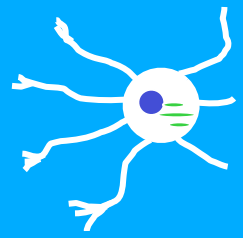
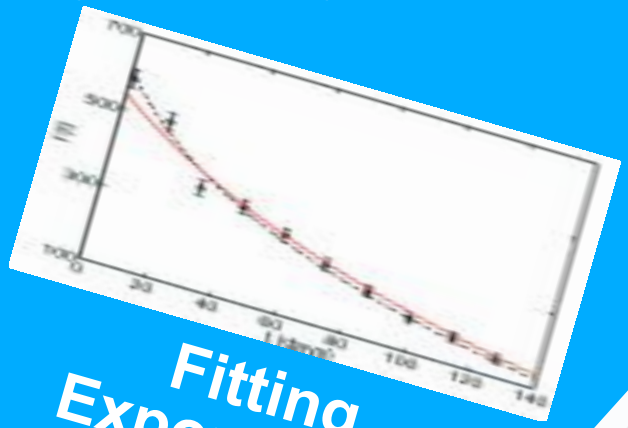
Sloppy Models, Differential geometry, and How science works

JPS, Katherine Quinn, Archishman Raju, Mark Transtrum, Ben Machta, Ricky Chachra, Kevin Brown, Ryan Gutenkunst, Josh Waterfall, Fergal Casey, Chris Myers, …

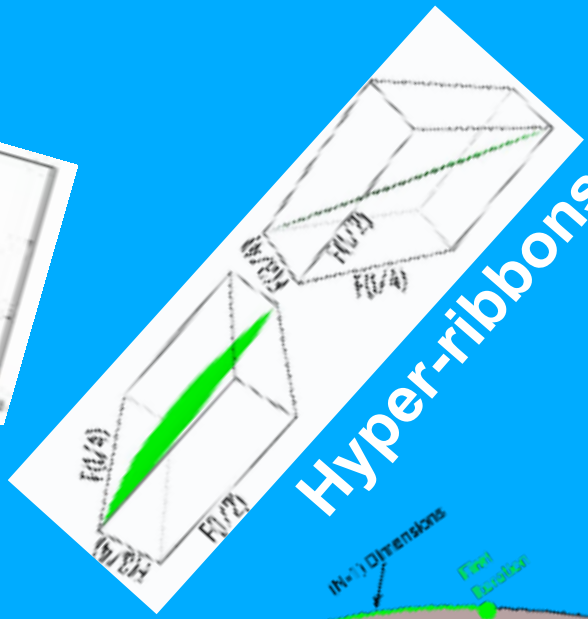


Cell Dynamics

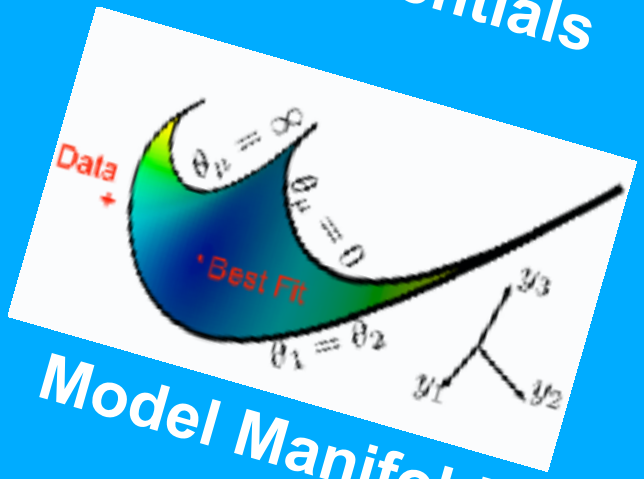
Sloppiness



Fitting Exponentials



Hyper-ribbons



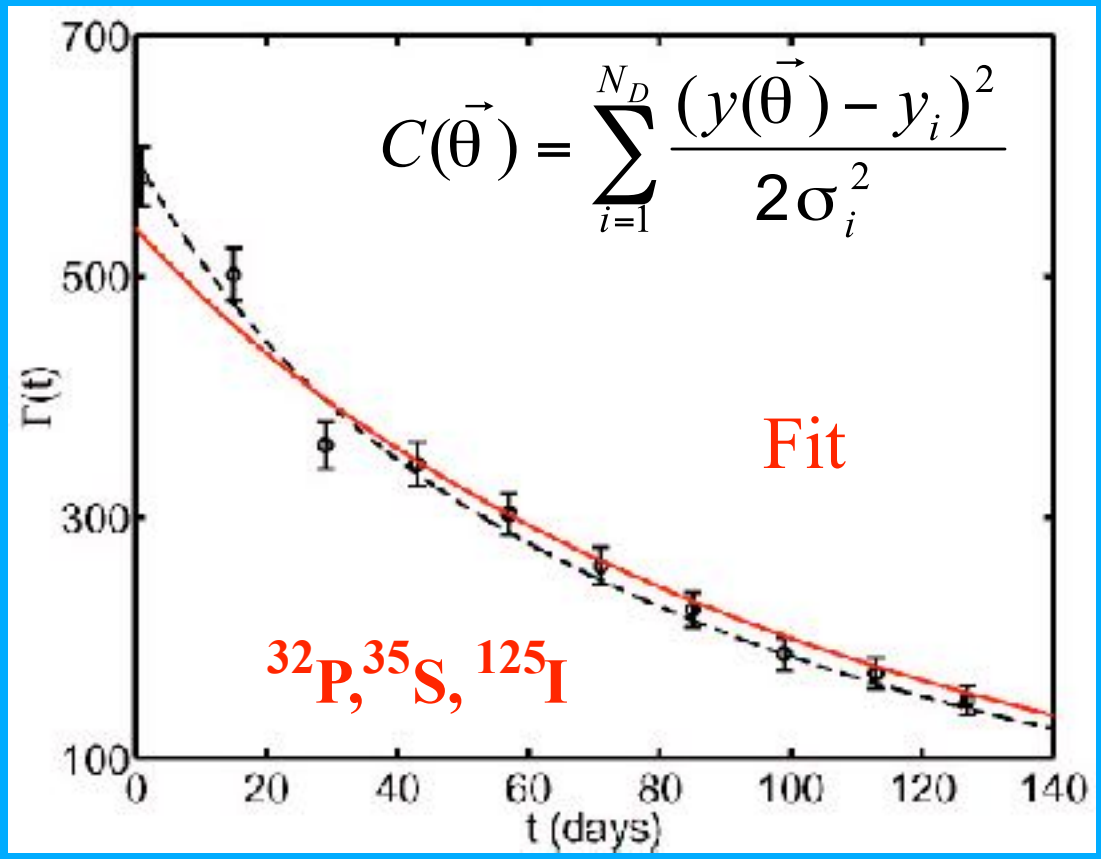
Model Manifold

Coarse-Grained Models

Fitting Decaying Exponentials

Classic ill-posed inverse problem

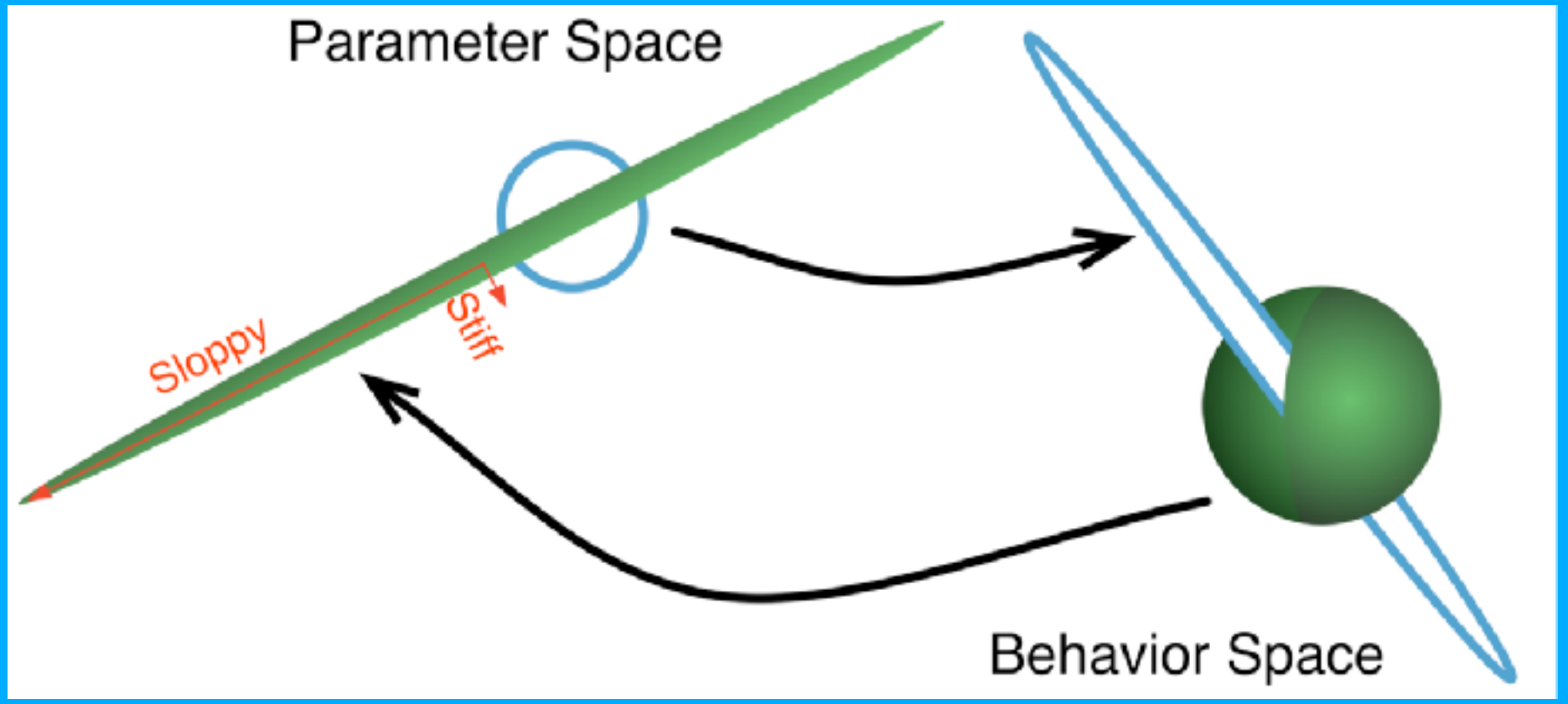
Given Geiger counter measurements from a radioactive pile, can we recover the identity of the elements and/or predict future radioactivity? Good fits with bad decay rates!



$$y(A,\gamma,t) = A_1 e^{-\gamma_1 t} + A_2 e^{-\gamma_2 t} + A_3 e^{-\gamma_3 t}$$

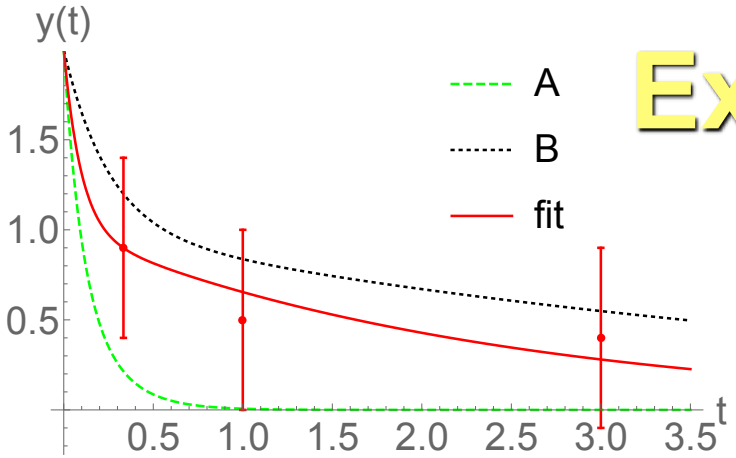
6 Parameter Fit

Models: Predictions about Data

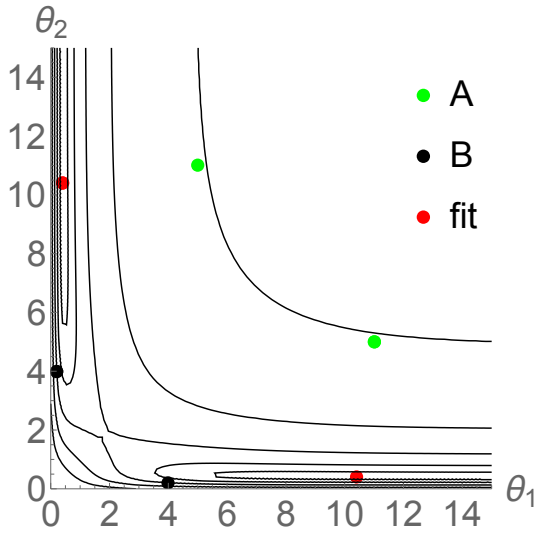


Scientific model: Predictions about behavior depend on physical constants (parameters) in the model.
Sloppiness: the behavior only depends on a few stiff parameter combinations.

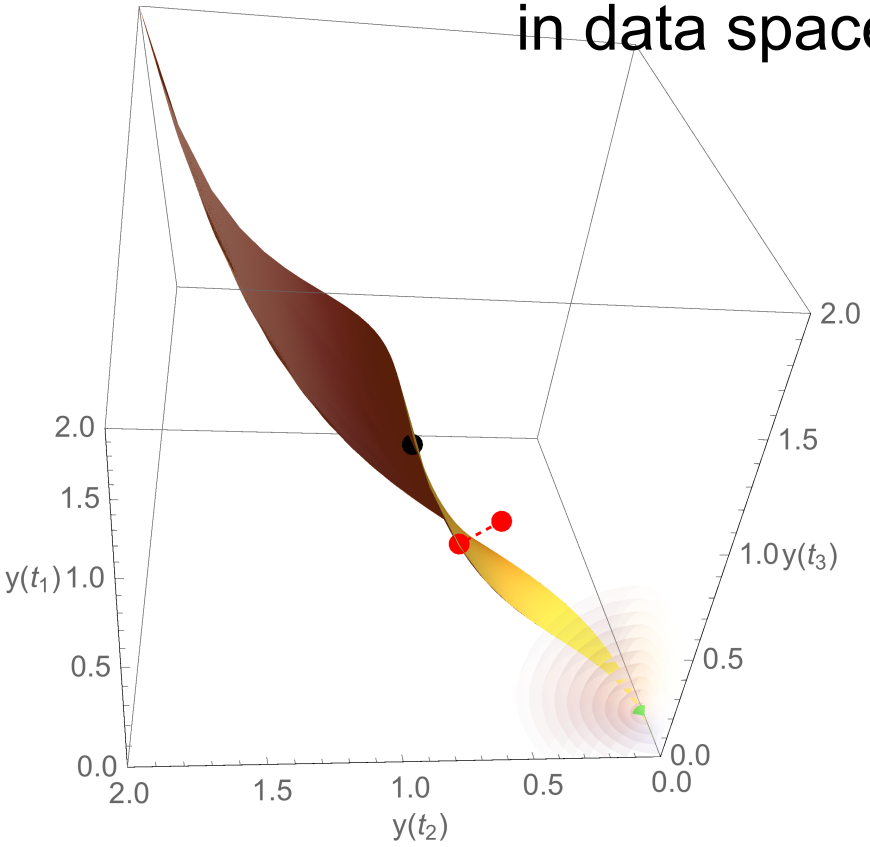
Exercise 1.14a



Two exponentials fit to data



Model manifold in data space



$$C(\theta) = \chi^2/2 = \sum_i (y_i(\theta) - d_i)^2/2\sigma_i^2$$

$$P(\theta) \propto \exp(-C(\theta))$$

[i] C is half the squared distance in data space. What is the metric g_{ij}?
[ii] How is the cost related to the log likelihood, if the errors are Gaussian?
[iii] View C as a Hamiltonian. What is the temperature?

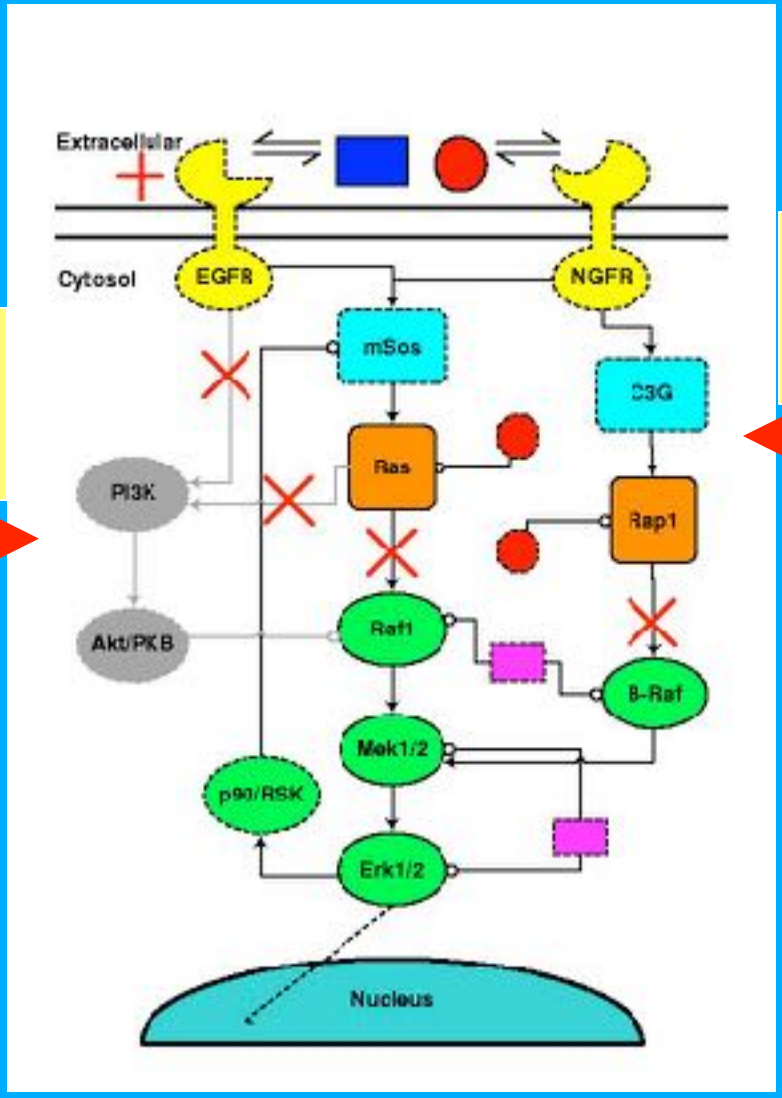
PC12 Differentiation

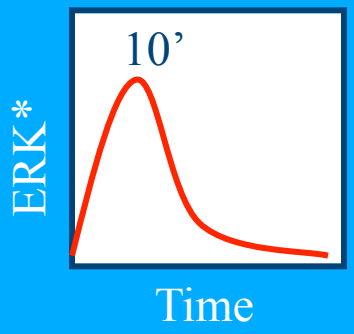
+EGF

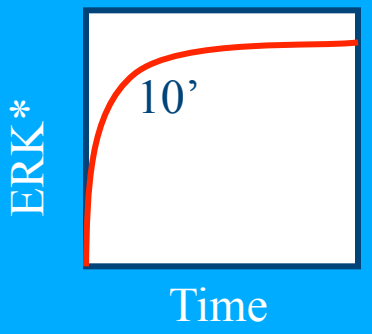
+NGF

Tunes down signal (Raf-1)

Pumps up signal (Mek)

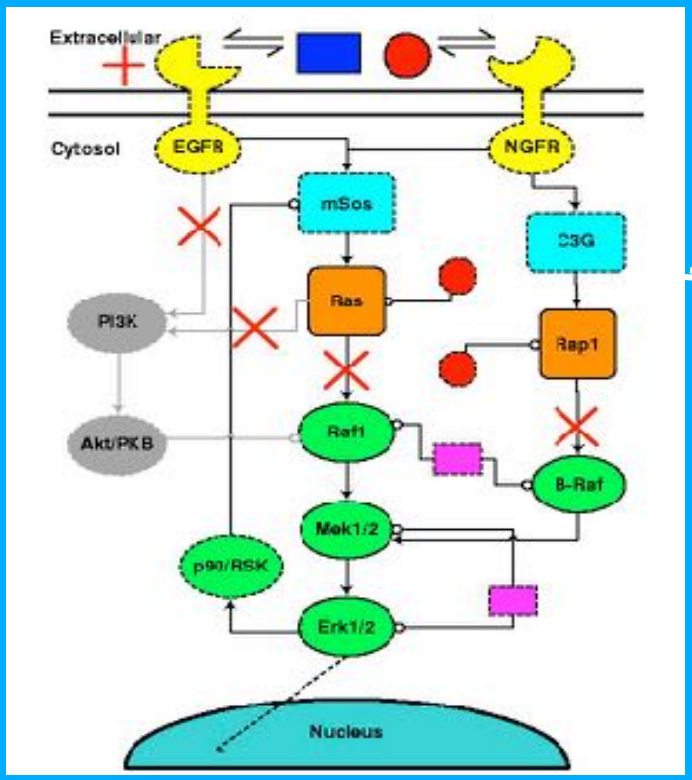


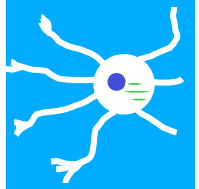




Biologists study which proteins talk to which. Modeling?

Systems Biology: Cell Protein Reactions





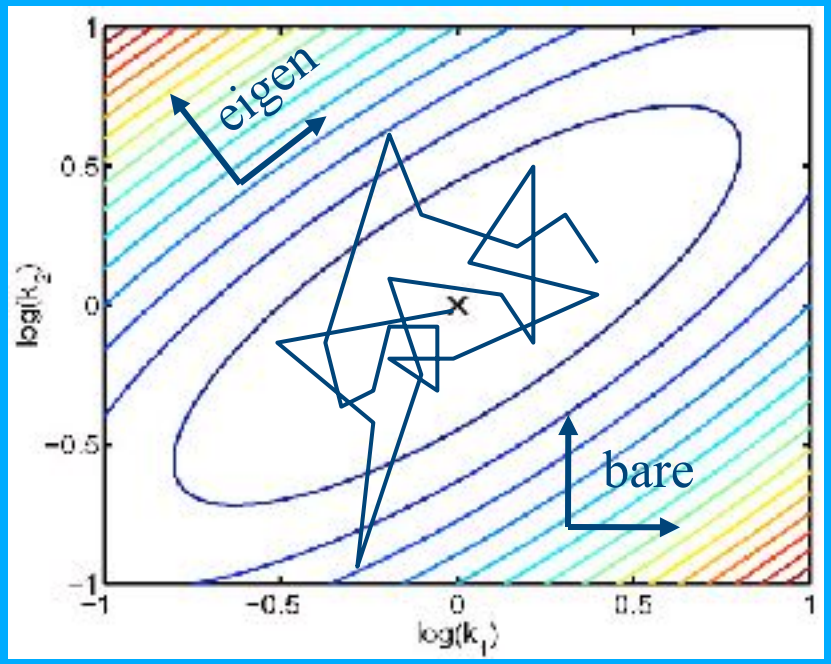
$$\frac{d\,[\mathrm{SosActive}]}{dt} = +k_{\mathrm{EGF}}\,[\mathrm{boundEGFR}]\,\frac{[\mathrm{SosInactive}]}{[\mathrm{SosInactive}] + K_{\mathrm{mEGF}}} + k_{\mathrm{NGF}}\,[\mathrm{boundNGFR}]\,\frac{[\mathrm{SosInactive}]}{[\mathrm{SosInactive}] + K_{\mathrm{mNGF}}} - k_{\mathrm{dSos}}\,[\mathrm{P90RskActive}]\,\frac{[\mathrm{SosActive}]}{[\mathrm{SosActive}] + K_{\mathrm{mdSos}}}$$

48 Parameter Fit!

Ensemble of Models

We want to consider not just minimum cost fits, but all parameter sets consistent with the available data. New level of abstraction: *statistical mechanics in model space*.

