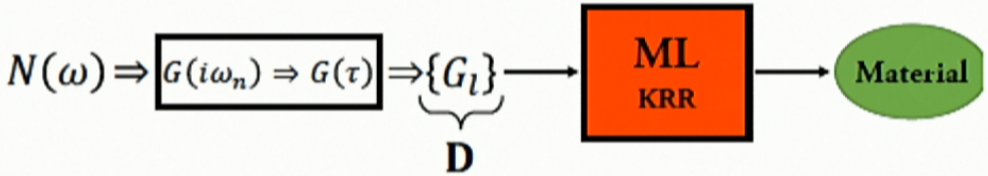
$$\mathbf{H} = \sum_{\mathbf{ij}\sigma} \mathbf{t_{ij}} \mathbf{c}^{\dagger}_{\mathbf{i}\sigma} \mathbf{c}_{\mathbf{j}\sigma} + \mathbf{U} \sum_{\mathbf{i}} \hat{\mathbf{n}}_{\mathbf{i}\uparrow} \hat{\mathbf{n}}_{\mathbf{i}\downarrow}$$

Question:

What t_{ij} should I use to get a desired many-body DOS?

Answer:

Run the `machine' backwards

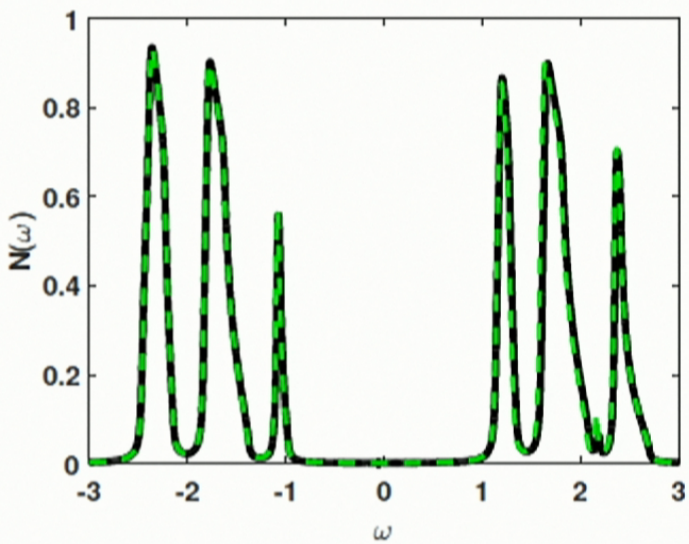


Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

reverse engineering a lattice many-body problem

$$N(\omega) \Rightarrow \boxed{G(i\omega_n) \Rightarrow G(\tau)} \Rightarrow \underbrace{\{G_l\}}_{\mathbf{D}} \longrightarrow \boxed{\begin{matrix}\text{ML}\\ \text{KRR}\end{matrix}} \longrightarrow \text{Material}$$



Target: gap of size 2. Method works with database of ~200 examples

Result

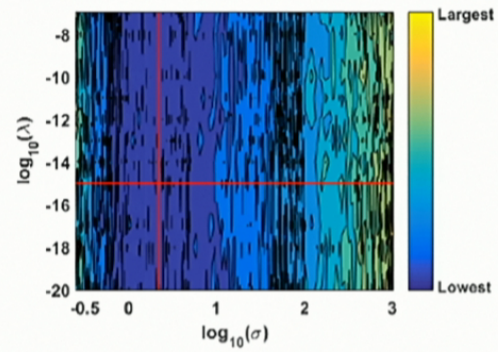
Target

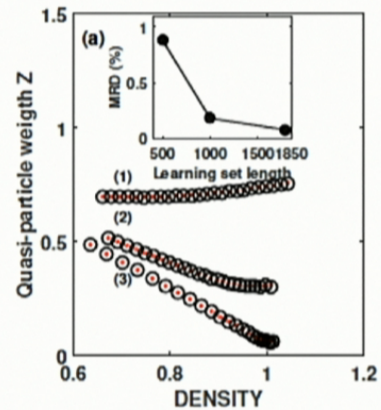
Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

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

Copyright A. J. Millis 2016

Coda: analytical continuation

$$\boxed{G(z) = \int \frac{dx}{\pi} \frac{A(x)}{z - x}}$$

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

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

you can reconstruct A.