Don't trust predictions that vary



$$H_{ij} = \partial^2 C / \partial\theta_i \partial\theta_j$$

Cost is least-squares fit

$$C(\vec{\theta}) = \frac{1}{2}\sum_{i=1}^{N_D} \frac{(y(\vec{\theta}) - y_i)^2}{\sigma_i^2}$$

Boltzmann weights exp(-C/T)

$$\langle O \rangle = \frac{1}{N_E}\sum_{i=1}^{N_E} O(\vec{\theta}_i)$$

$$\sigma_O^2 = \left\langle O^2(\vec{\theta}) \right\rangle - \left\langle O(\vec{\theta}) \right\rangle^2$$

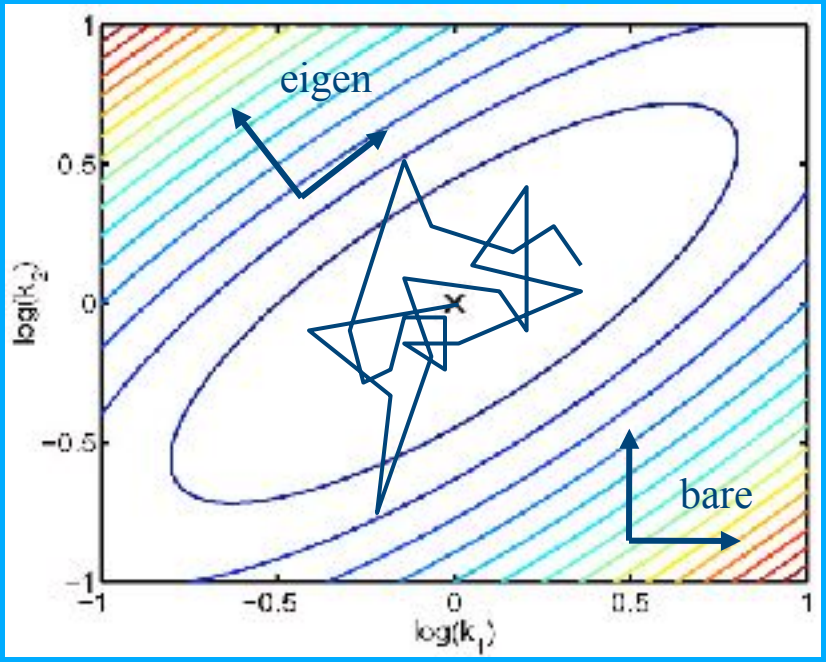
O is chemical concentration $y(t_i)$, or rate constant θ_n…

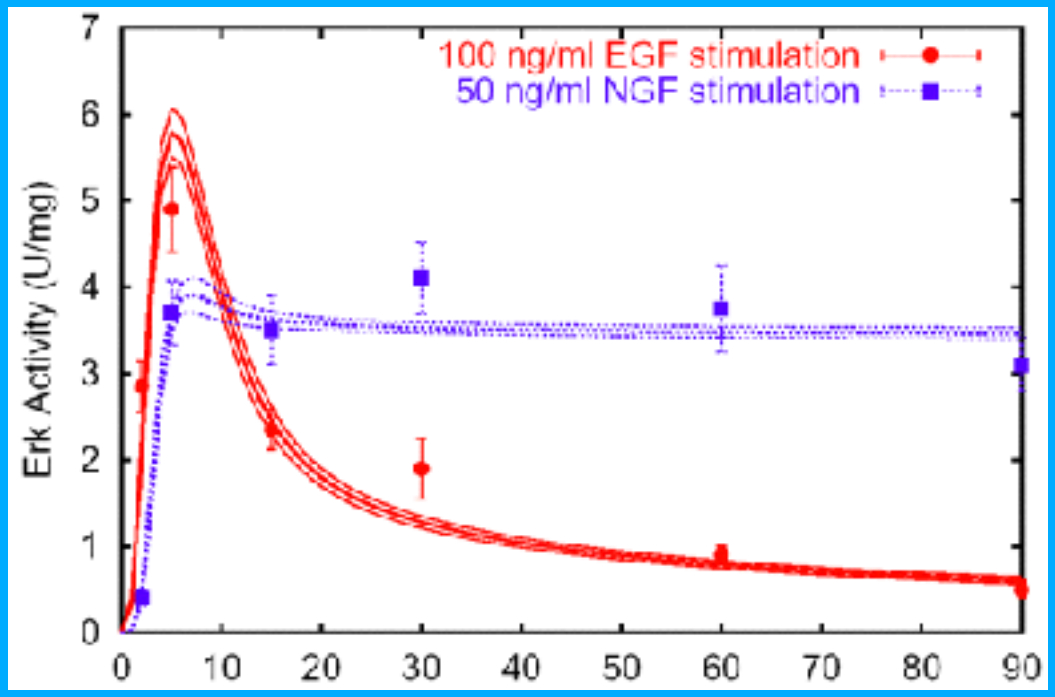
48 Parameter "Fit" to Data

Cost is Energy

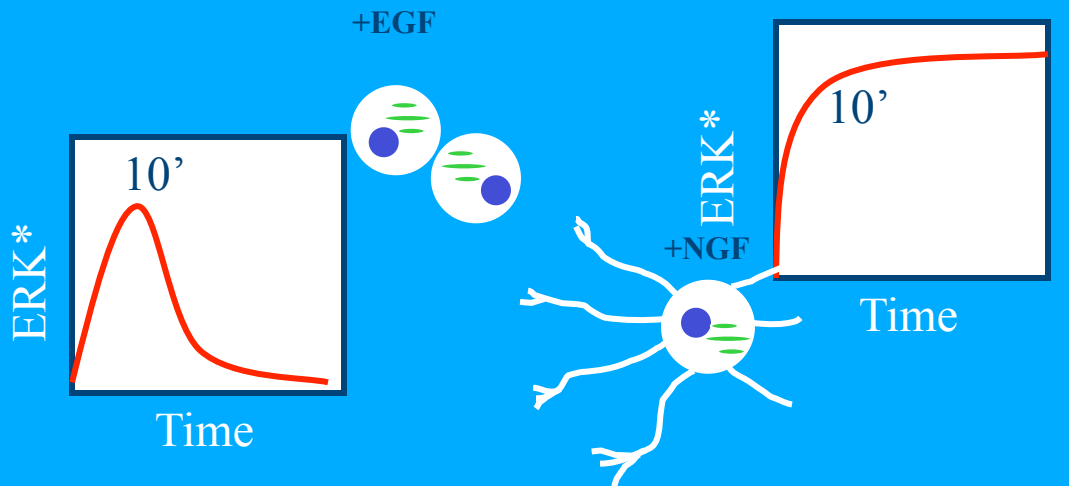
$$C(\vec{\theta}) = \frac{1}{2}\sum_{i=1}^{N_D}\frac{(y(\vec{\theta}) - y_i)^2}{\sigma_i^2}$$

Ensemble of Fits Gives Error Bars

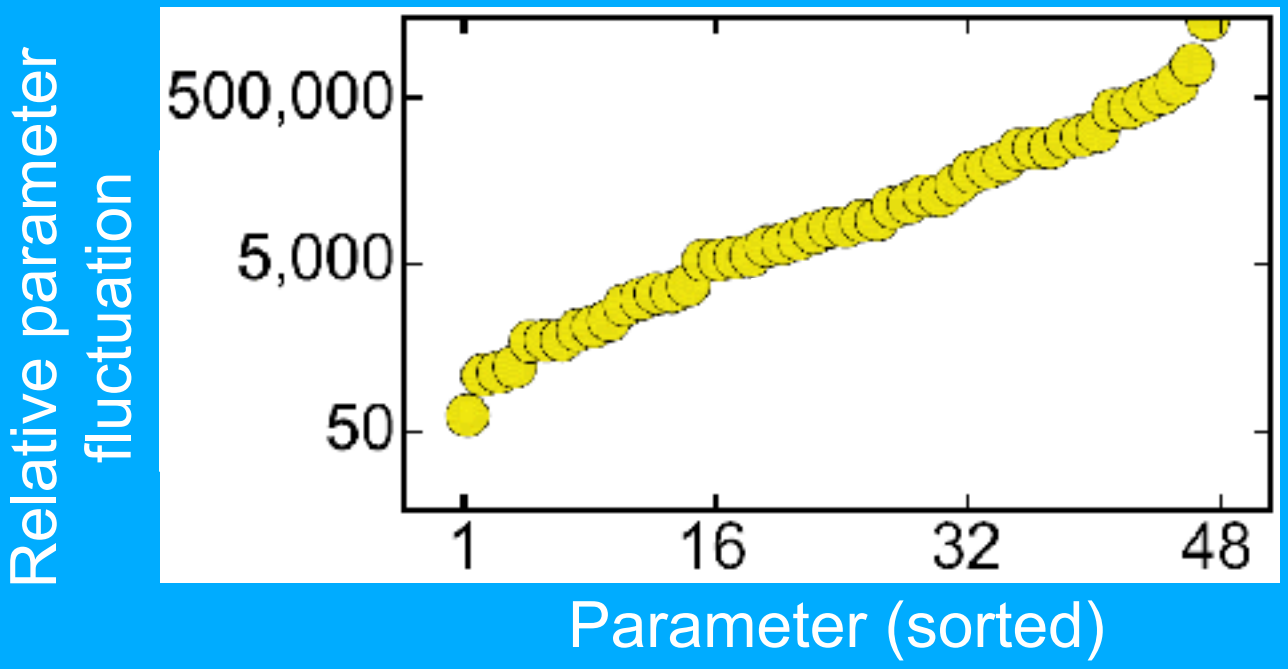




Error Bars from Data Uncertainty

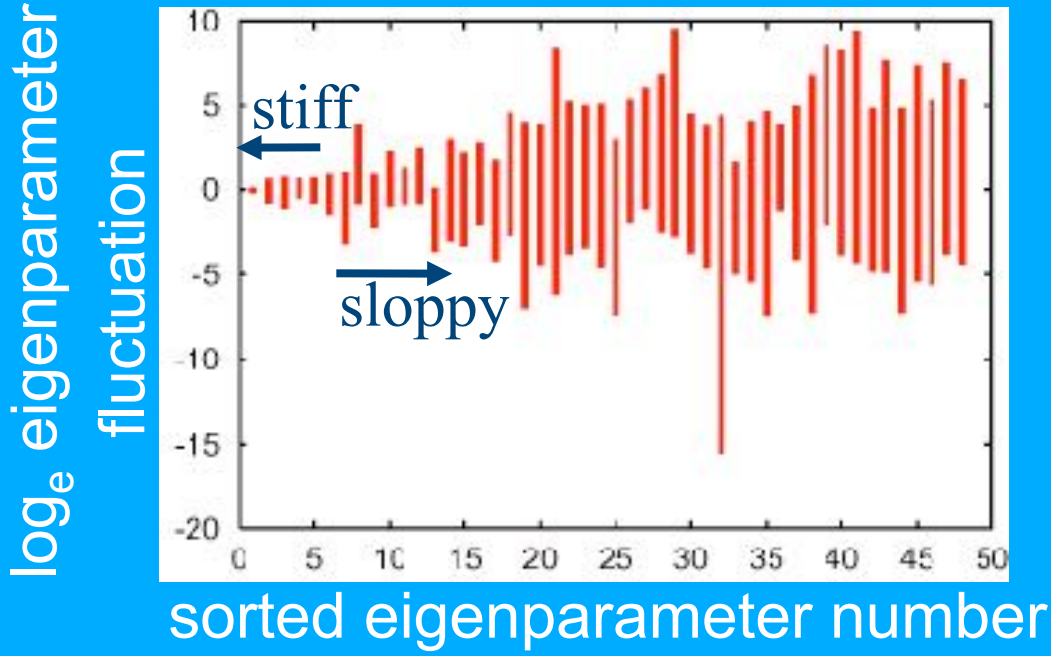


Parameters Fluctuate over Enormous Range

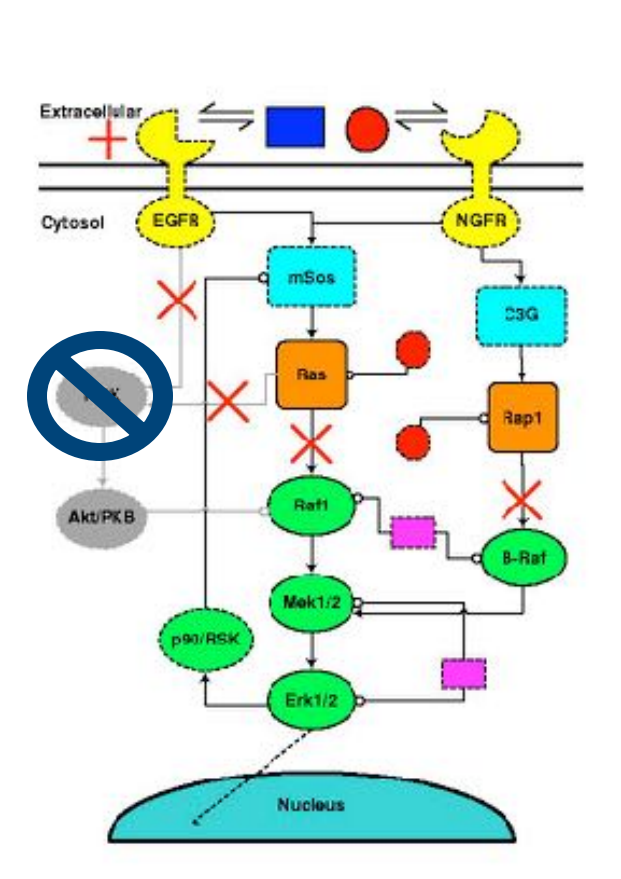


- All parameters vary by minimum factor of 50, some by a million
- Not robust: four or five "stiff" linear combinations of parameters; 44 sloppy

Are predictions possible?

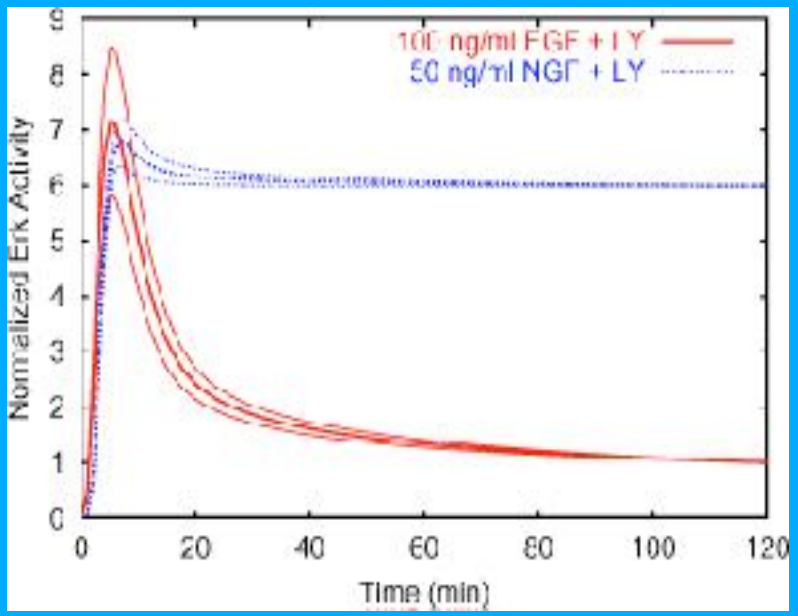


Predictions are Possible

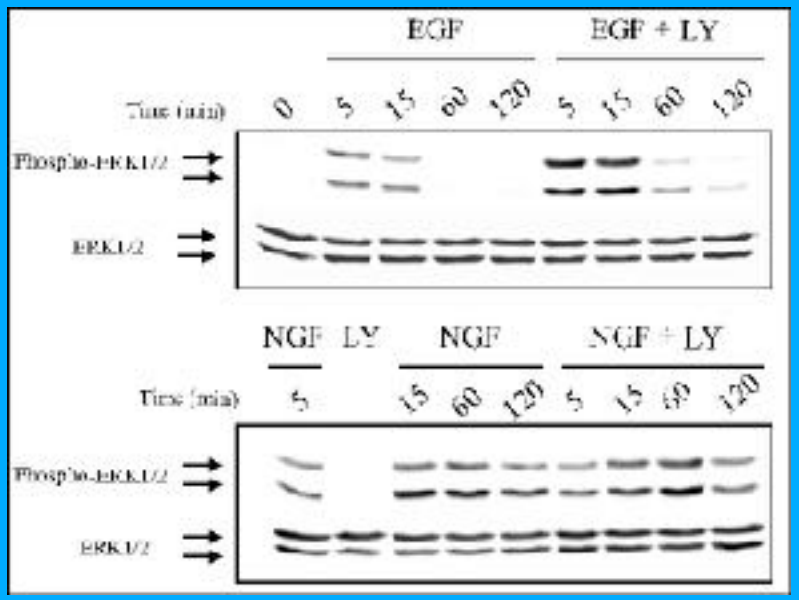


Model predicts that the left branch isn't important

Model Prediction

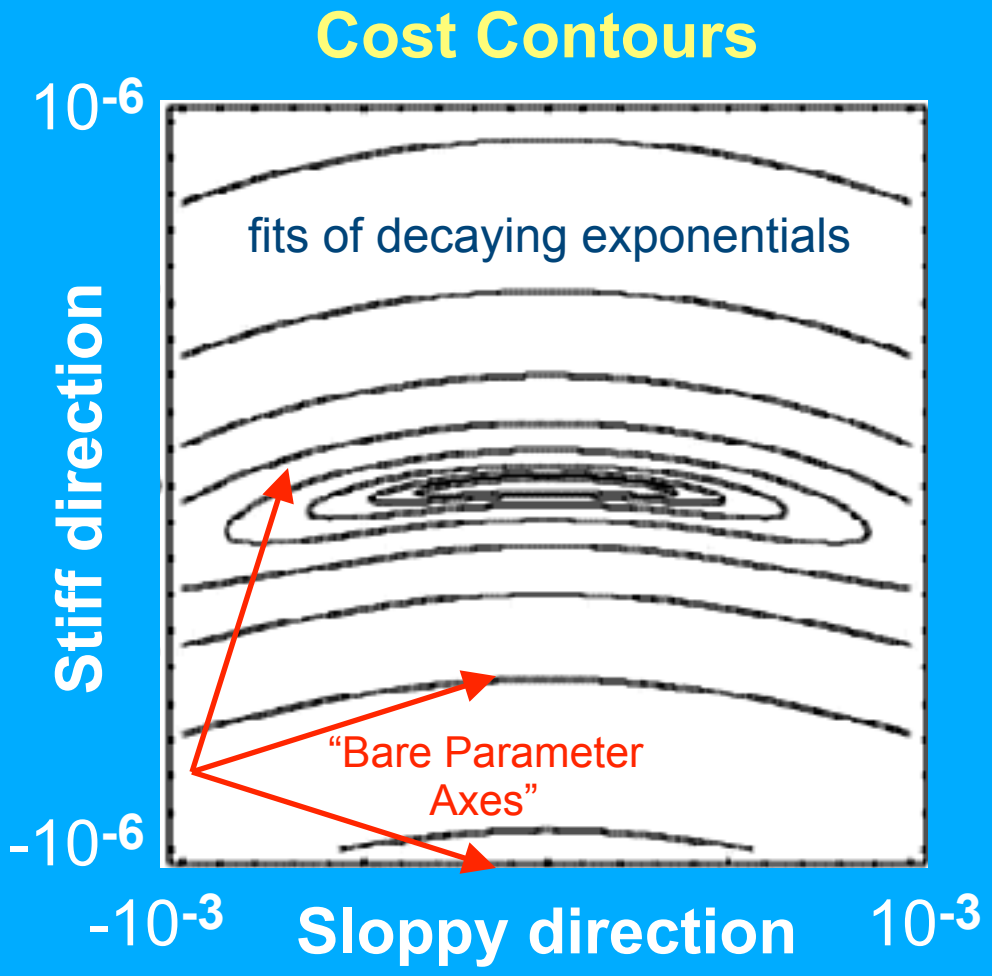


Brown's Experiment



Parameters fluctuate orders of magnitude, but still predictive!

Parameter Indeterminacy and Sloppiness



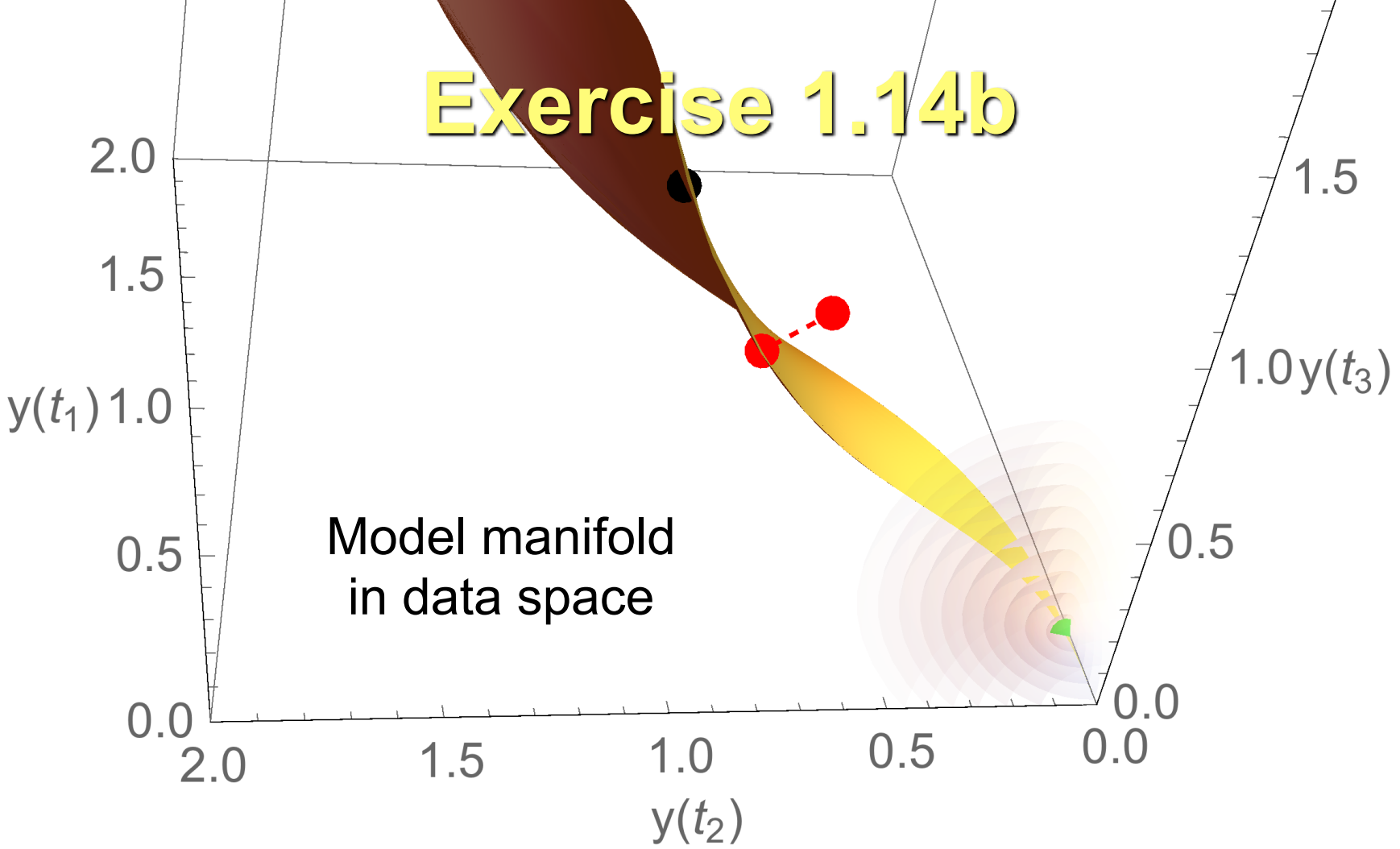
~5 stiff, ~43 sloppy directions

Note: Horizontal scale shrunk by 1000 times
Aspect ratio = Human hair

48 parameter fits are sloppy: Many parameter sets give almost equally good fits

A few 'stiff' constrained directions allow model to remain predictive

Exercise 1.14b



$$C(\theta) = \chi^2/2 = \sum_i (y_i(\theta) - d_i)^2/2\sigma_i^2 \qquad P(\mathbf{d}\,|\,\theta) \propto \exp(-C(\theta))$$

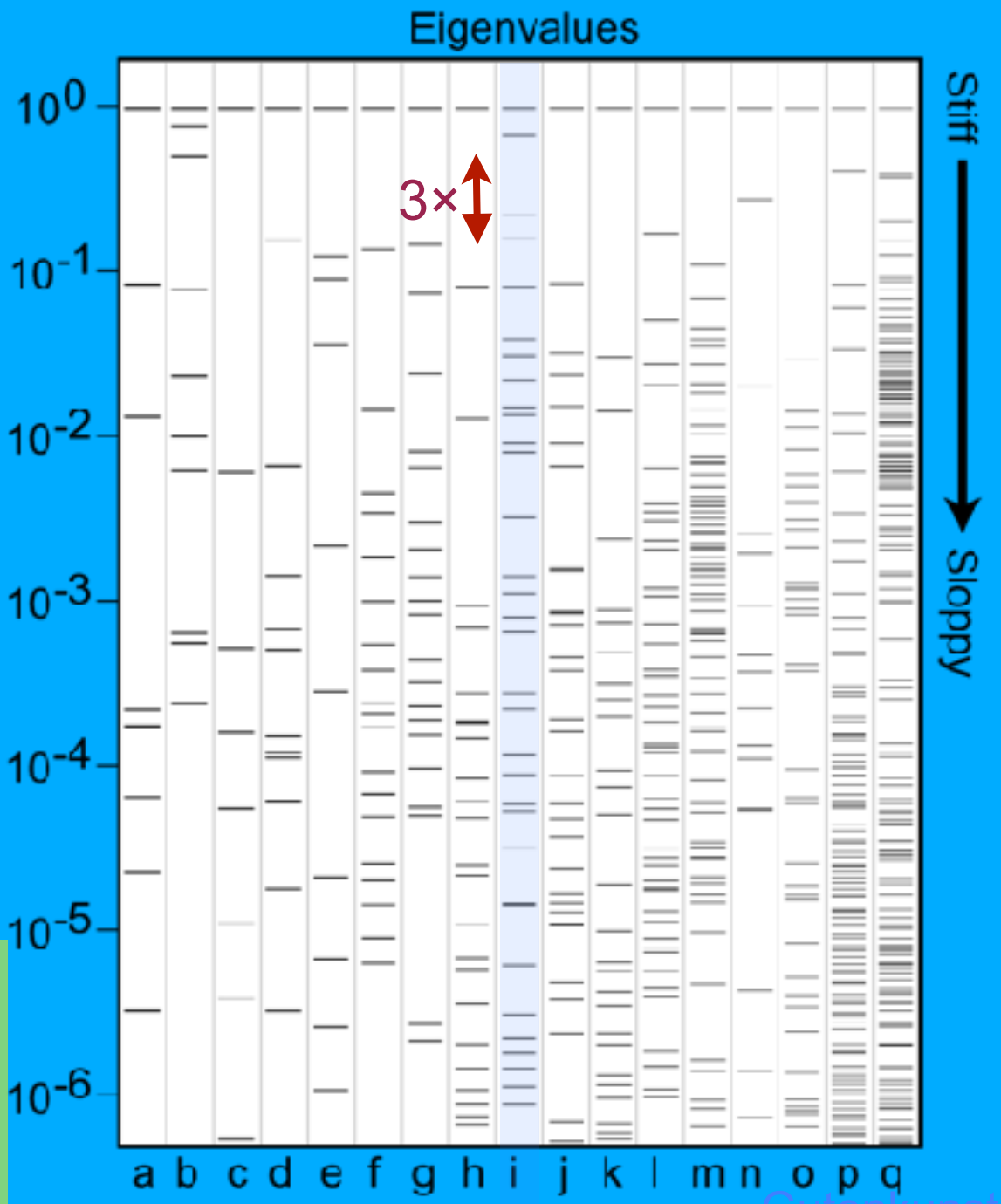
[i] Show $P(\mathbf{d}|\theta)$ is a blurred Gaussian sphere in data space.
[ii] Make an analogy to the momentum distribution of classical particles with different masses.

Systems Biology

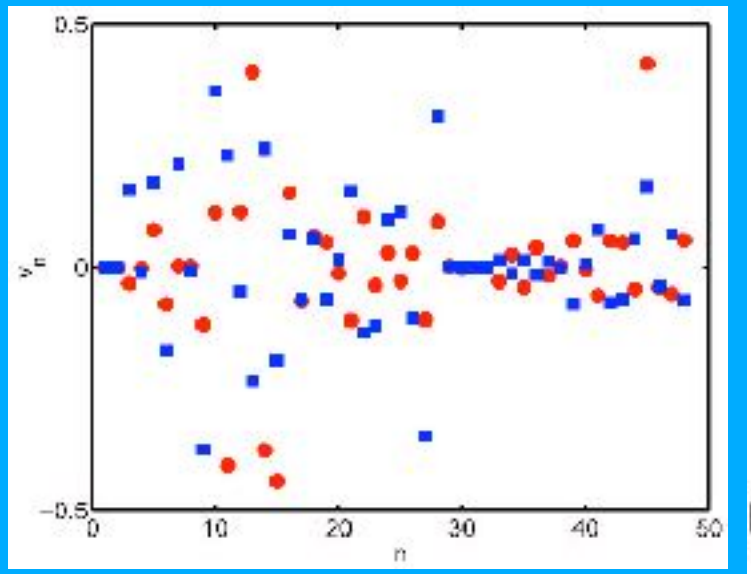
Seventeen models

(a) eukaryotic cell cycle
(b) Xenopus egg cell cycle
(c) eukaryotic mitosis
(d) generic circadian rhythm
(e) nicotinic acetylcholine intra-receptor dynamics
(f) generic kinase cascade
(g) Xenopus Wnt signaling
(h) Drosophila circadian rhythm
(i) rat growth-factor signaling
(j) Drosophila segment polarity
(k) Drosophila circadian rhythm
(l) Arabidopsis circadian rhythm
(m) in silico regulatory network
(n) human purine metabolism
(o) Escherichia coli carbon metabolism
(p) budding yeast cell cycle
(q) rat growth-factor signaling

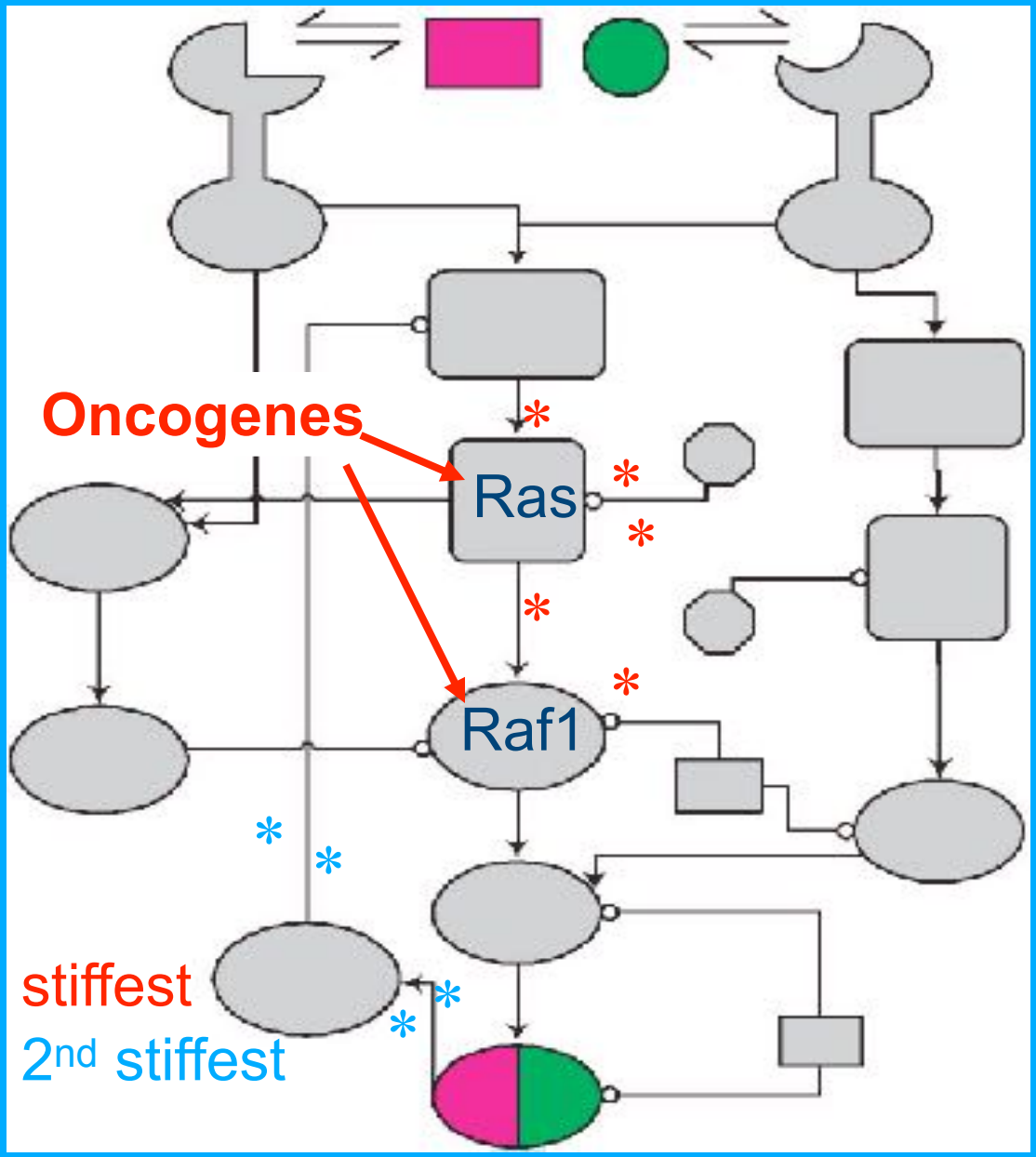
Enormous Ranges of Eigenvalues
(3^{48} is a big number)
Sloppy Range ~ $\sqrt{\lambda}$



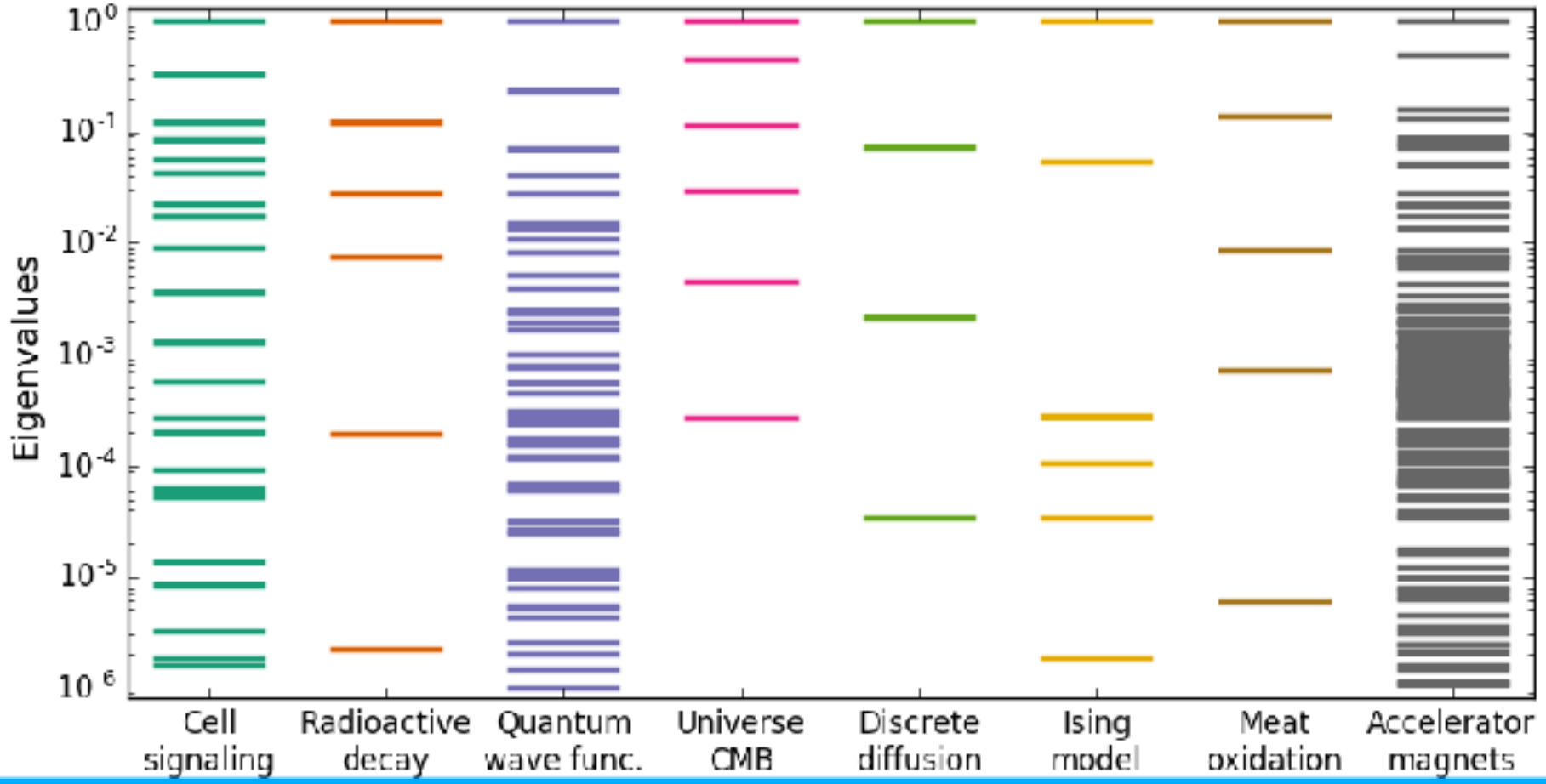
Which Rate Constants are in the Stiffest Eigenvector?



Eigenvector components along the bare parameters reveal which ones are most important for a given eigenvector.



Sloppy Universality Outside Bio

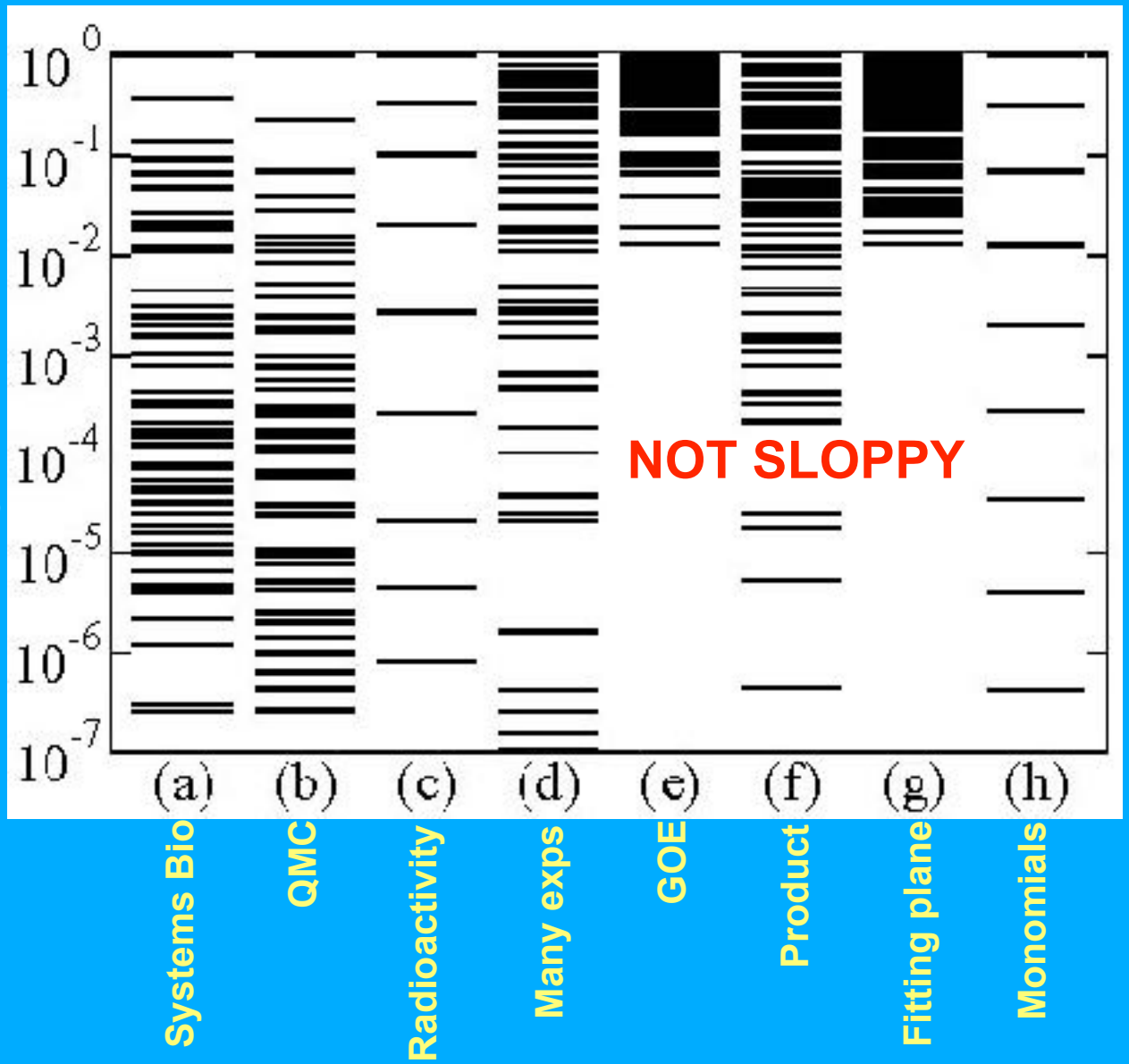


Enormous range of eigenvalues; Roughly equal density in log; Observed in broad range of systems

Some systems aren't sloppy

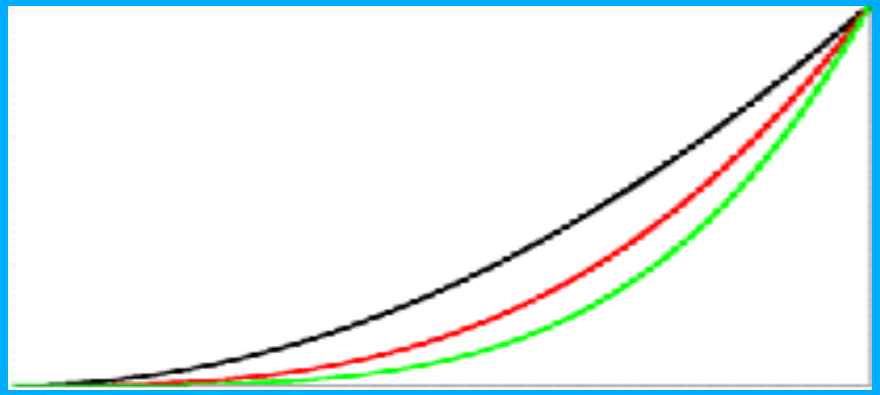
Sloppy Systems

- Quantum Monte Carlo (best molecule excitations)
- Monomial sums
- Not linear regression models
- Not random matrix theory



Where is Sloppiness From?

Fitting Polynomials to Data



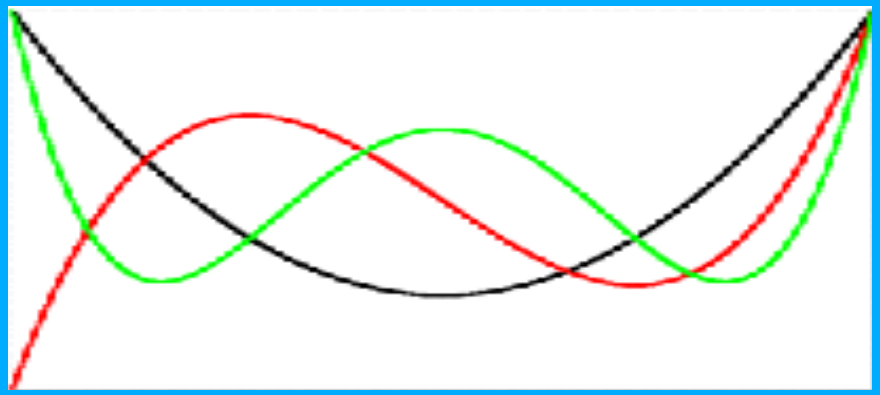
Fitting Monomials to Data

$y = \sum a_n x^n$

Functional Forms Same

Hessian $H_{ij} = 1/(i+j+1)$

Hilbert matrix: famous



Orthogonal Polynomials

$y = \sum b_n L_n(x)$

Functional Forms Distinct

Eigen Parameters

Hessian $H_{ij} = \delta_{ij}$

Sloppiness arises when bare parameters skew in eigenbasis

Small Determinant!