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

The difficulty:

The operator

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

has many very small eigenvalues.
=> inversion is ill-posed.
Small errors in $G(z_n)$ lead to big errors in inferred A

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

Coda: analytical continuation

$$\boxed{G(z) = \int \frac{dx}{\pi} \frac{A(x)}{z - x}}$$

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

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

you can reconstruct A.

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

the standard solution: maximum entropy

Minimize

$$\mathbf{F} = \sum_{\mathbf{n}} |\mathbf{G}(\mathbf{z_n}) - \mathbf{K}(\mathbf{z_n}, \mathbf{x}) * \mathbf{A}(\mathbf{x})|^2 + \alpha \int \mathbf{dx}\ \mathbf{A}(\mathbf{x}) \ln \frac{\mathbf{A}(\mathbf{x})}{\mathbf{M}(\mathbf{x})}$$

with respect to A(x).

***Need `model function' M(x) and parameter alpha**
***No method for assessing errors.**

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

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**

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

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**

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

Technical point: representation of functions

Represent A by conformal coefficients

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

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

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

Learning and Tuning Sets

Many examples, constructed as sums of Gaussians

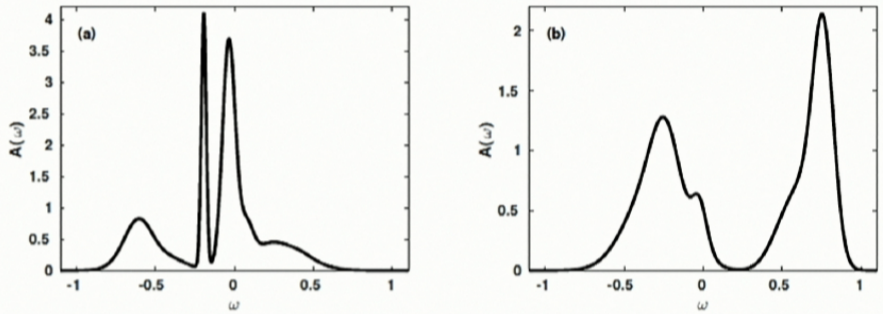
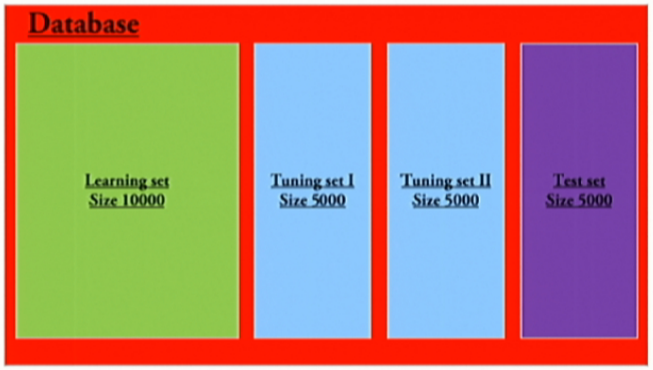


Figure 1: Two different (a) and (b) possible $A(\omega)$ in the database.



Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

Learning and Tuning Sets

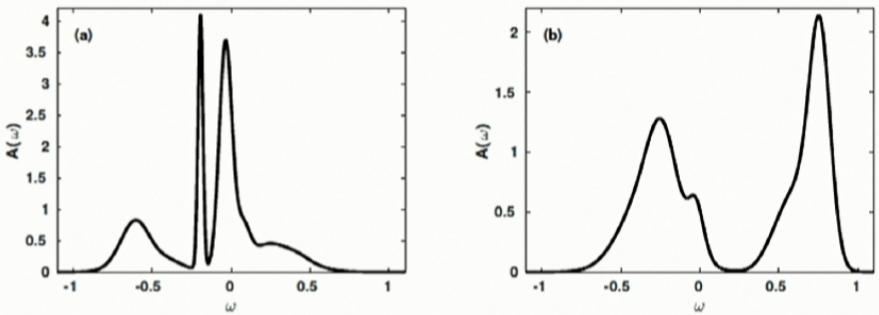
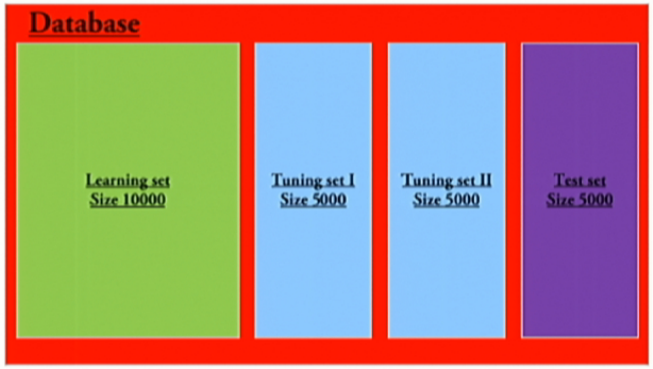


Figure 1: Two different (a) and (b) possible $A(\omega)$ in the database.



Note: learning involves fitting directly to real frequencies

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

Learning and Tuning Sets

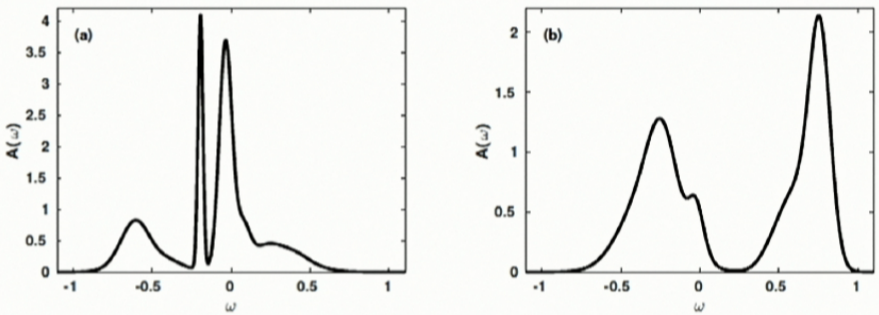
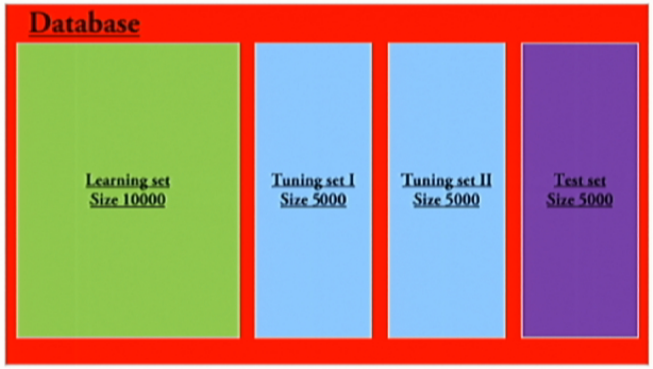


Figure 1: Two different (a) and (b) possible $A(\omega)$ in the database.