$|H| = \prod \lambda_n$

Exploring Parameter Space

Rugged? More like Grand Canyon (Waterfall)

Glasses: Rugged Landscape
- Metastable Local Valleys
- Transition State Passes

Optimization Hell: Golf Course

Sloppy Models
- Minima: 5 stiff, N-5 sloppy
- Search: Flat planes with cliffs

Variational Wavefunctions for Quantum Monte Carlo

Cyrus Umrigar, Josh Waterfall

The most accurate method for solving Schrödinger's equation for molecules rests on a variational wavefunction (followed by diffusion Monte Carlo):

$$\Psi(R) = \sum_m a_m \exp(-\sum_{i<j} u(r_{ij})) \begin{vmatrix} \phi_1(r_1) & \phi_1(r_2) & \dots & \phi_1(r_N) \\ \phi_2(r_1) & \phi_2(r_2) & \dots & \phi_2(r_N) \\ \dots & \dots & \dots & \dots \\ \phi_N(r_1) & \phi_N(r_2) & \dots & \phi_N(r_N) \end{vmatrix}_{m\uparrow} \begin{vmatrix} \phi_1(r_1) & \phi_1(r_2) & \dots & \phi_1(r_N) \\ \phi_2(r_1) & \phi_2(r_2) & \dots & \phi_2(r_N) \\ \dots & \dots & \dots & \dots \\ \phi_N(r_1) & \phi_N(r_2) & \dots & \phi_N(r_N) \end{vmatrix}_{m\downarrow} \dots$$

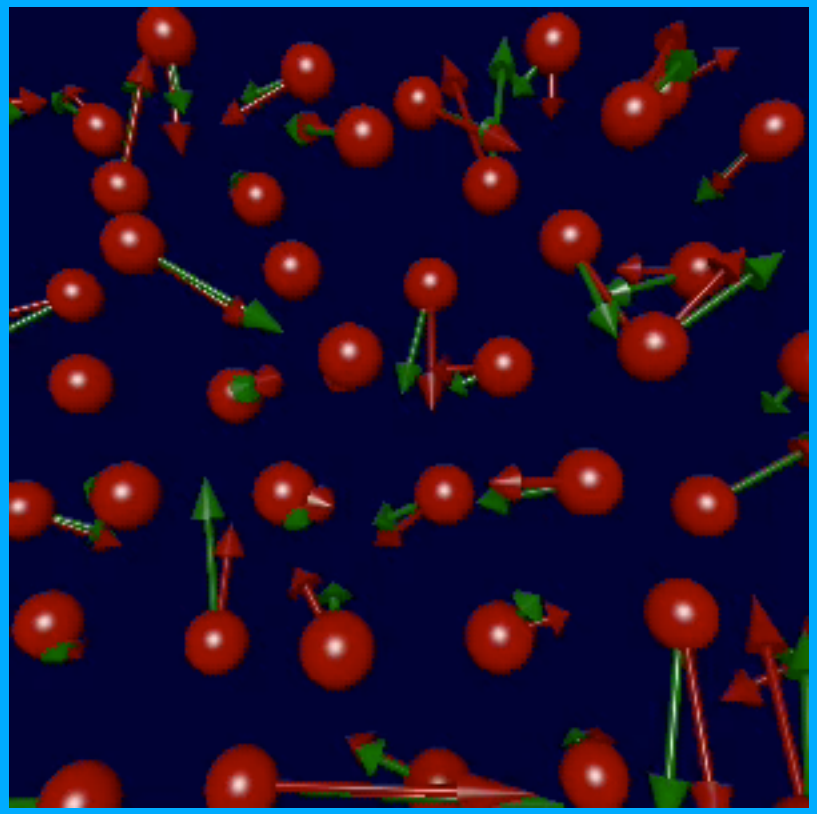
Often parameters θ are varied to minimize energy (find ground states)
Umrigar developed a method which fits the local energies at randomly chosen R_j

$$\text{Data} = \mathscr{H}\Psi_\theta(R_j) / |\Psi_\theta(R_j)| = E(R_j)$$

then varies parameters to minimize fluctuations in local E's away from eigen-energy

Interatomic Potentials

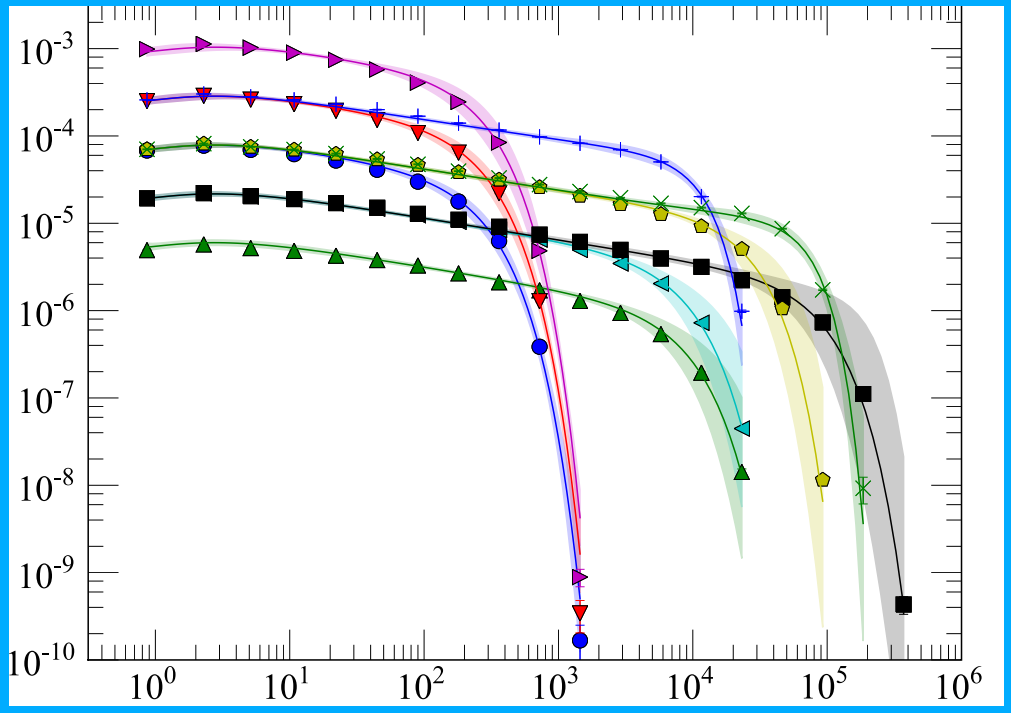
Søren Frederiksen, Karsten Jacobsen, Kevin Brown, JPS



- Need atomic forces and energies
- Guess functional form for potential (17 parameters for MEAM, 5 for Finnis-Sinclair)
- Use electronic ground state energy $U(R_1,R_2,...)$ (Born-Oppenheimer approximation)
- Least-squares fit to DFT calculations of energy, forces for a variety of “important” atomic configurations

Universal Scaling Functions

Yan-Jiun Chen, Zapperi, Durin, Papanikolaou

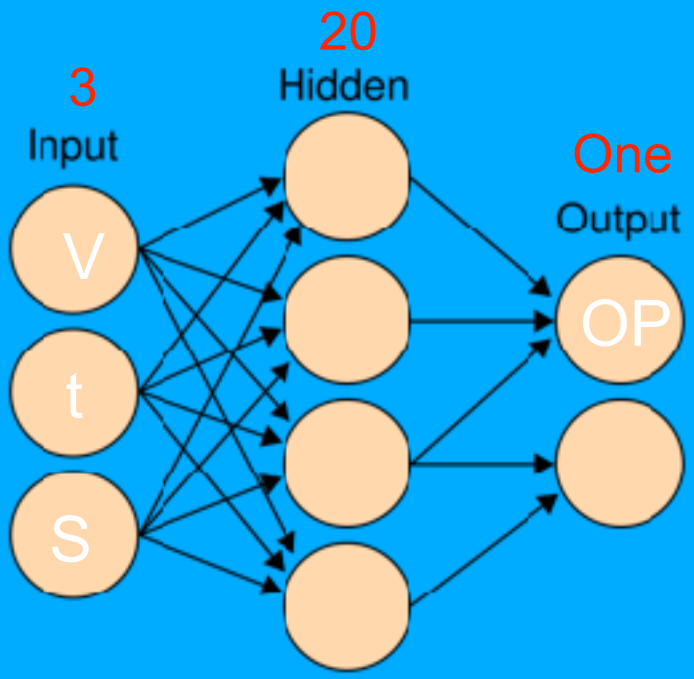


Avalanches through Windows

11 parameter universal scaling form depends on demag length L_K, window width W

Machine Learning

Mark Transtrum



V	t	S	OP
0.20	5.0	75.	25.0000
0.40	5.0	93.	7.2537
0.40	15.0	79.	21.0225
0.66	10.0	91.	10.3957

…

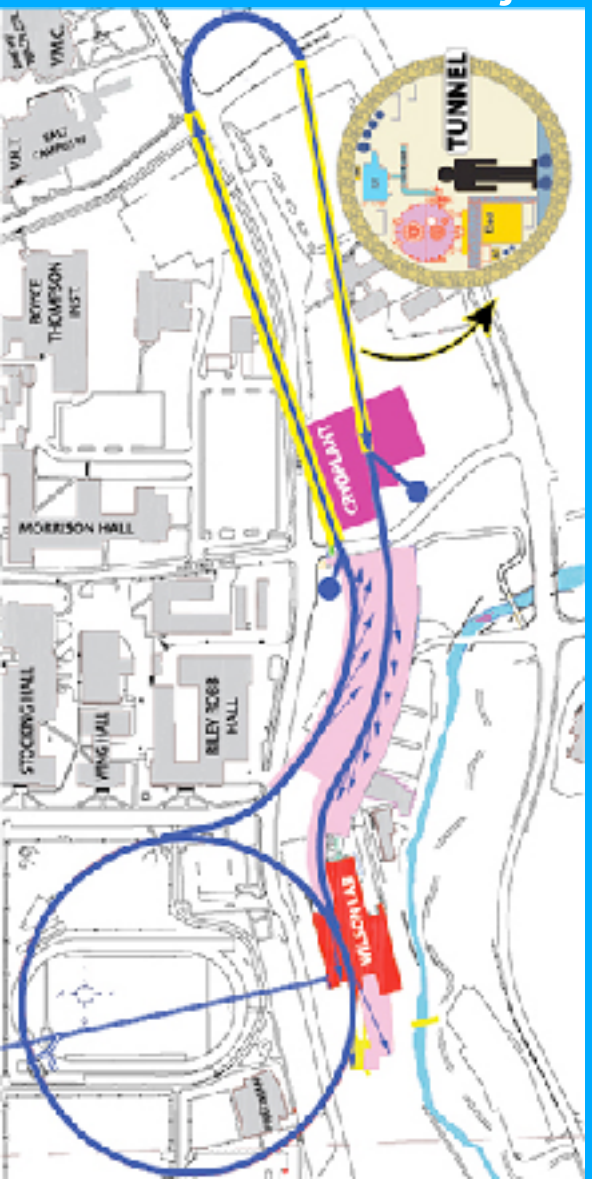
- Neural net “trained” to predict option price, or recognize handwriting
- Neural weights sloppy!
- Output can be viewed as small dimensional representation of Big Data set
- Is the inverse image of output a hyperribbon in Data space?

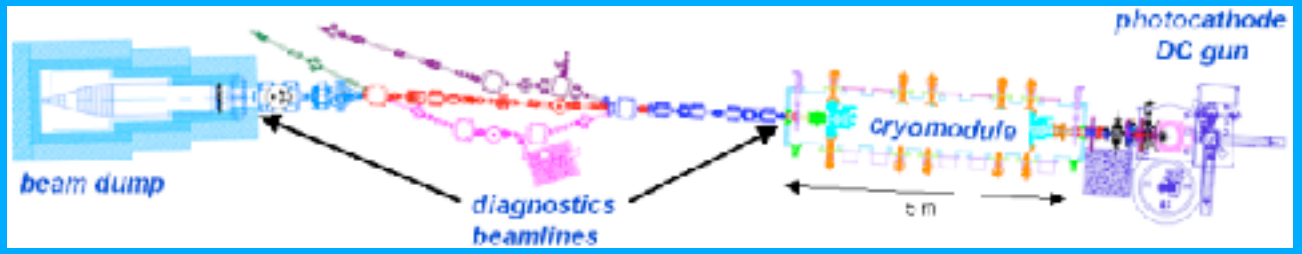
Alex Alemi

Thorsten Joachims CS

Accelerator Optimization

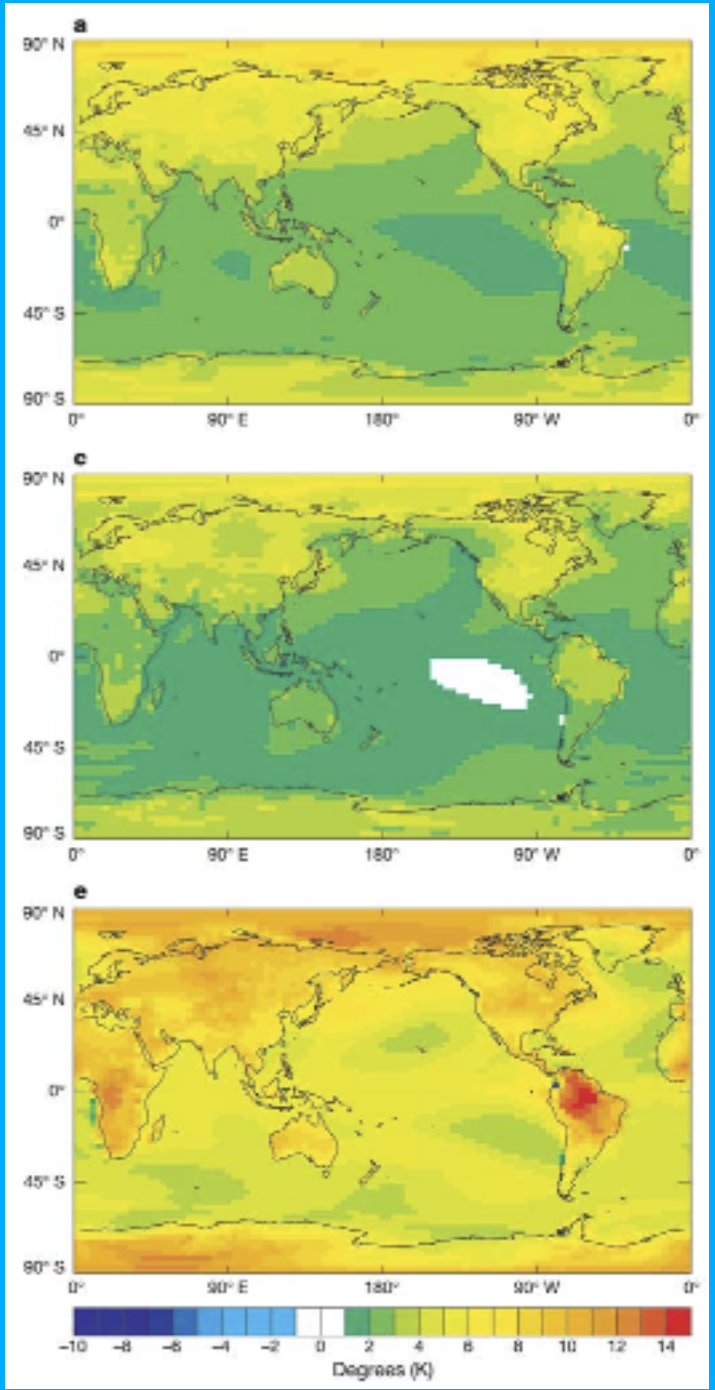
Photoinjector system for Energy Recovery Linac





- Laser hits flexible mirror, which shapes beam on target
- Beam of electrons nonlinear distortion through photoinjector
- Optimize beam profile for emittance
- Use 3D beam shape (point on model manifold), find “stiff knobs” to optimally adjust beam

Colin Clement
Ivan Bazarov (Physics)



Climate Change?

Climate models contain many unknown parameters, fit to data

- General Circulation Model (air, oceans, clouds), exploring doubling of CO_2
- 21 total parameters
- Initial conditions and (only) 6 "cloud dynamics" parameters varied
- Heating typically 3.4K, ranged from < 2K to > 11K

Stainforth et al., *Uncertainty in predictions of the climate response to rising levels of greenhouse gases,* **Nature 433**, 403-406 (2005)

Yan-Jiun Chen
Natalie Mahowald EAS

Macroeconomics?

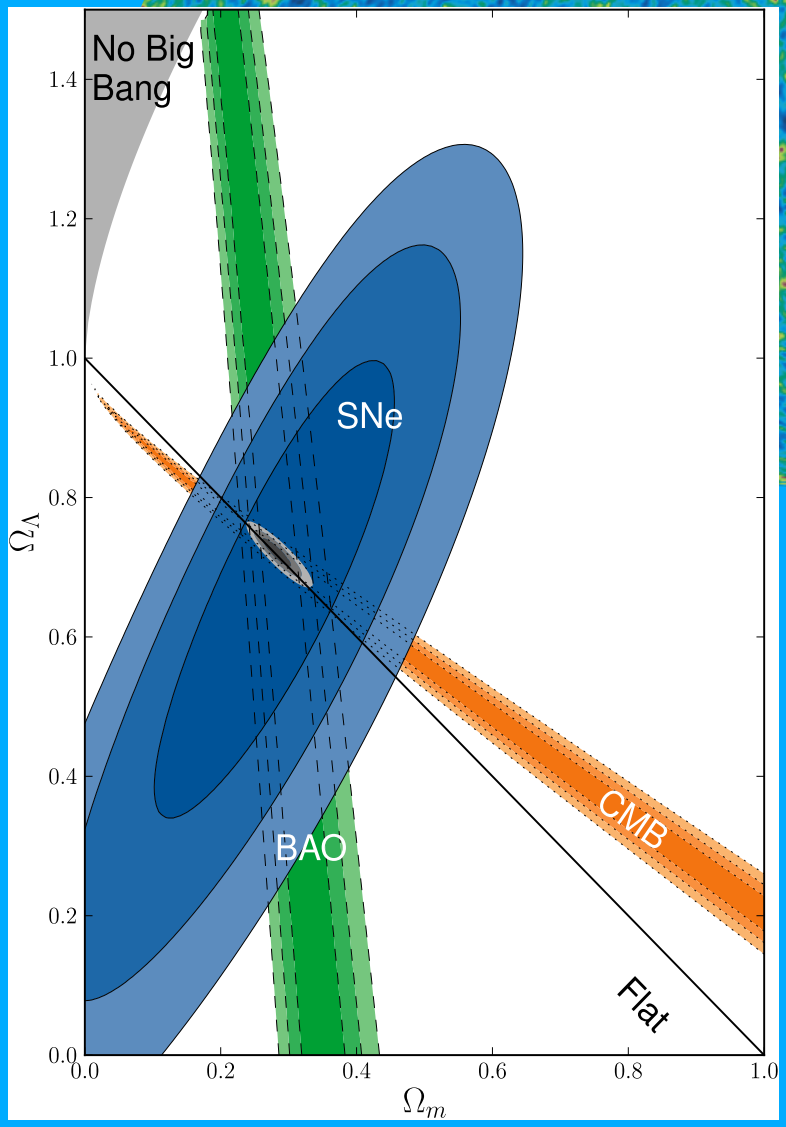
Many ‘structural’ parameters unknown

- Dynamic stochastic general equilibrium models (DSGE), such as Smets and Wouters
- ~25 parameters
- Larger uncertainties and less predictive than much more primitive models with many more parameters?
- Parameter sensitivities: is it sloppy?
- Model reduction: can we simplify it? Can we distill one from more primitive models?
- Are extensions of the model different?

Ricky Chachra
Karel Mertens Econ

The Universe

ΛCDM fit for cosmic microwave background radiation



Universe is flat, mostly unknown dark stuff

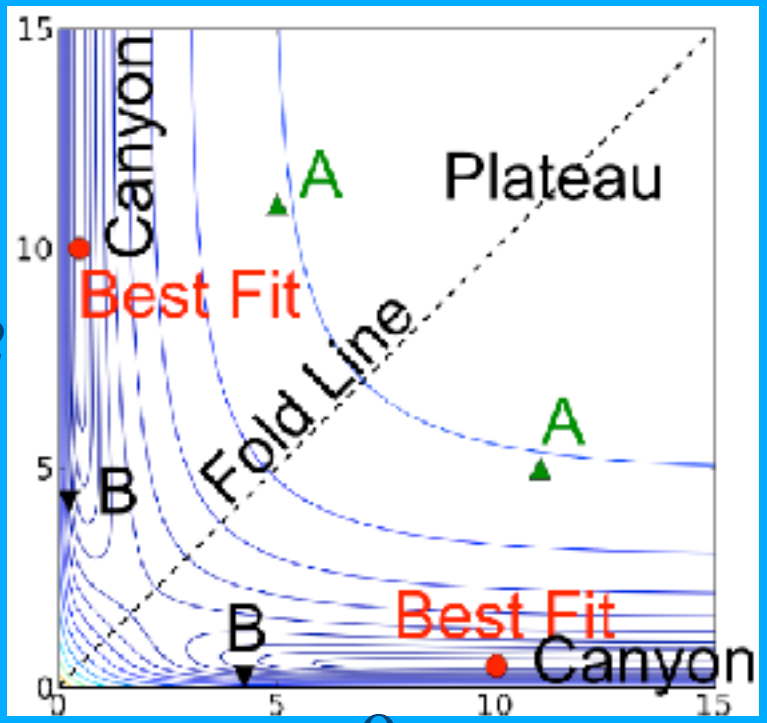
- Six parameter ΛCDM model is sloppy fit to CMB; SNe and BAO determine
- More general models introduce degeneracies

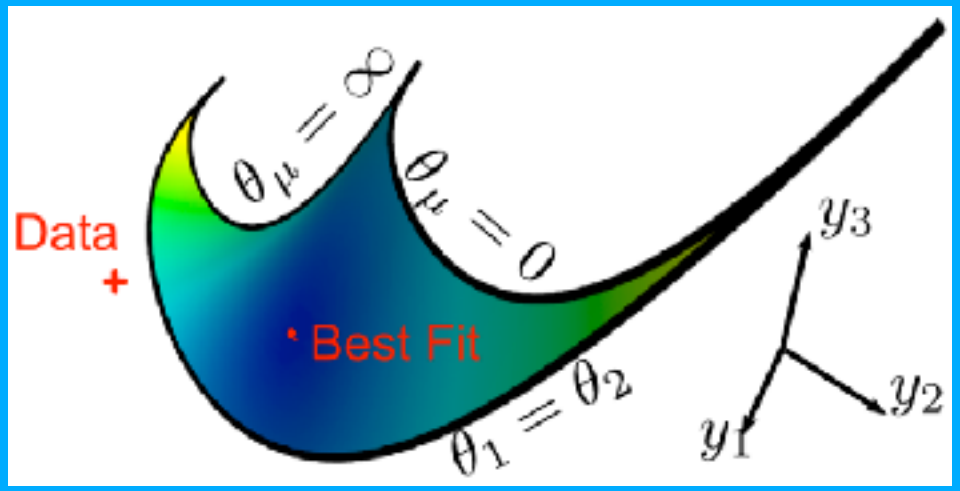
Katherine Quinn
Michael Niemack

The Model Manifold

Two exponentials θ_1, θ_2
fit to three data points y_1, y_2, y_3
$y_n = \exp(-\theta_1 t_n) + \exp(-\theta_2 t_n)$

Parameter space
Stiff and sloppy directions
Canyons, Plateaus





Data space
Manifold of model predictions
Parameters as coordinates
Model boundaries $\theta_n = \pm\infty$, θ_m
cause Plateaus
Metric $g_{\mu\nu}$ from distance to data

Exercise 1.14e

Metric for model manifold

Cost

$$C(\theta) = \chi^2/2 = \sum_i (y_i(\theta) - d_i)^2/2\sigma_i^2$$

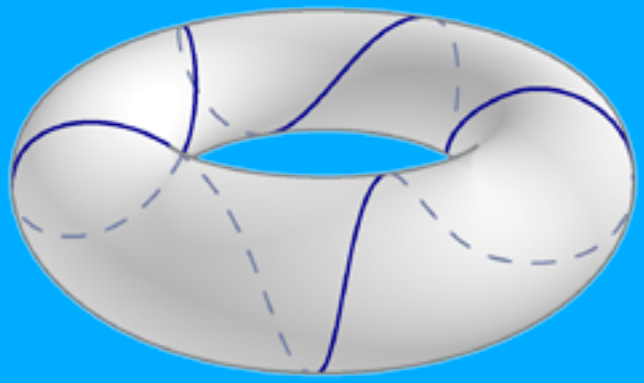
Cost Hessian

$$\begin{aligned} C(\theta) &= C(\theta^{\text{best}} + \boldsymbol{\Delta}) \\ &\approx C(\theta^{\text{best}}) + \Delta_\alpha \frac{\partial C}{\partial \theta_\alpha} + 1/2 \Delta_\alpha \Delta_\beta \frac{\partial^2 C}{\partial \theta_\alpha \partial \theta_\beta} \\ &\approx C^{\text{best}} + 1/2 H_{\alpha\beta} \Delta_\alpha \Delta_\beta \end{aligned}$$

[i] Write $H_{\alpha\beta}$ in terms of first and second derivatives of $y_i(\theta)$.
[ii] If $d_i = y_i(\theta)$, show $H = J^T J$ where $J_{i\alpha} = (1/\sigma_i)\partial y_i/\partial\theta_\alpha$ is the Jacobian.
[iii] Show that the squared distance in data space between $\mathbf{y}(\theta)$ and $\mathbf{y}(\theta + \boldsymbol{\Delta})$ is given by a metric tensor $g = J^T J$.

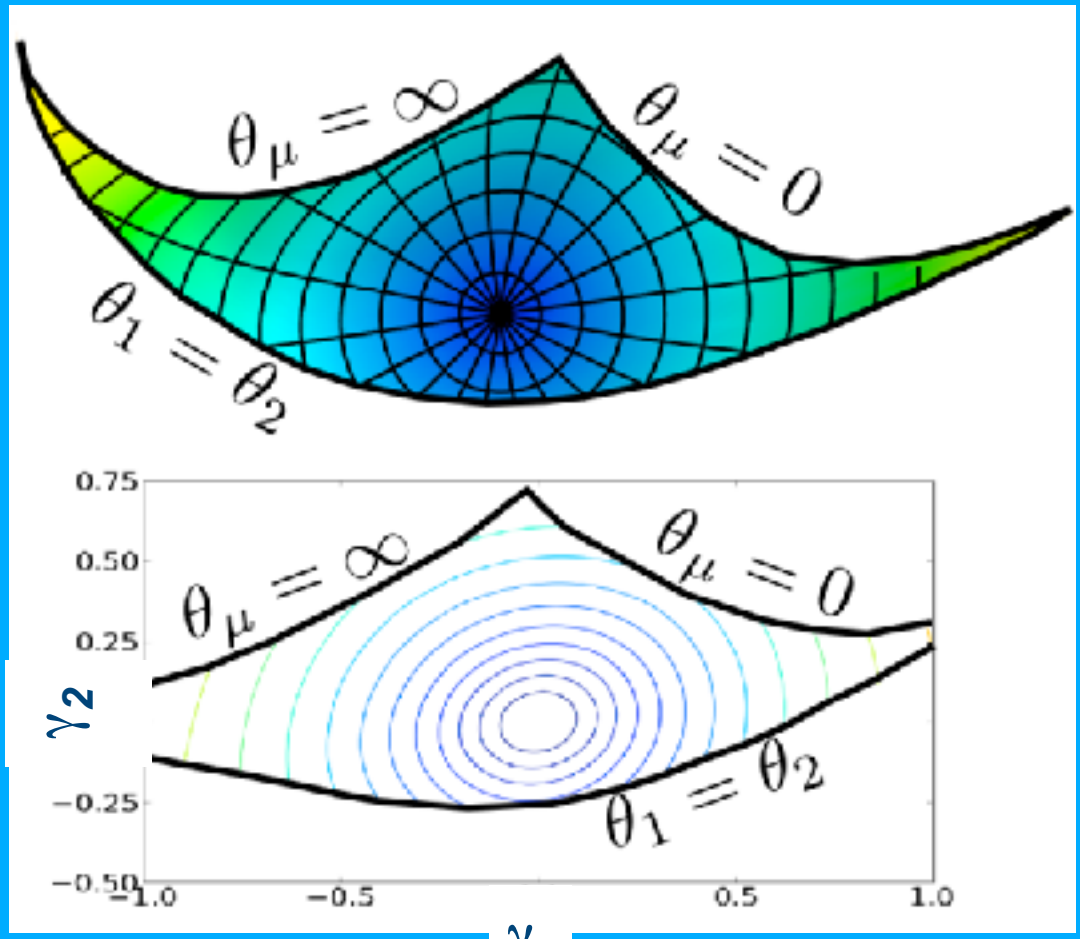
Geodesics

"Straight line" in curved space
Shortest path between points



Easy to find cost minimum using polar *geodesic coordinates*

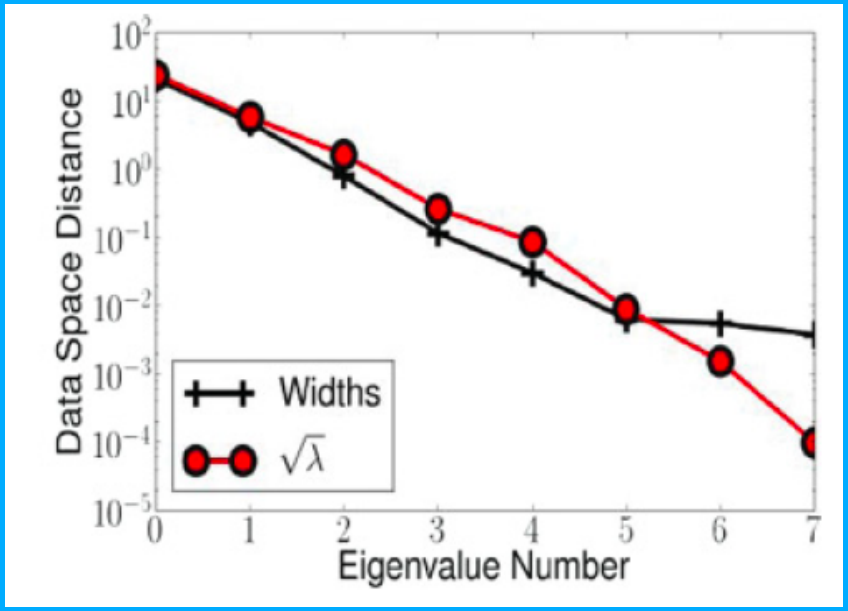
γ_1, γ_2



Cost contours in geodesic coordinates nearly concentric circles!
Use this for algorithms…

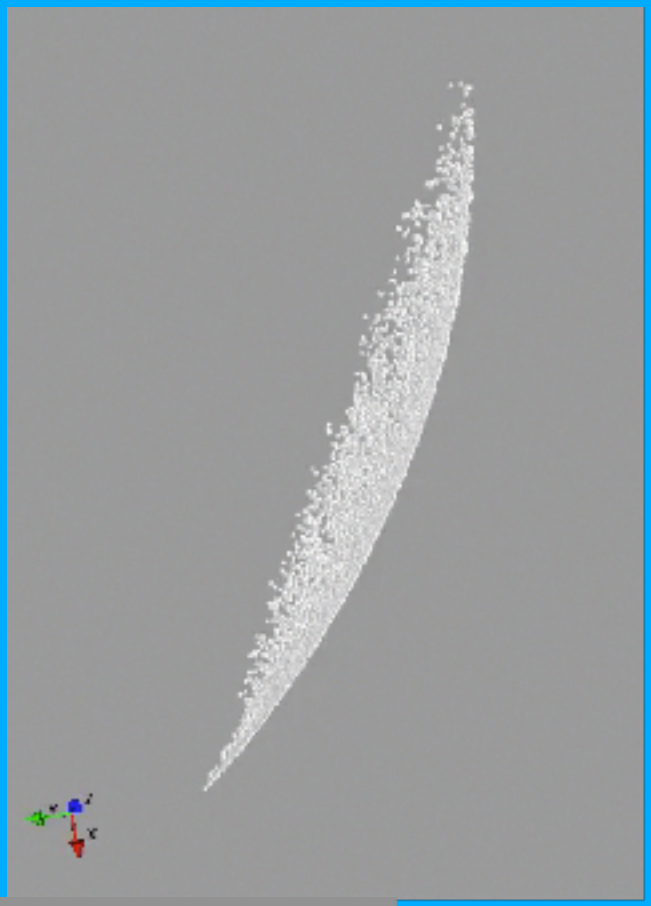
The Model Manifold is a *Hyper-Ribbon*

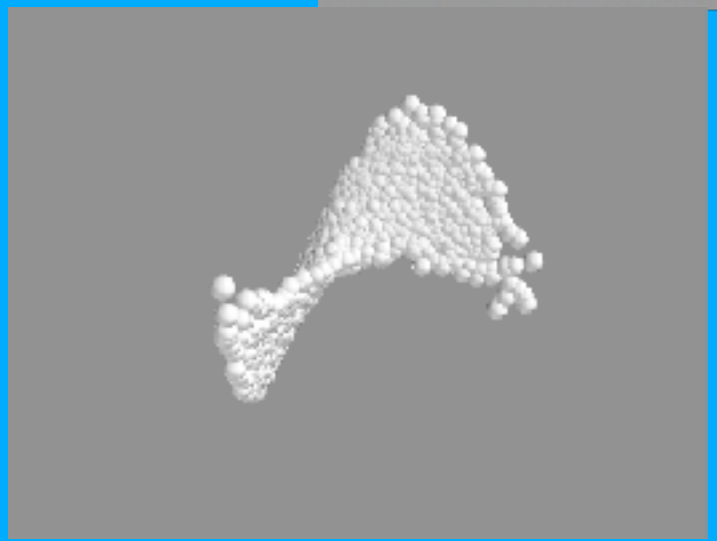
- Hyper-ribbon: object that is longer than wide, wider than thick, thicker than ...
- Thick directions traversed by stiff eigenparameters, thin as sloppy directions varied.



Widths along geodesics track eigenvalues almost perfectly!

Sum of many exponentials, fit to y(0), y(1) data predictions at y(1/4), y(1/2), y(3/4)





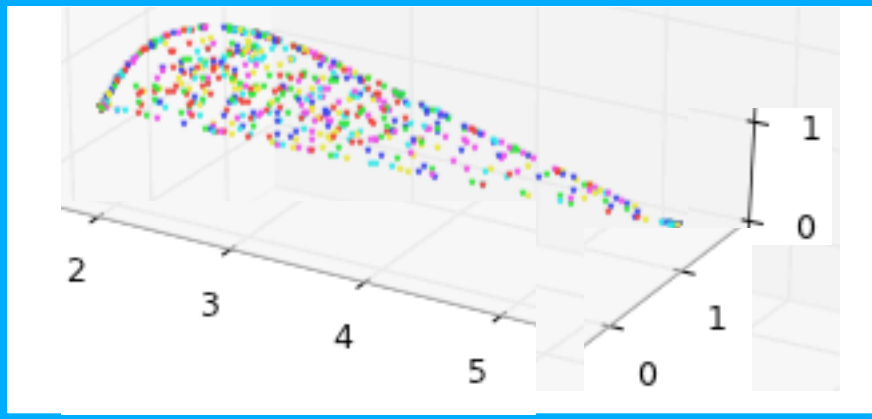
Diffusion equation after three time steps

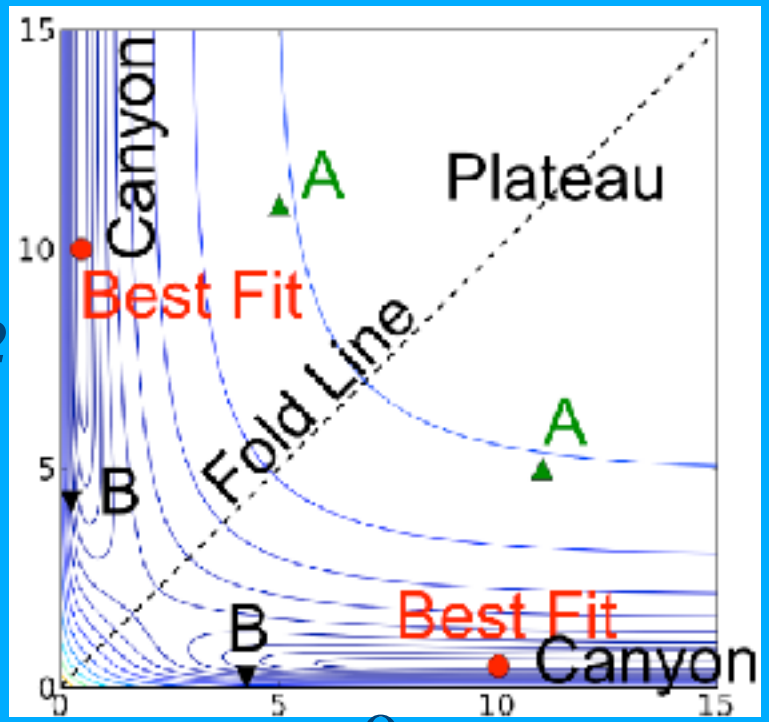
Edges of the model manifold

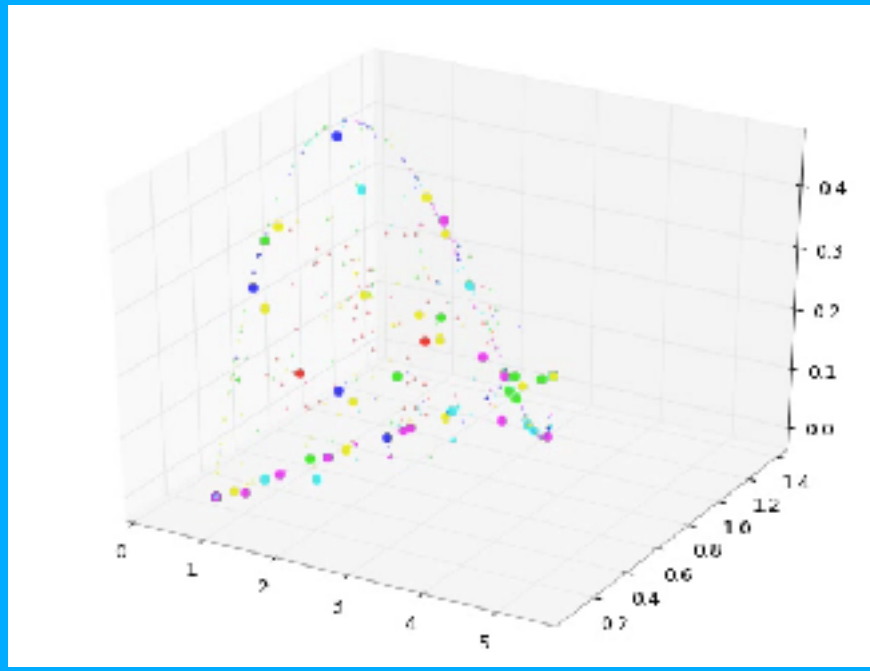
Fitting Exponentials

Top: Flat model manifold; articulated edges = plateau

Bottom: Stretch to uniform aspect ratio (Isabel Kloumann)



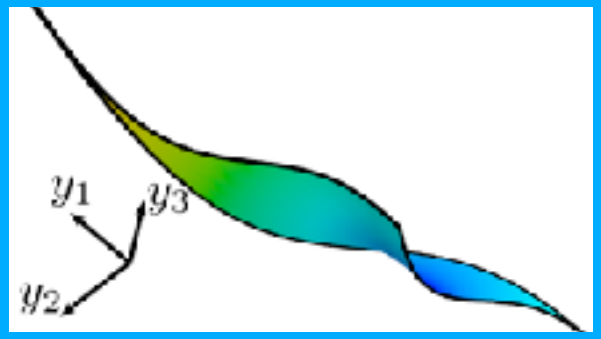




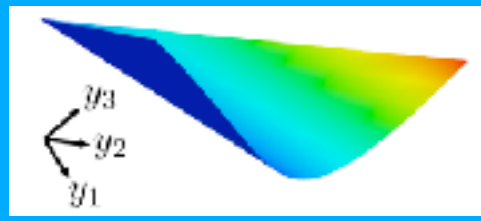
Curvatures

Intrinsic curvature $R^{\mu}{}_{\nu\alpha\beta}$

- determines geodesic shortest paths
- independent of embedding, parameters

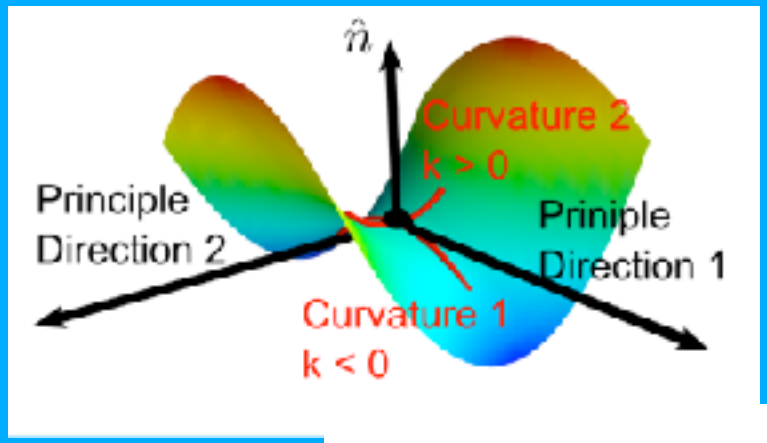


No intrinsic curvature



Extrinsic curvature

- also measures bending in embedding space (i.e., cylinder)
- independent of parameters
- Shape operator, geodesic curvature

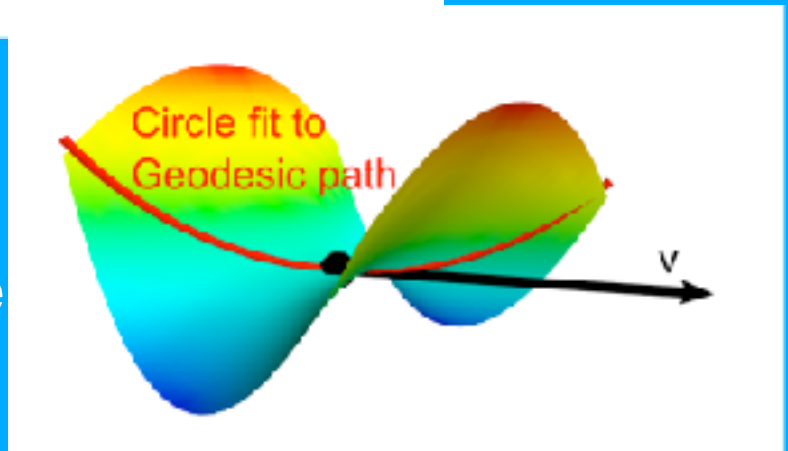


Shape Operator

Parameter effects "curvature"

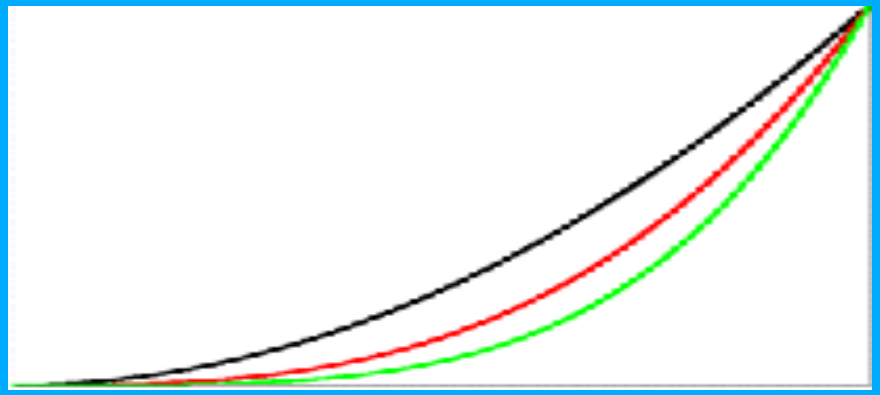
- Usually much the largest
- Defined in analogy to extrinsic curvature (projecting out of surface, rather than into)

Geodesic Curvature



Where is Sloppiness From?

Fitting Polynomials to Data



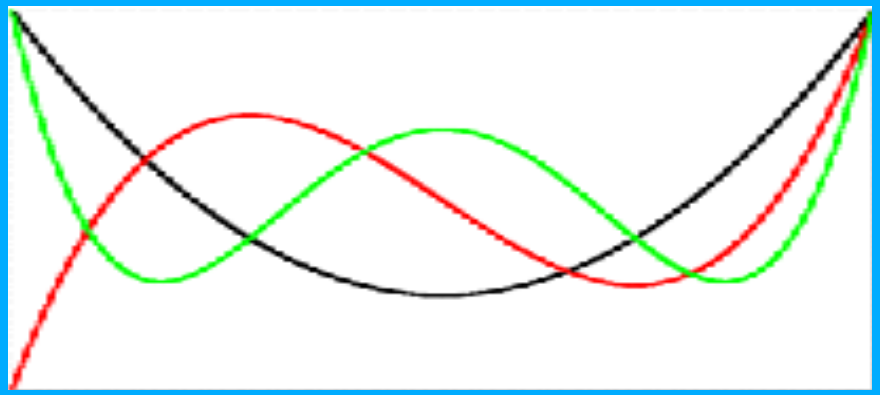
Fitting Monomials to Data

$y = \sum a_n x^n$

Functional Forms Same

Hessian $H_{ij} = 1/(i+j+1)$

Hilbert matrix: famous



Orthogonal Polynomials

$y = \sum b_n L_n(x)$

Functional Forms Distinct

Eigen Parameters

Hessian $H_{ij} = \delta_{ij}$

Sloppiness arises when bare parameters skew in eigenbasis

Small Determinant!

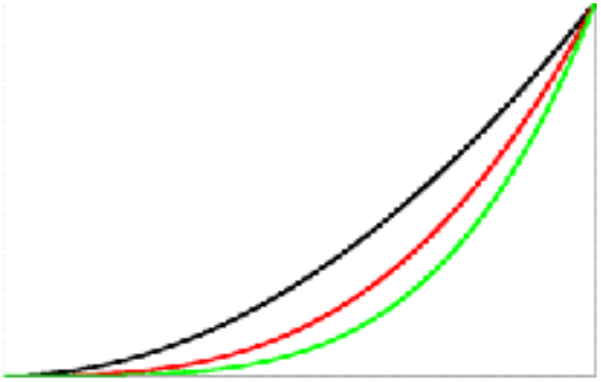
$|H| = \prod \lambda_n$

Exercise Sloppy Monomials

Fitting Polynomials to Data

Sloppy eigenvalues log-equally spaced; $\lambda_m \propto \Delta^{M-1}$. Determinant varies with what power of Δ?

$$C^{\rm poly} = (1/2) \int_0^1 (f(x) - \sum_{m=0}^{M} \theta_m x^m)^2 \, dx$$



For discrete data at $(t_1, \ldots, t_N)$, show $J_{i\alpha} = \partial y(t_i)/\partial \theta_\alpha$ forms the Vandermonde matrix,

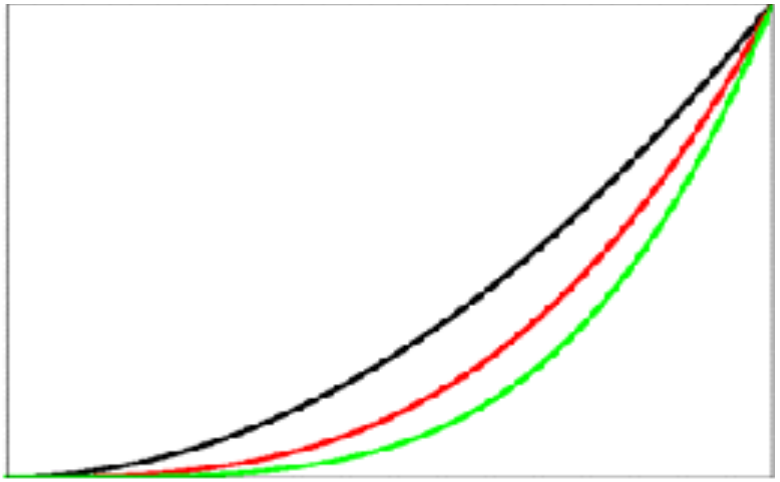
$$J = \begin{pmatrix} 1 & t_1 & t_1^2 & \ldots & t_1^M \\ 1 & t_2 & t_2^2 & \ldots & t_2^M \\ 1 & t_3 & t_3^2 & \ldots & t_3^M \\ \ldots & & & & \end{pmatrix}$$

For M=N (a square matrix), calculate det(Vandermonde). (Hint: It is a polynomial, with roots when $t_i = t_j$.) If $\lambda_m \propto \Delta^{M-1}$, what sets Δ?