Note: learning involves fitting directly to real frequencies

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

Learning and Tuning Sets

Many examples, constructed as sums of Gaussians

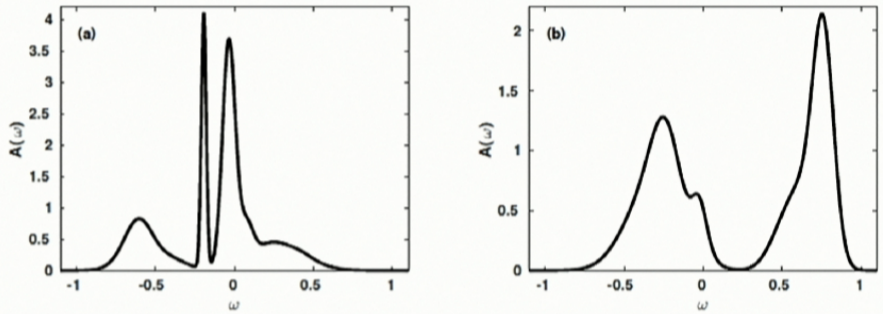
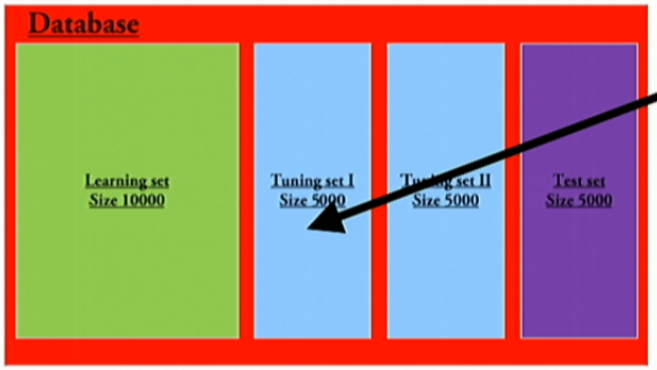


Figure 1: Two different (a) and (b) possible $A(\omega)$ in the database.

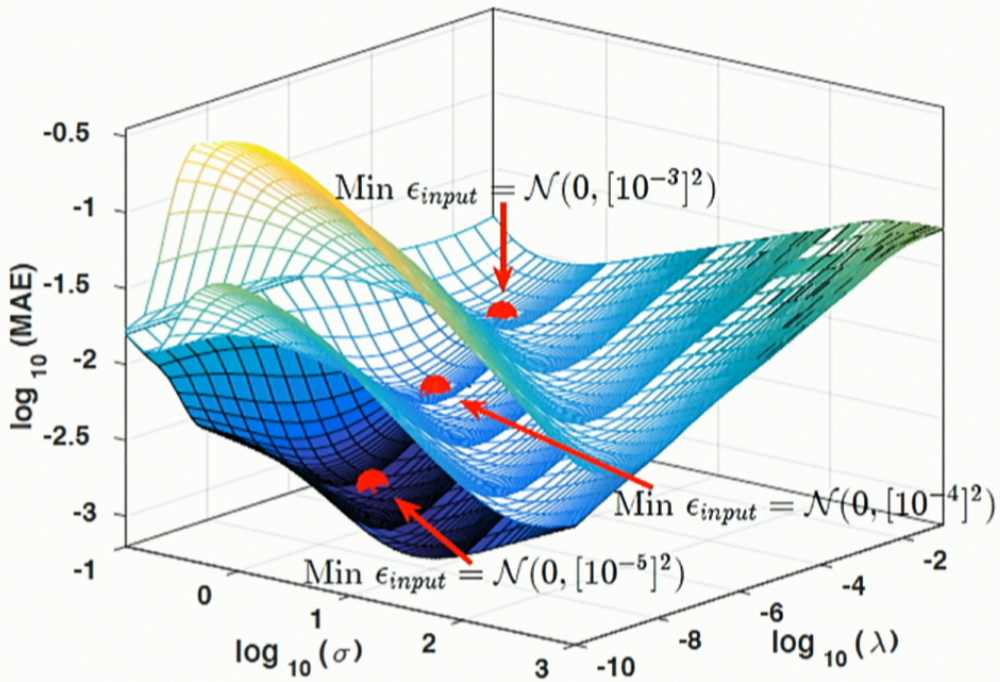
Tune hyperparameters



Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

Tuning the hyperparameters (1st level regression)