The Vandermonde matrix is famous for being ill-conditioned.

Sloppy Monomials (a)

Fitting Polynomials to Data



$$C^{\text{poly}} = (1/2) \int_0^1 (f(x) - \sum_{m=0}^{M} \theta_m x^m)^2 \, dx$$

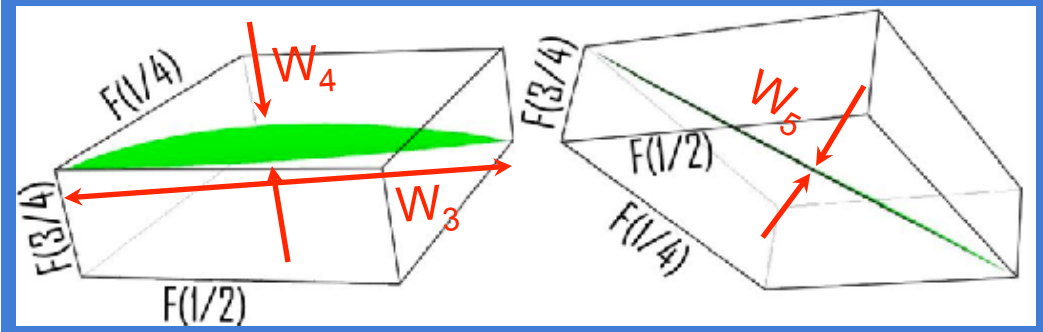
(a) Note that the first derivative of the cost C^{poly} is zero at the best fit. Show that the Hessian second derivative of the cost is

$$H_{mn}^{\text{poly}} = \frac{\partial^2 C^{\text{poly}}}{\partial \theta_m \partial \theta_n} = \frac{1}{m+n+1}. \tag{1}$$

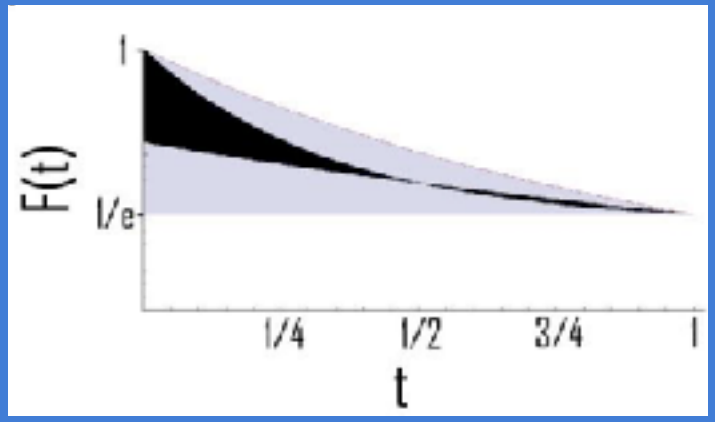
The Hilbert matrix is famous for being ill-conditioned too. The monomial coefficients in a polynomial fit are remarkably poorly determined.

Hierarchy of widths and curvatures

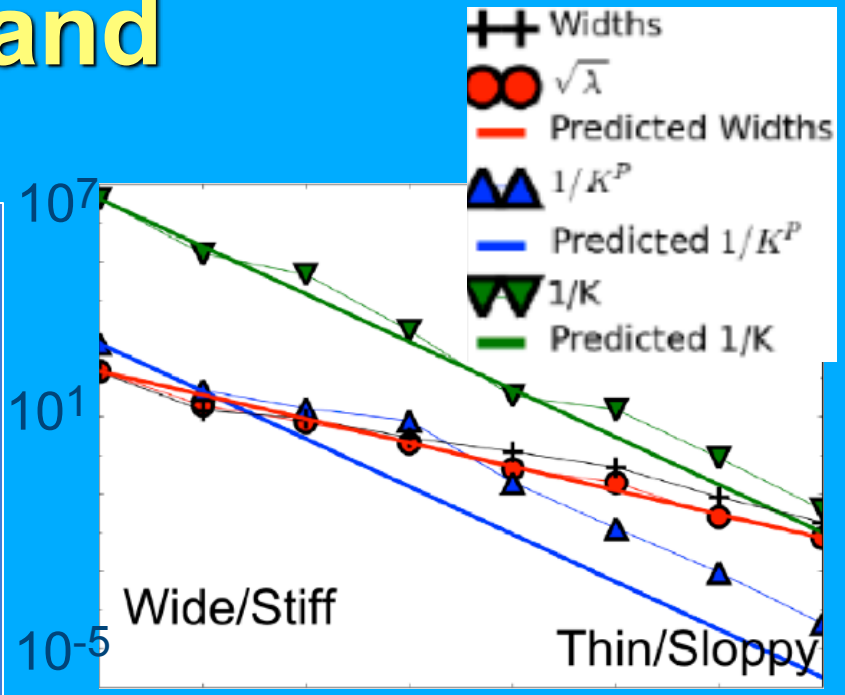
Hierarchy of widths



Cross sections: fixing f at 0, ½, 1



Theorem: interpolation good with many data points
Geometrical convergence



Eigendirection at best fit

Multi-decade span of widths, curvatures, eigenvalues

Widths ~ $\sqrt{\lambda}$ sloppy eigs

Parameter curvature
$K^P = 10^3 \times K$
>> extrinsic curvature

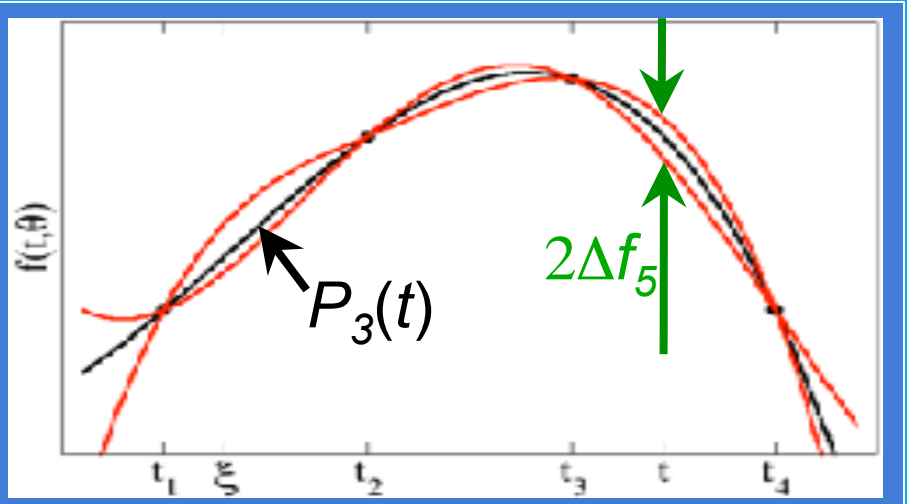
Why is it so thin and flat?

Follows from model analyticity in control parameters

- Model $f(t,\theta)$ analytic:
  $f^{(n)}(t)/n! \le R^{-n}$
- Polynomial fit $P_{m-1}(t)$ to $f(t_1),\ldots,f(t_m)$
- Interpolation convergence theorem

$$\Delta f_{m+1} = f(t)-P_{m-1}(t) < (t-t_1)\,(t-t_2)\ldots(t-t_m)\,f^{(m)}(\xi)/m! \sim (\Delta t / R)^m$$

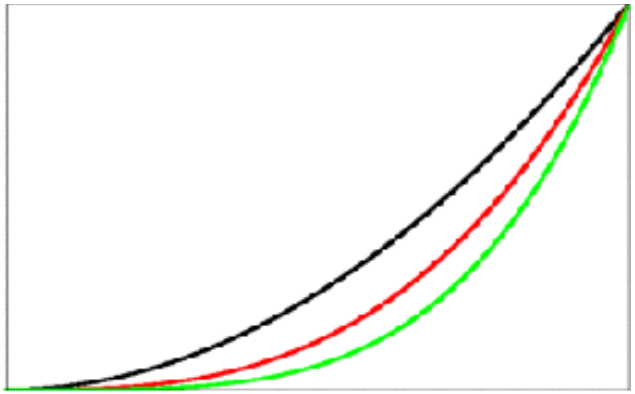
- More than one data per R



Hyper-ribbon: Cross section constraining m points has width $W_{m+1} \sim \Delta f_{m+1} \sim (\Delta t / R)^m$

Sloppy Monomials

Fitting Polynomials to Data



$$C^{\rm poly} = (1/2) \int_0^1 (f(x) - \sum_{m=0}^{M} \theta_m x^m)^2 \, dx$$

The interpolation theorem says that a fit to m data points by a theory with bounded derivative $|f^{(m)}(\xi)| < Am!$ can differ from the $m-1$ degree polynomial fit $P_{m-1}(t)$ by at most

$$\Delta f_{m+1} = |f(t) - P_{m-1}(t)| < A(t-t_1)(t-t_2)\ldots(t-t_m) \qquad (1)$$

[i] Why can m points be fit by an $m-1$ degree polynomial? Why is the difference zero at all the data points? [ii] What does the bound on $|f^{(m)}(\xi)|$ tell us about the maximum value of t^m in a polynomial fit to the function? [iii] Given this bound on the derivative, can any fit have bigger range (thicker model manifold cross-section) than polynomials?

Hyperellipsoid bounds on model manifold

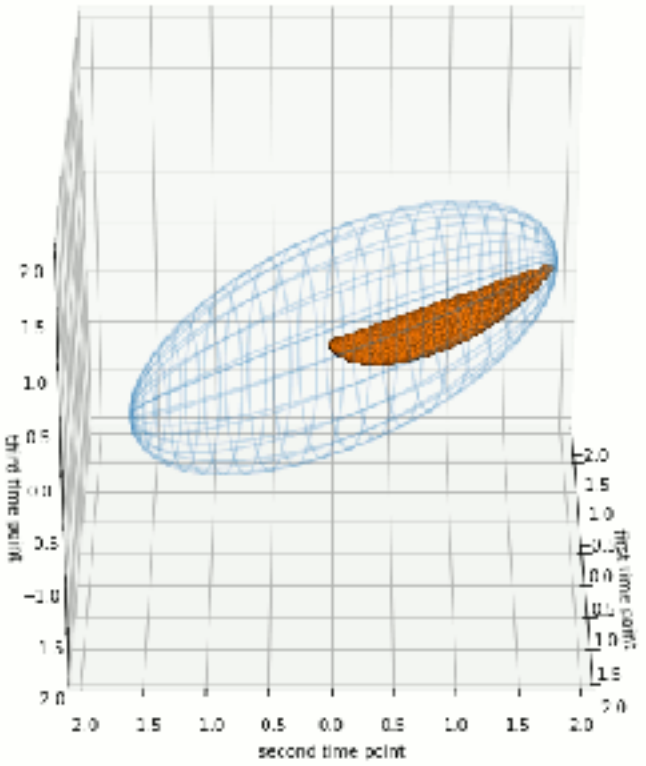
Katherine Quinn

Consider models f(t) with similar ‘radius of variation’

$$\sum (f^{(n)}/n!)^2 < NR^{-n}$$

Any prediction must be contained in a hyper-ellipsoid whose eigenvalues are exponentially separated

Triangle = Model manifold, fitting exponentials
Ellipsoid = Bound for any theory with R=1

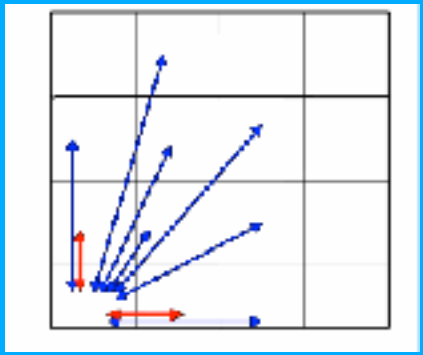


If R=1, parameter sphere of radius N maps via Vandermonde J to hyperribbon ellipsoid, axes = singular values = $\sqrt{\lambda}$

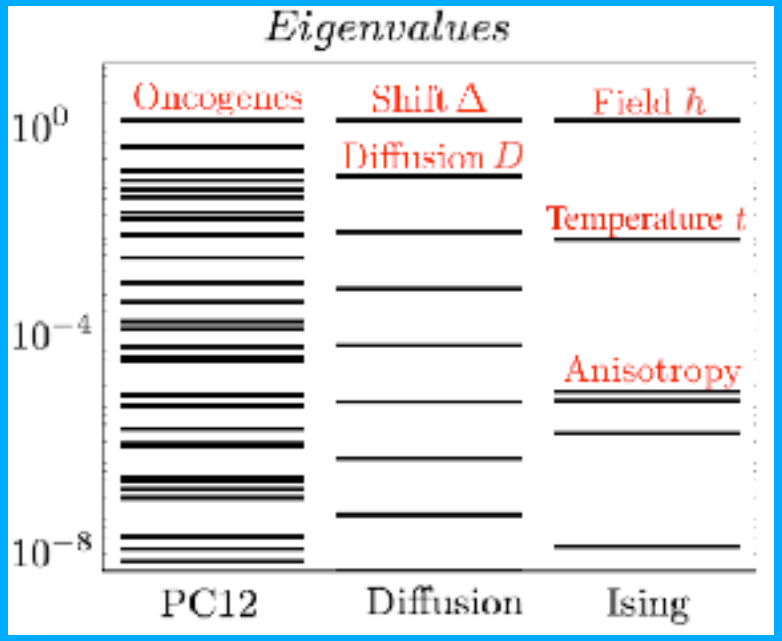
Physics: Sloppiness and Emergence

Ben Machta, Ricky Chachra, Mark Transtrum

Emergence of distilled laws from microscopic complexity

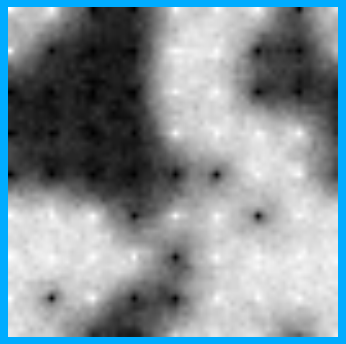


Ising: long bonds
Diffusion: long hops
Irrelevant on macroscale



Both sloppy at long-wavelengths



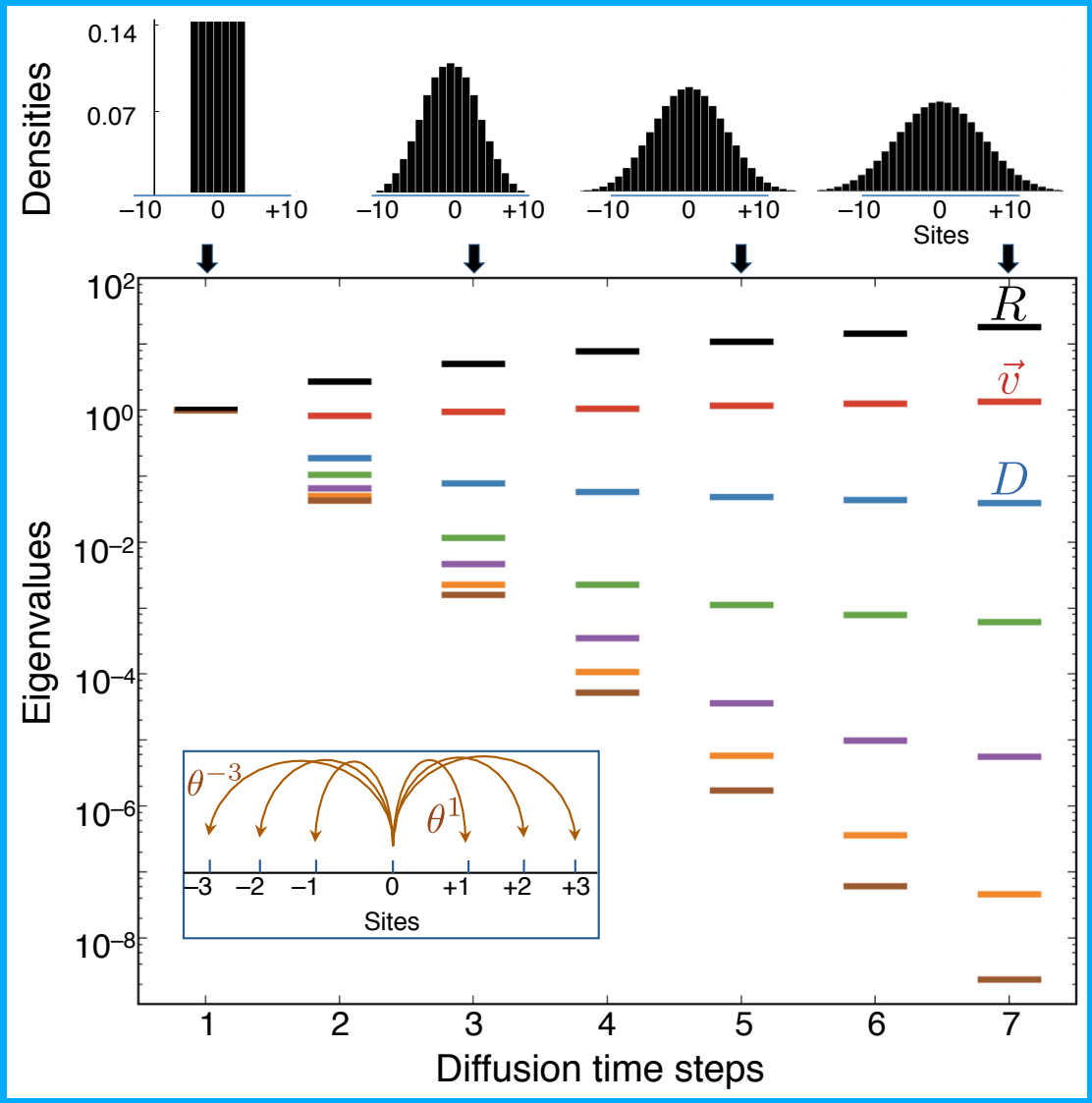



Sloppiness and Continuum Limits

Ben Machta, Ricky Chachra, Mark Transtrum

Microscopic hopping gives continuum diffusion equation

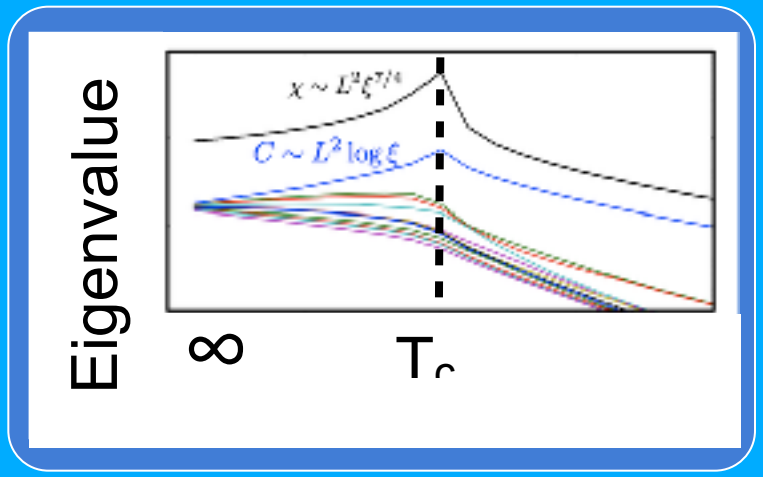
Sloppy after coarse graining in time

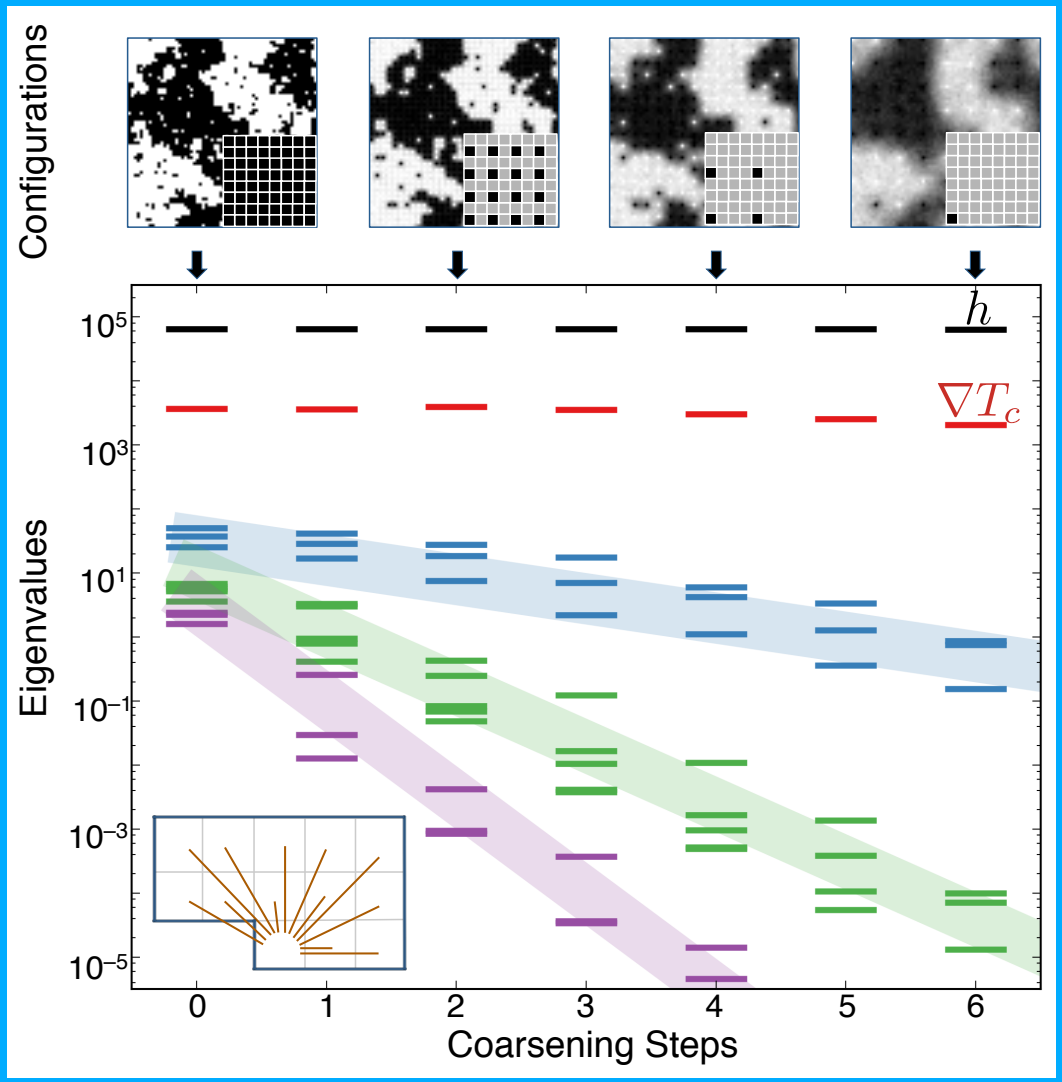


Sloppiness and Critical Points

Ben Machta, Ricky Chachra, Mark Transtrum

Ising model with long-range bonds
Fisher information metric
Sloppy after coarse graining in space

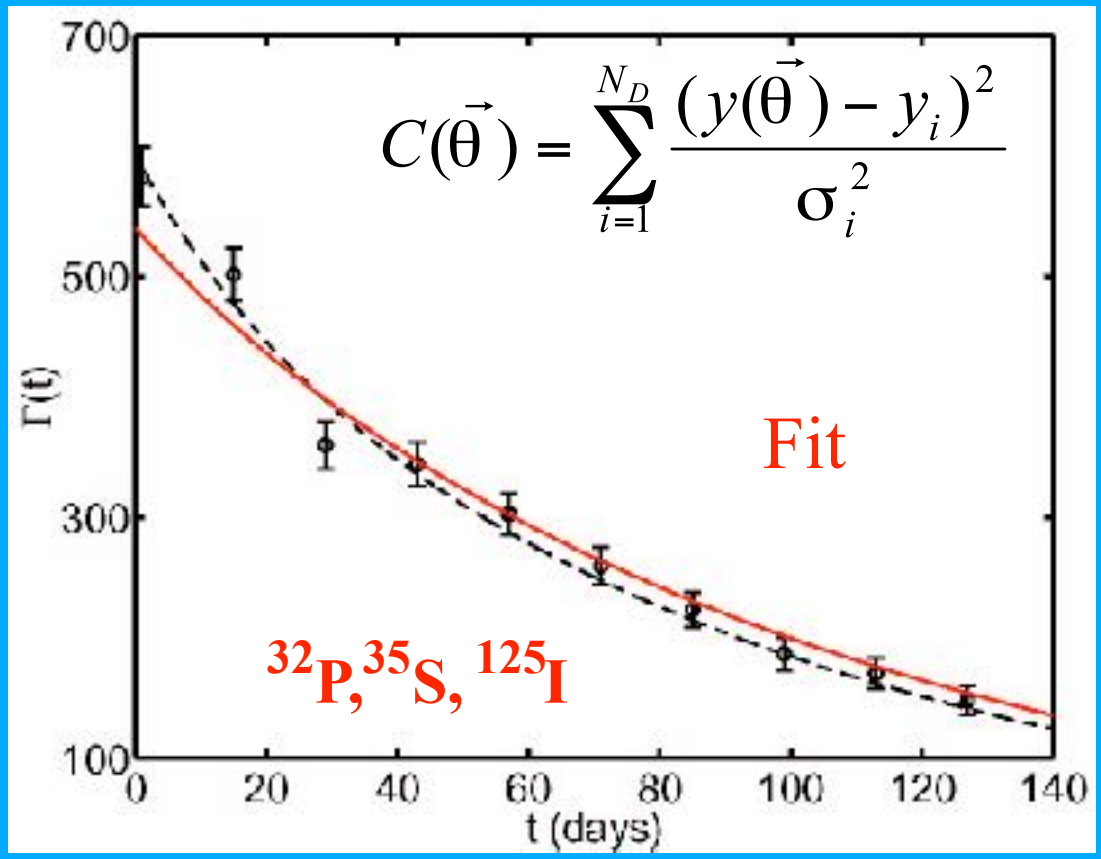




Fitting Decaying Exponentials

Classic ill-posed inverse problem

Given Geiger counter measurements from a radioactive pile, can we recover the identity of the elements and/or predict future radioactivity? Good fits with bad decay rates!

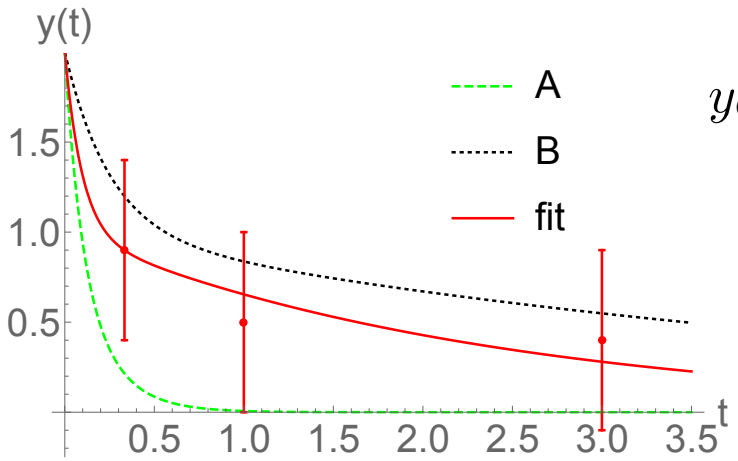


$$y(A,\gamma,t) = A_1 e^{-\gamma_1 t} + A_2 e^{-\gamma_2 t} + A_3 e^{-\gamma_3 t}$$

6 Parameter Fit

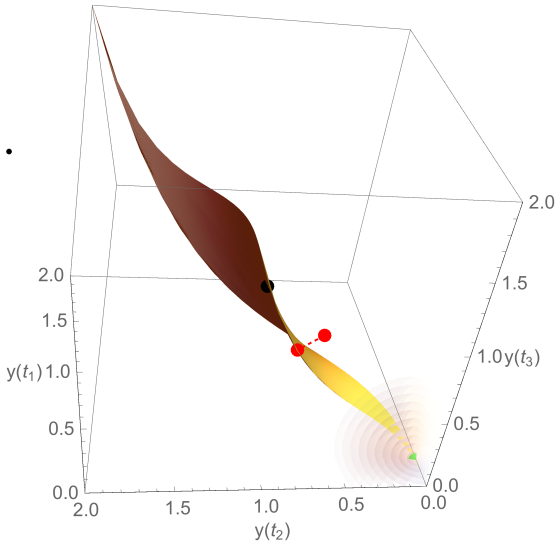
Sloppy Minimization (c)

Decaying exponentials



$$y_{\mathbf{\Theta}}(t) = \sum_{\alpha=0}^{N-1} \theta_\alpha \exp(-\theta_\alpha t).$$

$$J_{i\alpha} = \partial y(t_i)/\partial \theta_\alpha$$



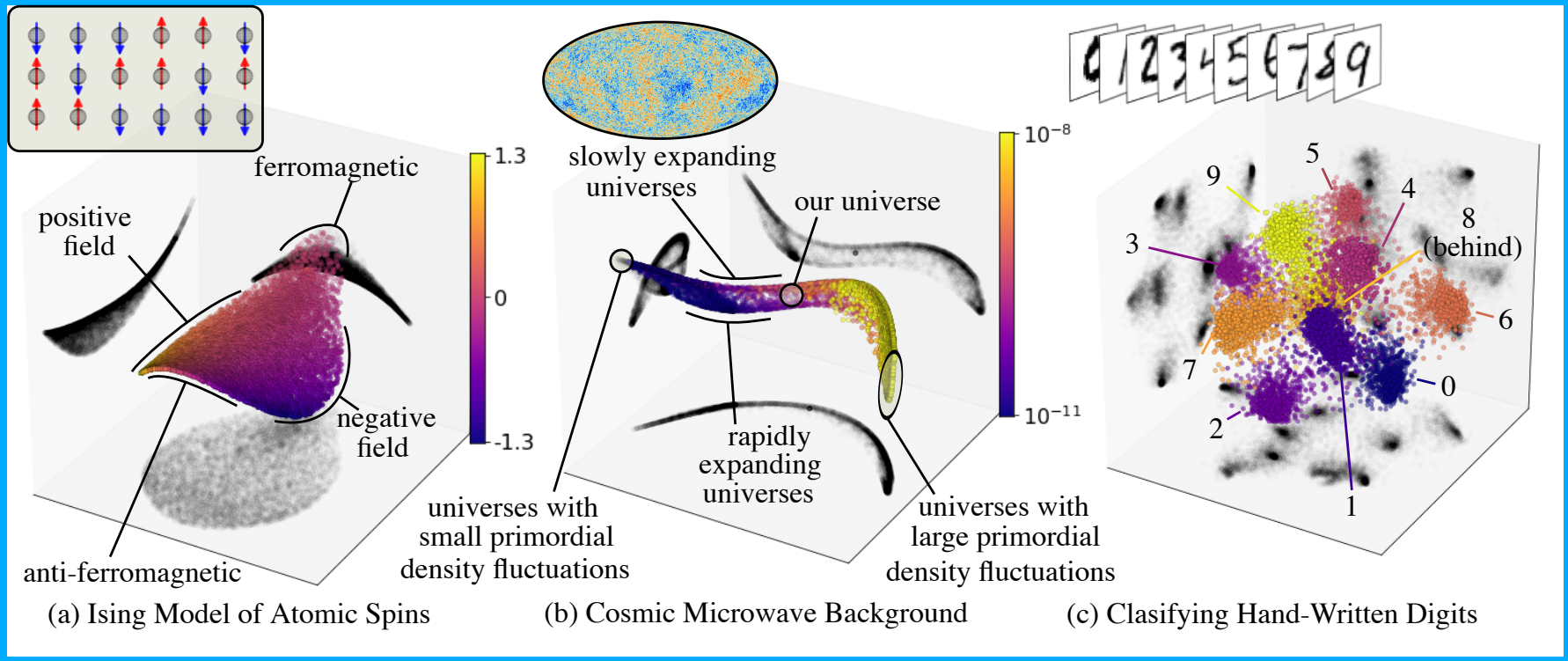
(c) For our exercise, where the data are perfectly fit by $\mathbf{\Theta} = \mathbf{\Theta}_0$, show that the cost Hessian is a continuous integral

$$H_{\alpha\beta} = (J^T J)_{\alpha\beta} = J_{t\alpha} J_{t\beta} = \int_0^\infty J(t,\alpha) J(t,\beta) dt$$

where the Jacobian is now the $\infty \times N$ 'matrix' $J(t,\alpha) = \exp(-\theta_\alpha t)(1 - \theta_\alpha t)$.

Intensive embeddings: Ising, CMB, digits

Katherine Quinn



Probabilistic models

- $g_{\mu\nu}$ = Fisher Information Metric
- Isometric embedding
- Orthogonality catastrophe
- Replica theory — extrapolate to zero data

Exercise 1.15

Metrics in Probability Space

Fisher information matrix

$$g_{\alpha\beta}(\boldsymbol{\Theta}) = -\left\langle \frac{\partial^2 \log P(\mathbf{x})}{\partial\theta_\alpha \partial\theta_\beta} \right\rangle_{\mathbf{x}} = -\int d\mathbf{x}\, P(\mathbf{x}) \frac{\partial^2 \log P(\mathbf{x})}{\partial\theta_\alpha \partial\theta_\beta}$$

Natural metric in space of probability distributions.
Gives distance between nearby probabilities
$(d(\Theta+\Delta)-d(\Theta))^2 = D_\alpha\, g_{\alpha\beta}\, D_\beta$
Agrees with least-squares distance.
'Ease of distinguishing' along different parameter directions

Exercise 1.15b, 1.16a

Fisher Information Metric in Probability Space

Fisher information matrix

$$g_{\alpha\beta}(\mathbf{\Theta}) = -\left\langle \frac{\partial^2 \log P(\mathbf{x})}{\partial\theta_\alpha \partial\theta_\beta} \right\rangle_{\mathbf{x}} = -\int d\mathbf{x}\, P(\mathbf{x}) \frac{\partial^2 \log P(\mathbf{x})}{\partial\theta_\alpha \partial\theta_\beta}$$

Least-squares cost $P_{\mathrm{LS}}(\mathbf{x}|\mathbf{\Theta}) = \exp(-\sum_i (y_i(\theta) - x_i)^2/2\sigma_i^2)/\prod_i \sqrt{2\pi\sigma_i^2}$

(b) Calculate the FIM $g_{\alpha\beta}$ for the least-squares probability P_{LS}. Show that it is $J^T J$, where $J_{i\alpha} = (\partial y_i/\partial\theta_\alpha)/\sigma$

Exercise 1.16c

Hellinger distance

$\sqrt{P(\mathbf{x})}$ is a point on the unit sphere. Hellinger defines a *dot product* between probability distributions P and Q: $P \cdot Q = \sum_{\mathbf{x}} \sqrt{P(\mathbf{x})}\sqrt{Q(\mathbf{x})}$
The Hellinger distance between P and Q is the straight-line distance between them on the sphere:

$$d^2_{\text{Hel}}(P,Q) = (P-Q)^2 = (P \cdot P - 2P \cdot Q + Q \cdot Q)$$
$$= \int d\mathbf{x}(\sqrt{P(\mathbf{x})} - \sqrt{Q(\mathbf{x})})^2 = 2 - 2P \cdot Q.$$

[i] Argue that Hellinger is a valid distance.
[ii] Show that it gives the correct FIM as its metric, up to a constant factor.
[iii] Show that $d_{\mathbf{Hel}}(P,Q) \leq \sqrt{2}$.

Large systems (thermodynamic limit), modest parameter changes
all are near maximum distance

Exercise 1.16d

Hellinger distance

$$d^2_{\rm Hel}(P,Q) = 2 - 2P\cdot Q$$

$$P\cdot Q = \int d\mathbf{x}\sqrt{P(\mathbf{x})}\sqrt{Q(\mathbf{x})}$$

[iv] If $P = \{1/6,\ldots,1/6\}$ are the probabilities of rolling different numbers for a fair die, and $Q = \{1/10,\ldots,1/10,1/2\}$ are the probabilities for a loaded die, what would the Hellinger distance be between P and Q? (A formula is fine.)
[v] Our gambler keeps using the loaded die. Can the casino catch him? Let $P_N(\mathbf{j})$ be the probability that rolling the die N times gives the sequence $\mathbf{j} = \{j_1,\ldots,j_N\}$. Show that

$$P_N\cdot Q_N = (P\cdot Q)^N$$

and hence

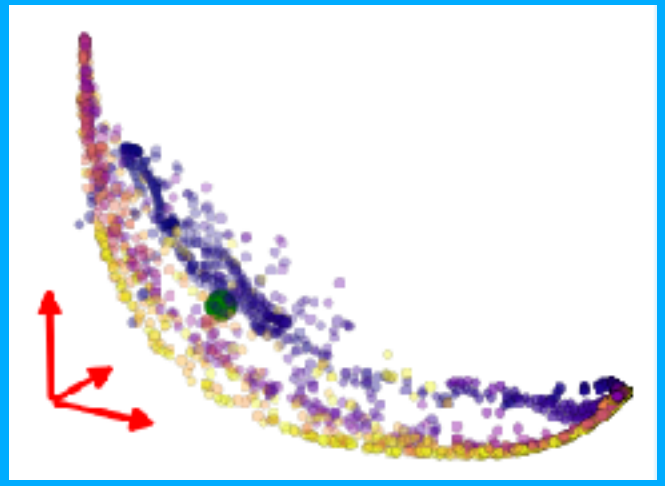
$$d^2_{\rm Hel}(P_N,Q_n) = 2 - 2(P\cdot Q)^N.$$

After $N = 100$ rolls, how close is the Hellinger distance from its maximum?

Large systems have Hellinger distances that saturate rapidly as parameters are changed — hard to use for visualization

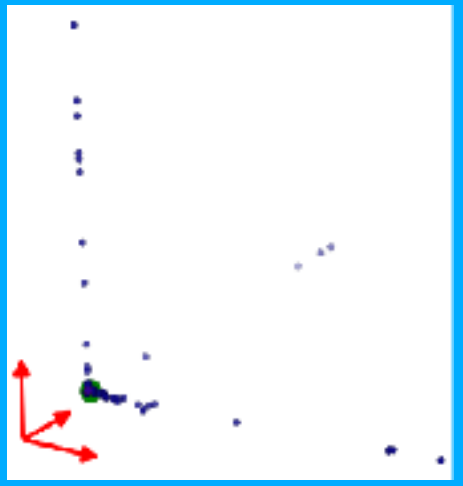
Intensive embeddings: Ising and CMB

Katherine Quinn

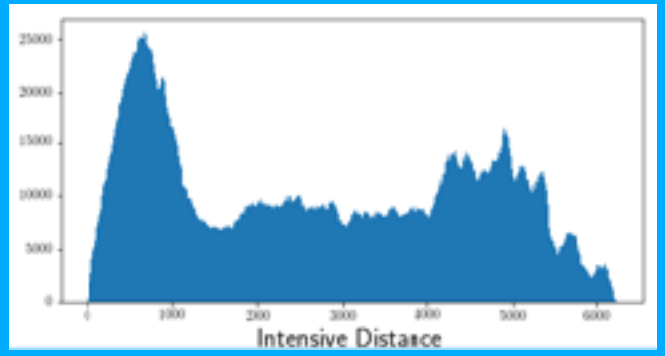


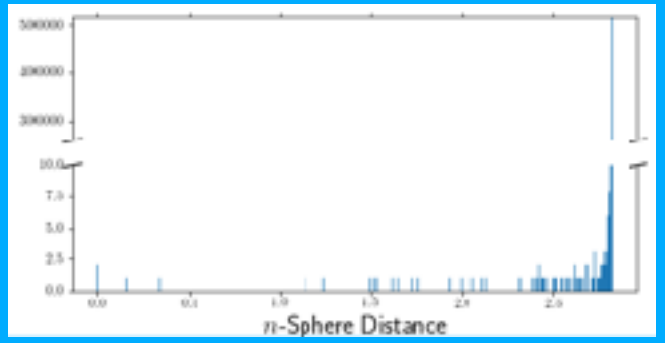
CMB: Intensive embedding

Hellinger distance useless for visualizing CMB spectra. Thousands of Fourier components, millions of data points — all pairs are nearly orthogonal.



CMB: Hellinger embedding





Exercise 1.16e

Bhattacharyya intensive distance (Quinn)

Hellinger distance for N replicas (samples) is $P_N \cdot Q_N = \exp(N \log(P \cdot Q))$; it is extensive in the amount of data. Katherine Quinn takes $N \to 0$ to define our *intensive distance*

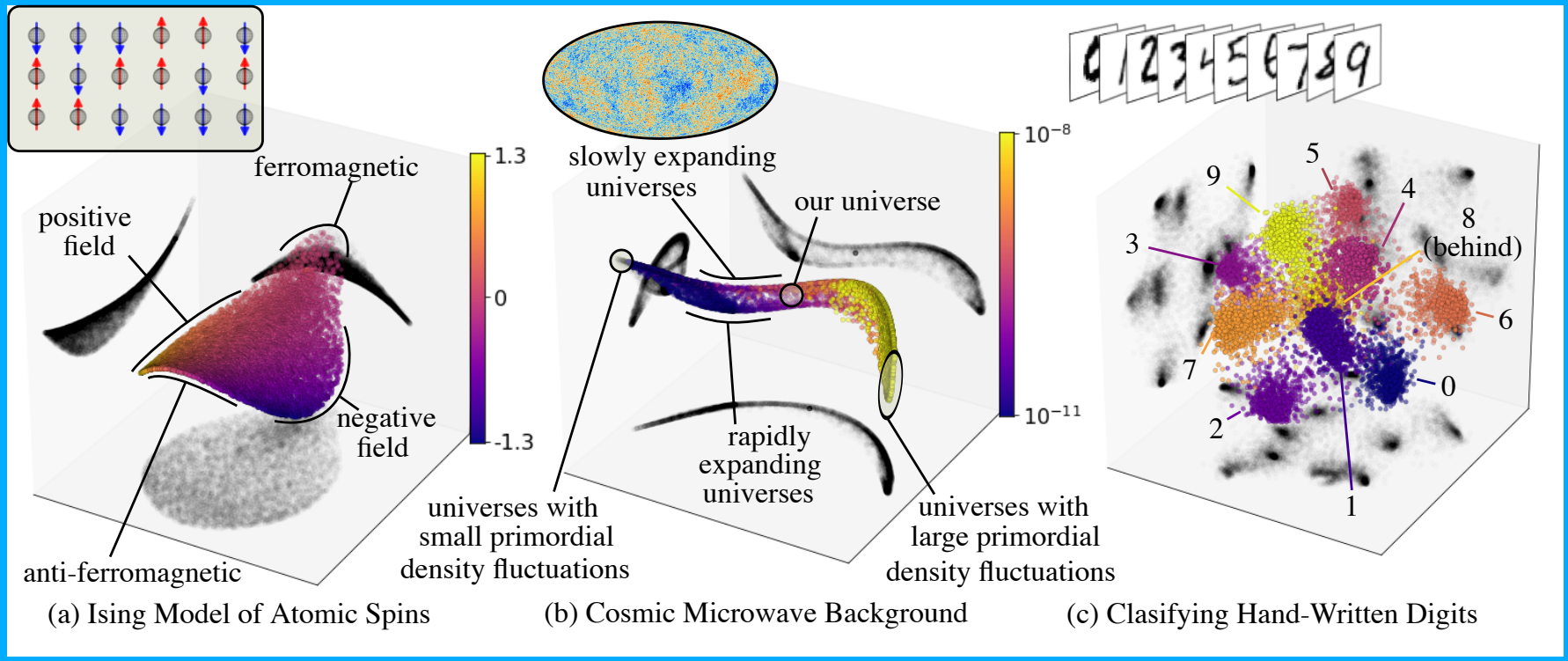
$$d^2_{\text{intensive}} = \lim_{N \to 0} d^2_{\text{Hel}}(P_N, Q_N)/N = -2 \log(P \cdot Q).$$

(f) Derive this equation. *(Hint:* $Z^N \approx \exp(N \log Z) \approx 1 + N \log Z$ *for small* N*.)* Show that the intensive distance does not satisfy the triangle inequality. *(Hint: Find two distributions of loaded dice which do not overlap (so* $P \cdot Q = 0$*), but both overlap with a third.)* Show that it does satisfy the other conditions for a distance. Show, for the nonlinear least-squares model of part (b), that the intensive distance equals the distance in data space between the two predictions.

The intensive distance allows for great visualizations. But it's not a metric, and it embeds the manifold in a Minkowski-like space.