The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016

Learning and Tuning Sets

Many examples, constructed as sums of Gaussians

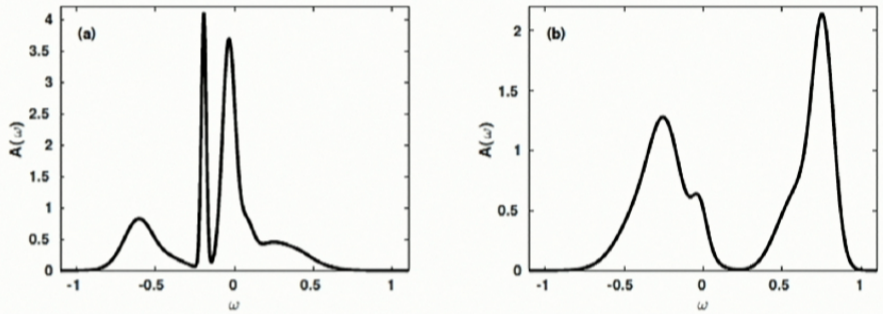
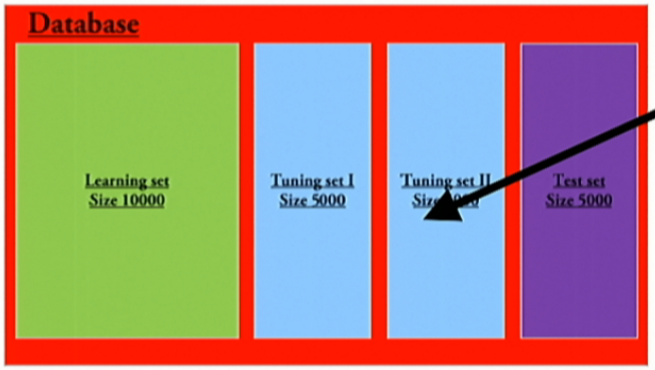


Figure 1: Two different (a) and (b) possible $A(\omega)$ in the database.



Tune covariance matrix (projection and error estimates)

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

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

$$\hat{A}_{\text{projected}} = \underset{\{A^* : A^* \text{ satisfies constraints (22) to (25)}\}}{\arg\min} \|A^* - \hat{A}\|_{\mathcal{M}}.$$

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

Subtlety: some regions of frequency space are better determined than others

=> define 'distance' via (principal components of) empirical covariance matrix

$$\| \bullet \|_{\mathcal{M}} = \bullet^{\mathsf{T}} \Sigma^{-1} \bullet$$

chosen by training

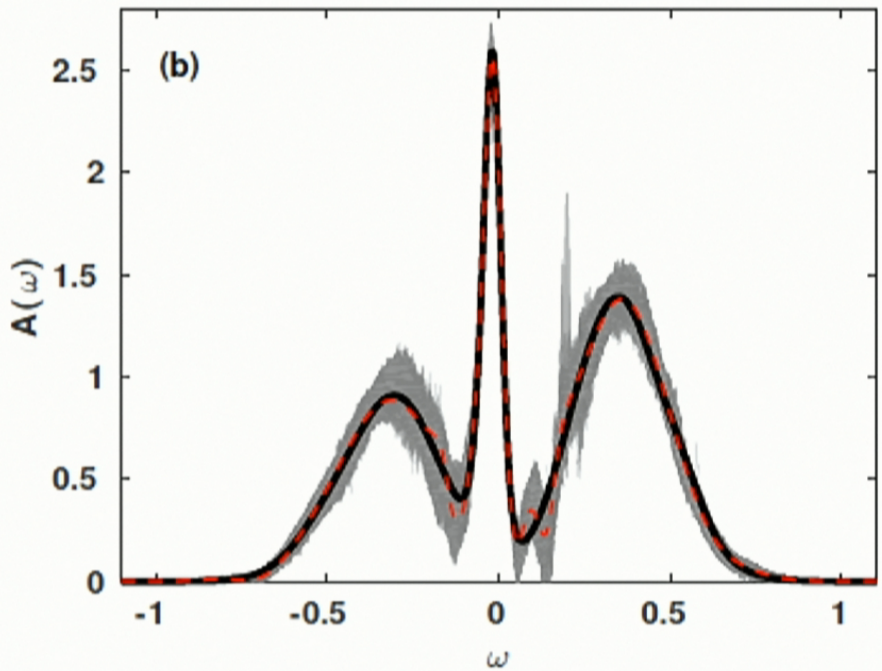
$$\hat{\Sigma} = \sum_{k=1}^{p} \lambda_k \mathrm{v}_k \mathrm{v}_k^{\mathsf{T}} + \mathrm{diag}\Big(\sum_{k=p+1}^{|\{\omega\}|} \lambda_k \mathrm{v}_k \mathrm{v}_k^{\mathsf{T}} \Big)$$

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'

Copyright A. J. Millis 2016

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)

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

Comparison to MaxEnt: modest noise

Red: statistical regression has lower error

Blue: Maxent has lower error

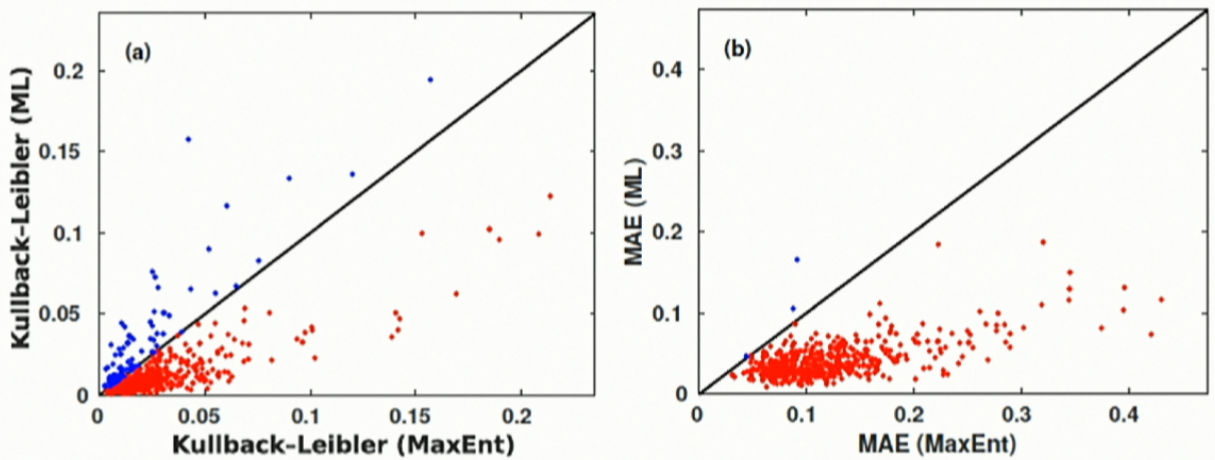


Figure 5: (Color online) Error for $\epsilon = \mathcal{N}(0, [10^{-3}]^2)$.

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

Comparison to MaxEnt: modest noise

Red: statistical regression has lower error

Blue: Maxent has lower error

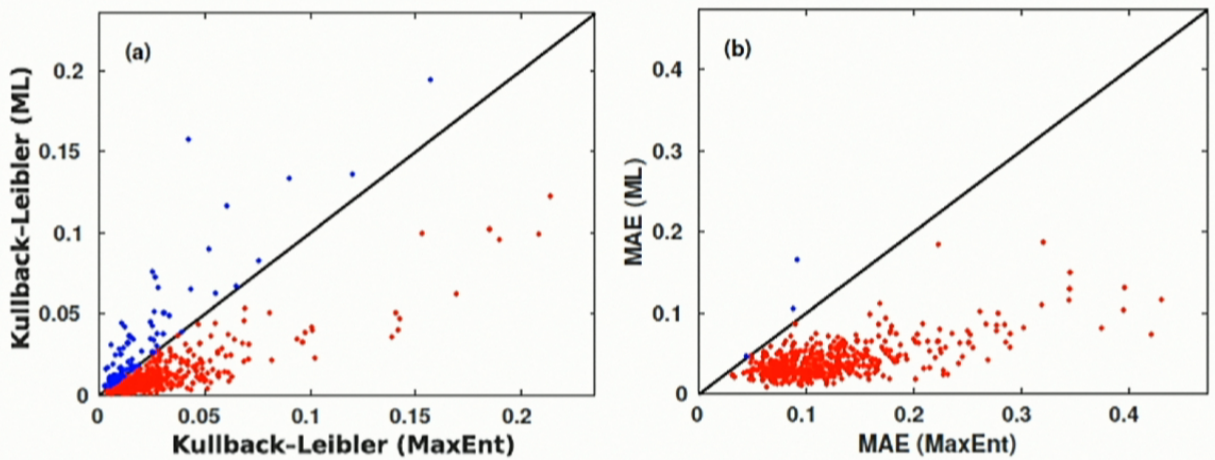


Figure 5: (Color online) Error for $\epsilon = \mathcal{N}(0, [10^{-3}]^2)$.

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University

Tiny noise (10^{-5})

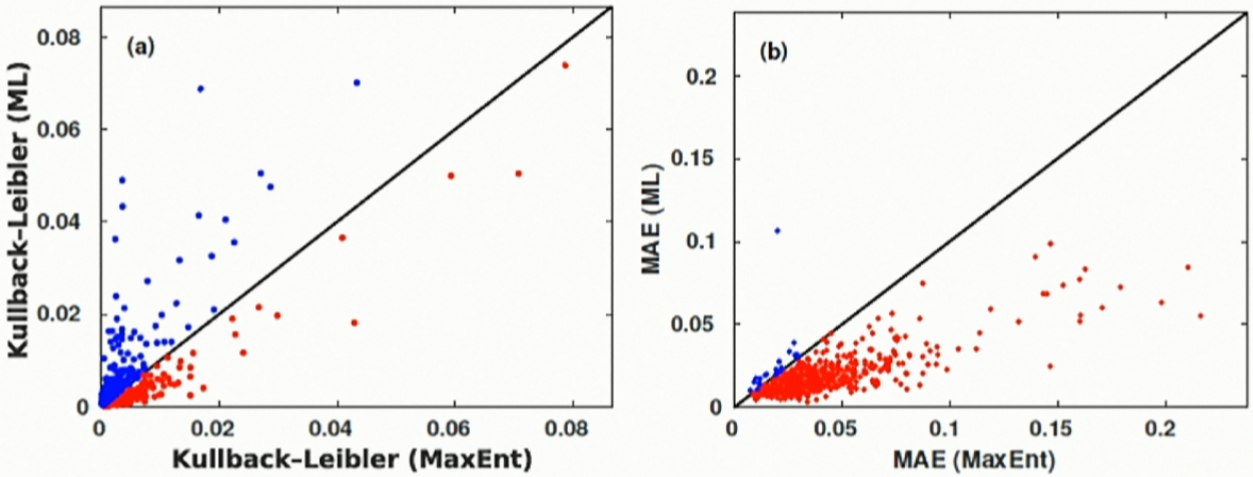


Figure 7: (Color online) Error for $\epsilon = \mathcal{N}(0, [10^{-5}]^2)$.

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

The Simons foundation
Department of Physics, Columbia University

Copyright A. J. Millis 2016

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

Copyright A. J. Millis 2016

The Simons foundation
Department of Physics, Columbia University