Intensive embeddings: Ising, CMB, digits

Katherine Quinn

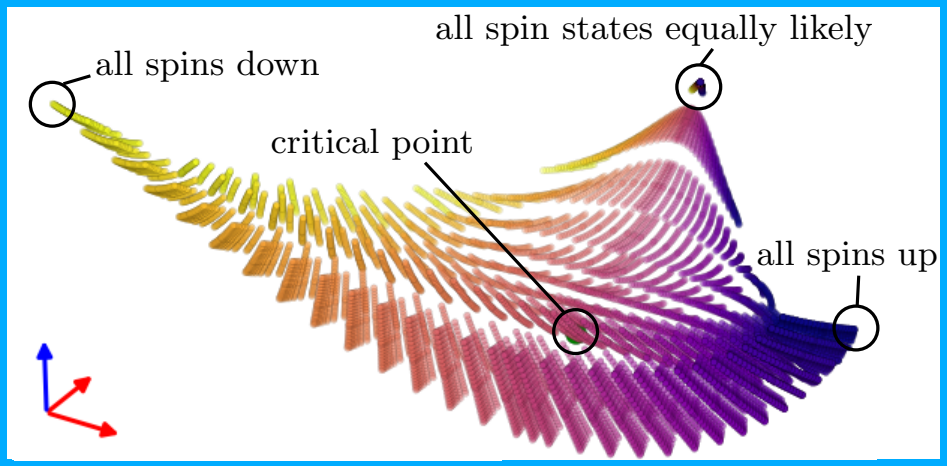


Probabilistic models

- $g_{\mu\nu}$ = Fisher Information Metric
- Isometric embedding
- Replica theory — extrapolate to zero data
- Embedded in Minkowski space

Renormalization group and the model manifold

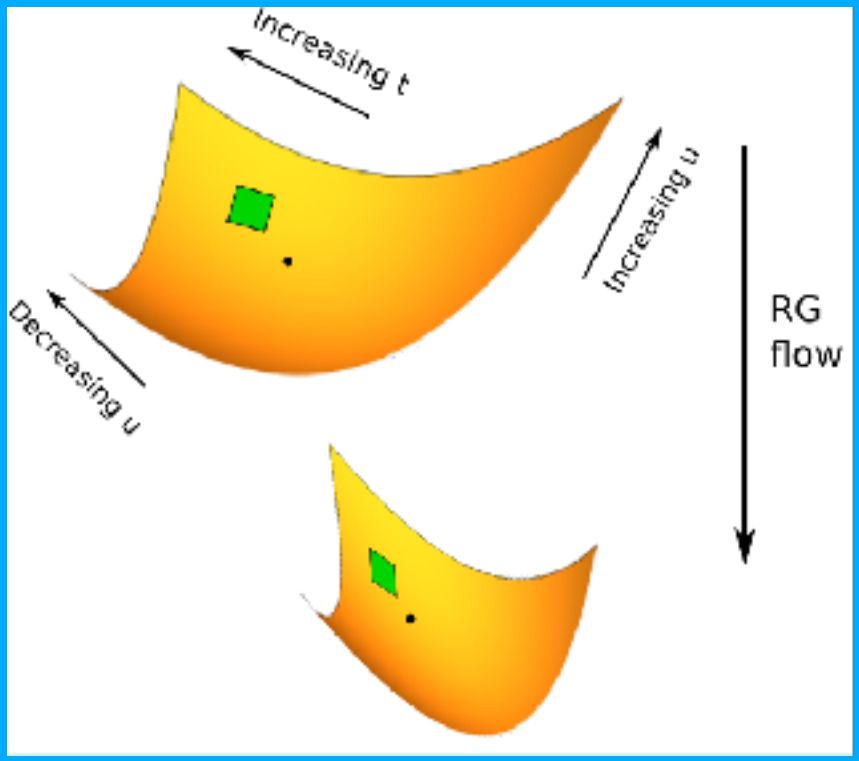
Archishman Raju, Ben Machta



RG on model manifold:
Relevant distances ***unchanged****; irrelevant shrink*

Emergent variables are those which remain equally distinguishable at long length scales

RG in parameter space:
Relevant t, h grow, irrelevant shrink

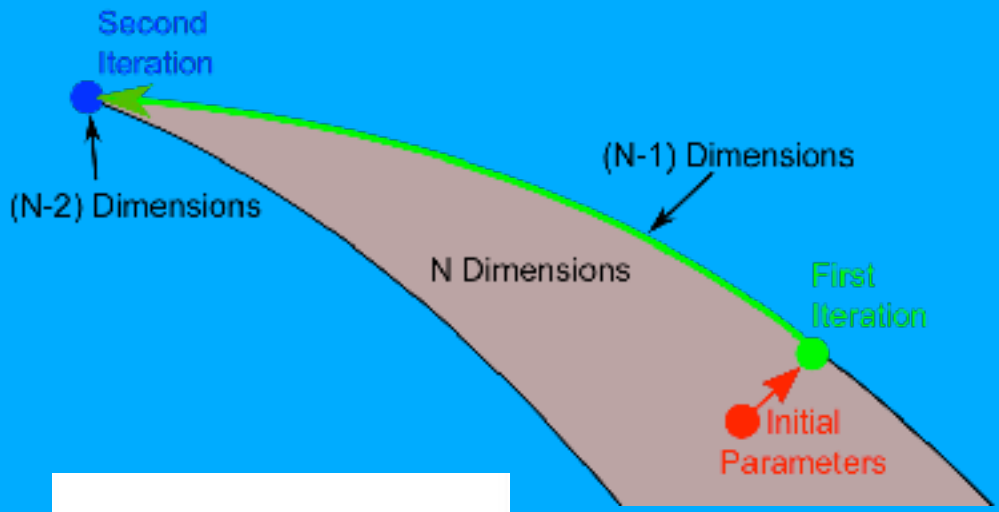


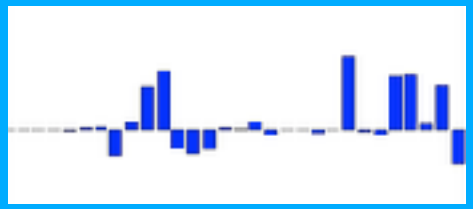
Generation of Reduced Models

Mark Transtrum (not me)

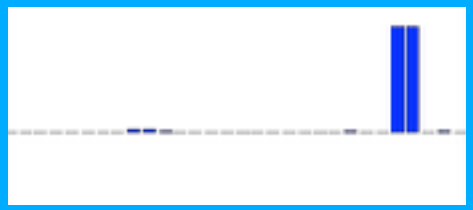
Can we coarse-grain sloppy models? If most parameter directions are useless, why not remove some?
Transtrum has *systematic* method!
(1) Geodesic along sloppiest direction to nearby point on manifold boundary
(2) Eigendirection simplifies at model boundary to chemically reasonable simplified model

Coarse-graining = boundaries of model manifold.





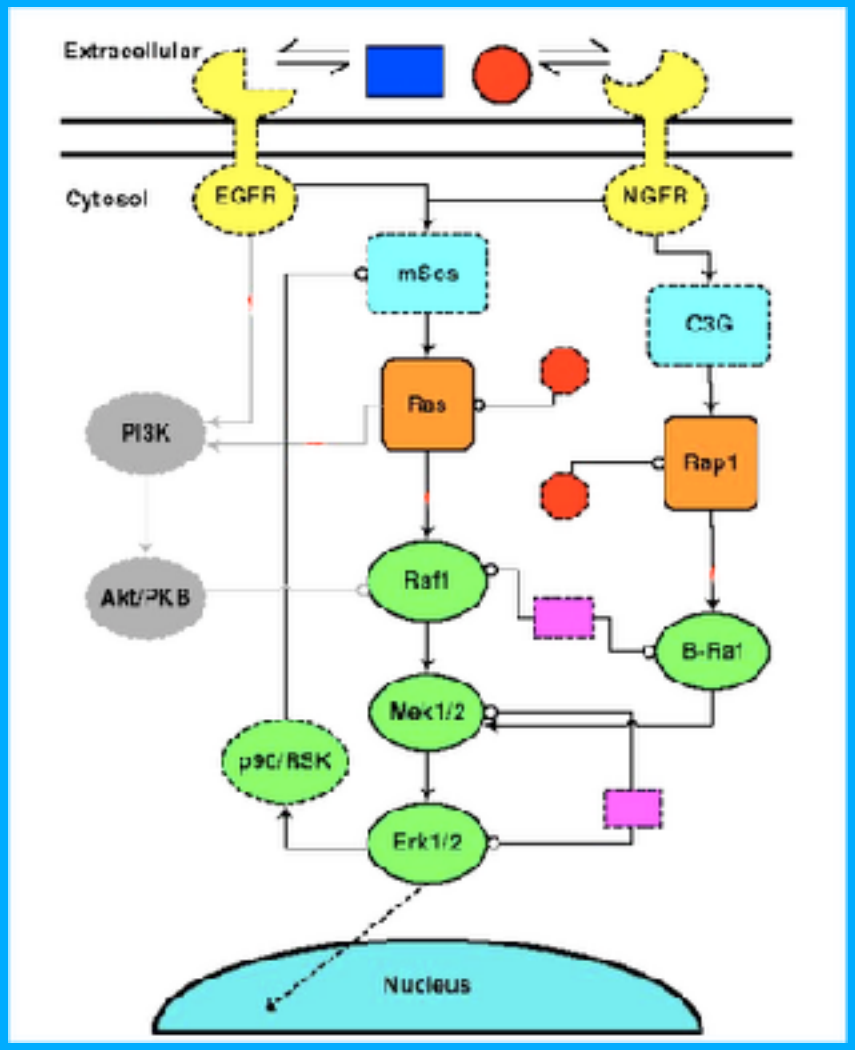
Sloppiest Eigendirection



Simplified at Boundary
(Unsaturable reaction)

Generation of Reduced Models

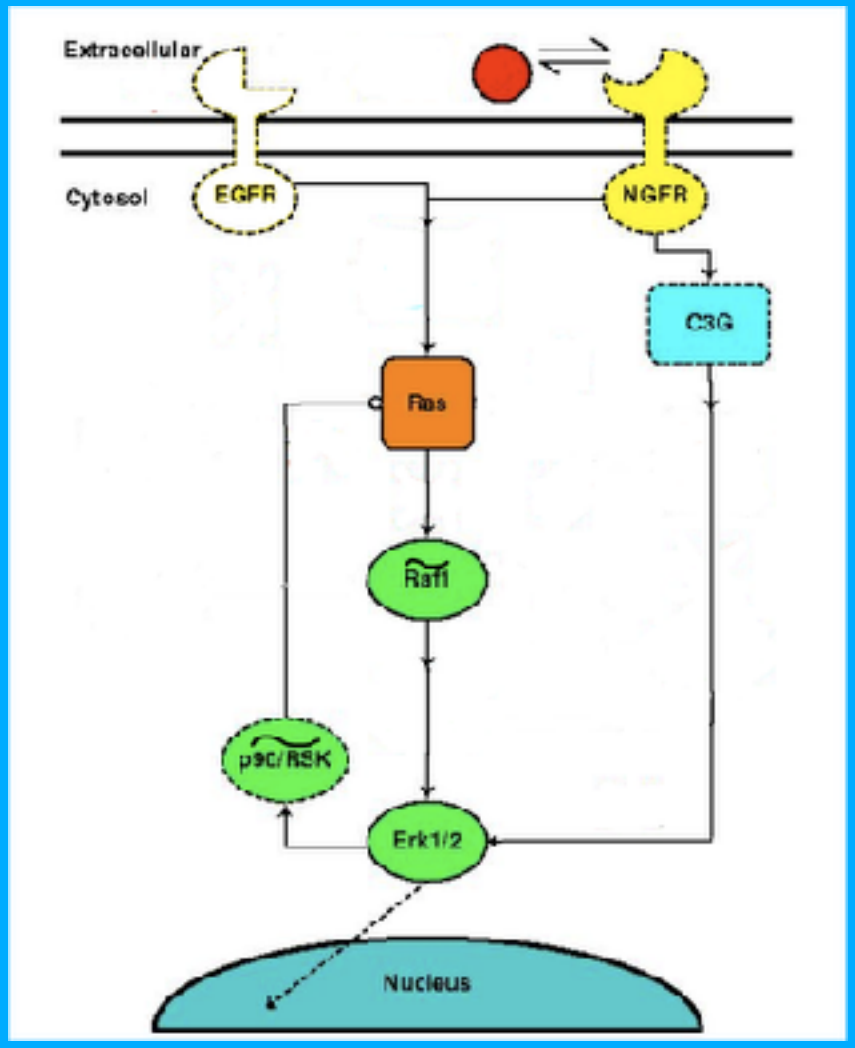
Mark Transtrum (not me)



48 params
29 ODEs

Generation of Reduced Models

Mark Transtrum (not me)



12 params
6 ODEs

$$
\begin{aligned}
[\mathrm{bEGFR}] &= \begin{cases} 1 & \text{EGF Present} \\ 0 & \text{Otherwise} \end{cases} \\
\frac{d}{dt}[\mathrm{bNGFR}] &= \theta_1[\mathrm{NGF}][\mathrm{tNGFR}] \\
\frac{d}{dt}[\mathrm{NGF}] &= -\theta_1[\mathrm{NGF}][\mathrm{tNGFR}] \\
\\
\frac{d}{dt}[\mathrm{RasA}] &= -[\mathrm{RasA}][\widetilde{\mathrm{P90RskA}}] - \theta_2[\mathrm{bEGFR}] - \theta_3[\mathrm{bNGFR}] \\
\frac{d}{dt}[\widetilde{\mathrm{Raf1A}}] &= \theta_4[\mathrm{RasA}] - \theta_5[\widetilde{\mathrm{Raf1A}}]/([\widetilde{\mathrm{Raf1A}}] - \theta_6) \\
\frac{d}{dt}[\mathrm{C3GA}] &= \theta_7[\mathrm{bNGFR}][\mathrm{C3GI}] \\
[\mathrm{Rap1A}] &= \theta_9[\mathrm{C3GA}] \\
[\mathrm{MekA}] &= [\widetilde{\mathrm{Raf1A}}][\mathrm{MekI}] + \theta_9[\mathrm{Rap1A}] \\
\frac{d}{dt}[\mathrm{Erk}] &= -\theta_{10}[\mathrm{ErkA}] + \theta_{11}[\mathrm{MekA}][\mathrm{ErkI}] \\
\frac{d}{dt}[\widetilde{\mathrm{P90RskA}}] &= \theta_{12}[\mathrm{ErkA}]
\end{aligned}
$$

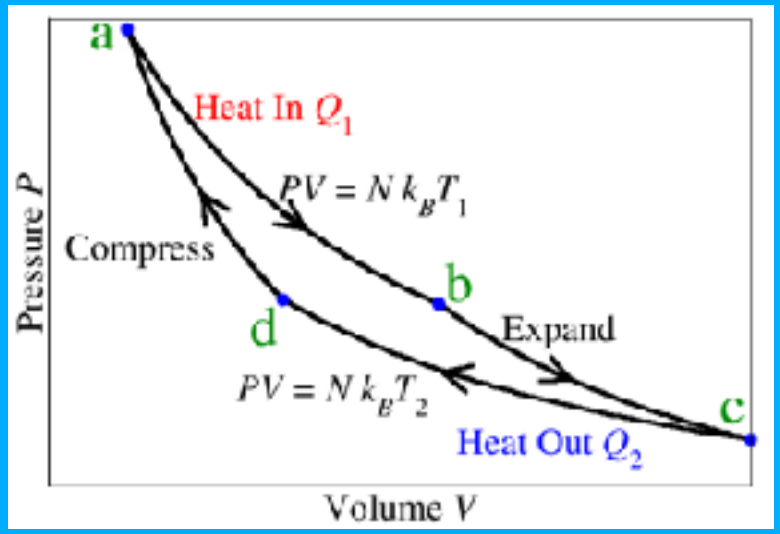
Reduced model fits all experimental data

$$\theta_9 = \frac{[\mathrm{BRafI}]\,\mathrm{kRap1toBRaf}\,\mathrm{KmdBRAF}\,\mathrm{kpBRaf}\,\mathrm{KmdMek}}{[\mathrm{PP2AA}]\,[\mathrm{Raf1PPtase}]\,\mathrm{kdBRaf}\,\mathrm{KmRap1toBRaf}\,\mathrm{kdMek}}$$

Effective ‘renormalized’ params

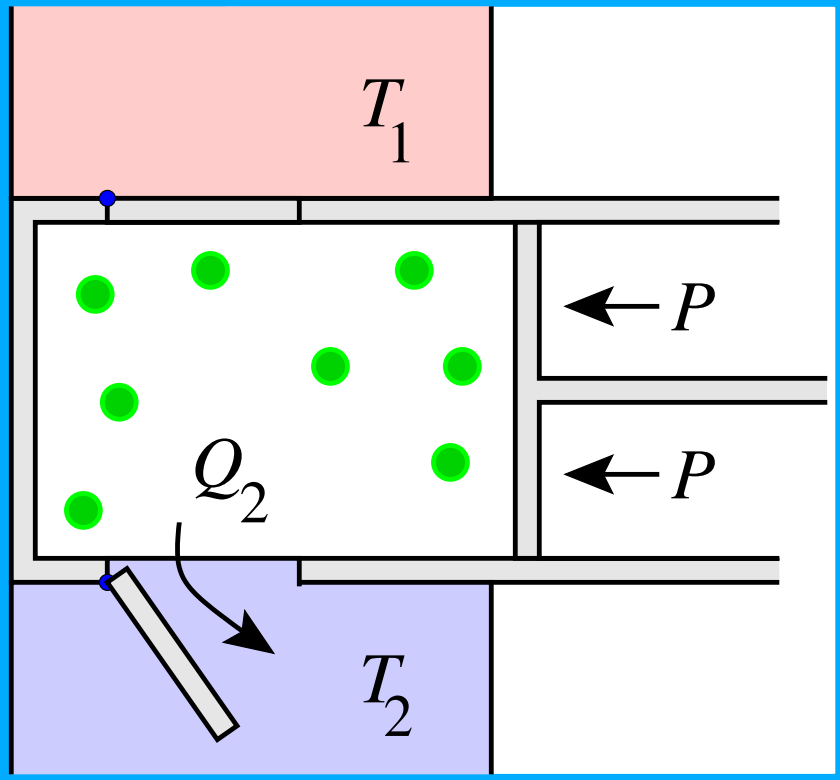
Control and Carnot Corrections

Ben Machta (not me)



Complex Carnot cycle: Isothermal Expansion, Adiabatic, Change Bath, Isothermal, Adiabatic

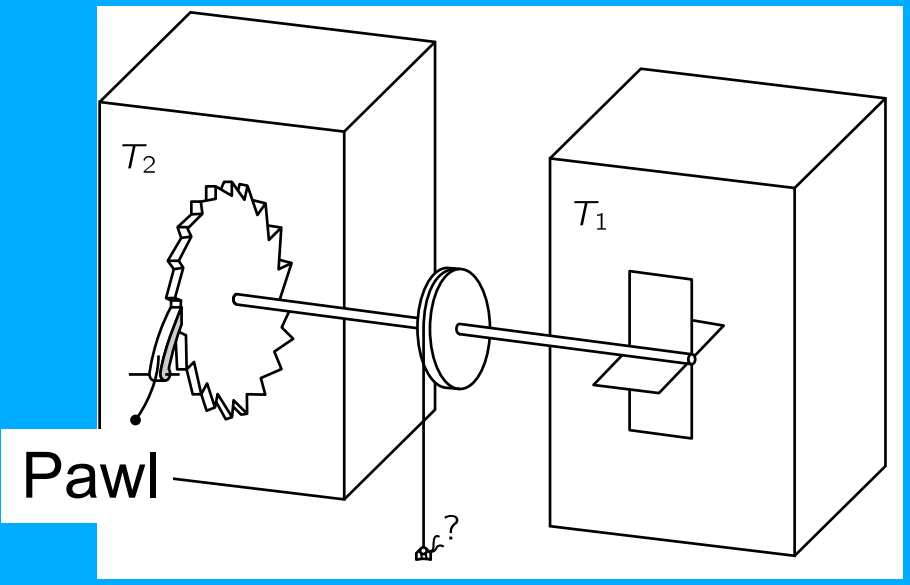
What is the entropy cost for the control system?



Incorporate control system into Hamiltonian
Move P(t), β(t) = 1/kT around controlled loop

Feynman Ratchet & Pawl

and Molecular Motors



Pawl rectifies random thermal motion

Hot gas $T_1 > T_2$ mass rises

But cold gas $T_1 < T_2$, pawl hops, mass falls!

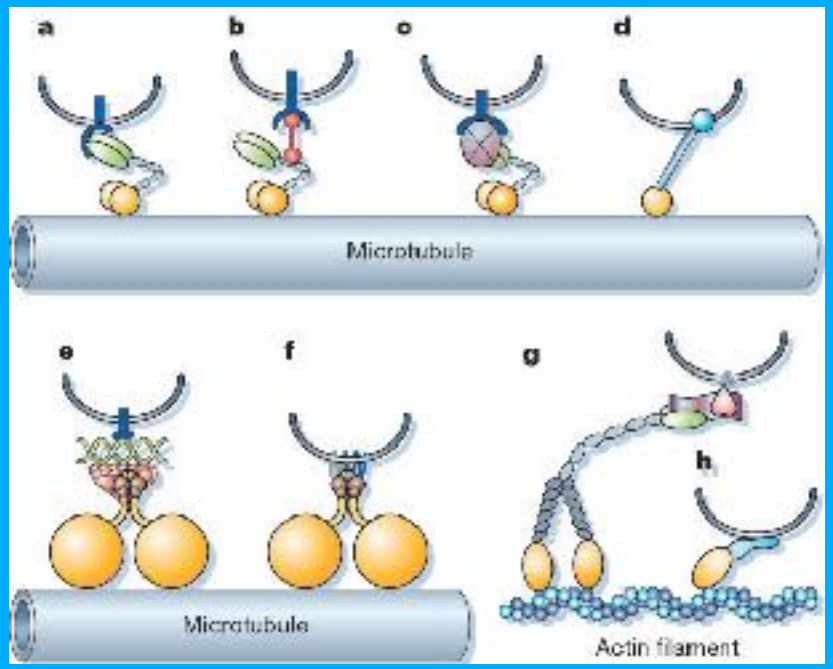
$T_1 = T_2$ Ratchet satisfies detailed balance

Molecular motors

Driven by ATP

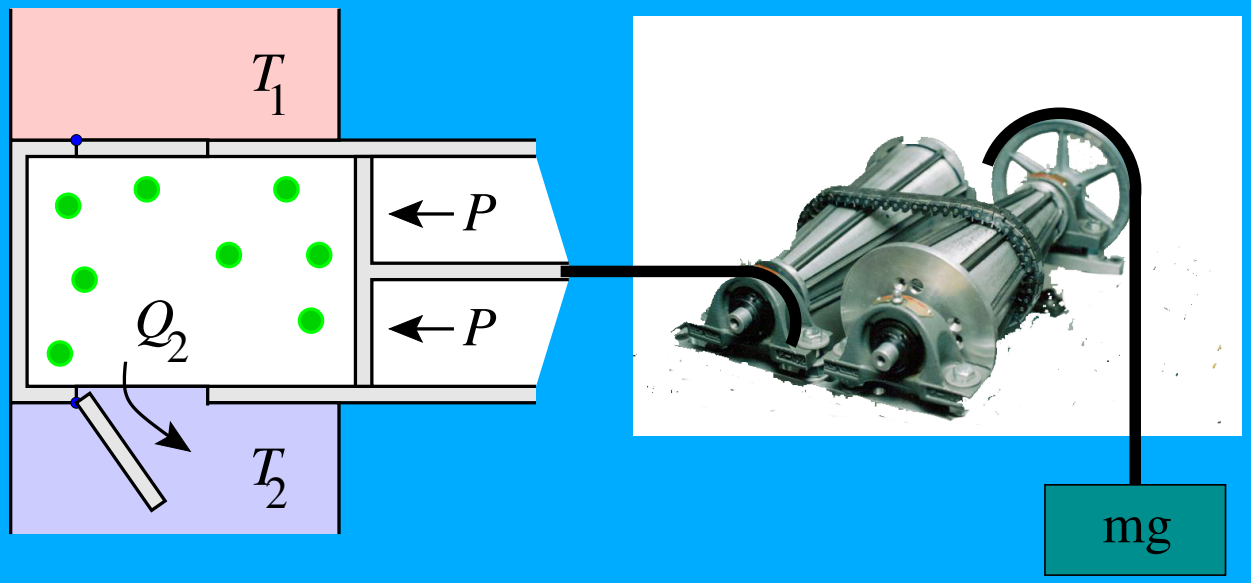
Weak driving — random hops back and forth

Strong driving, steady walk



Control and Carnot Corrections

Ben Machta (not me)



Pressure Control
Continuously variable transmission
Mass falls to drive piston (fast costs S)
Pressure fluctuations allow mass to fall (slow costs S)

$$\langle \Delta S_{\text{control}} \rangle = 2 \int_{P_i}^{P_f} \sqrt{g_{PP}} dP$$

$$g_{\mu\nu} = -\left\langle \frac{\partial^2 \log(\rho)}{\partial\theta_\mu \partial\theta_\nu} \right\rangle = \text{FIM}$$

Fisher Information Matrix:
Natural metric (distance) between probability distributions

Exercise 6.23

Entropy cost for running engines (Machta, Raju, Quinn, Clement)

$$\rho_{\mathrm{Gibbs}} = (1/\Gamma)\exp(-\beta\mathcal{H}(\mathbb{P},\mathbb{Q}) - \beta PV).$$

Show $\log(\rho)$ is $\beta G(P,\beta)$ plus terms linear in $p = \beta P$ and β.

Exponential family.

(e) *For a collection of particles interacting with Hamiltonian* $\mathcal{H}$*, using* $\mathbf{\Phi} = (p,\beta) = (P\beta,\beta)$*, relate the four terms*

$$g^{\mathrm{p}\beta}_{\alpha\beta} = -\langle \partial^2 \log(\rho)/\partial\phi_\mu \partial\phi_\nu \rangle$$

in terms of physical quantities given by the second derivatives of G*.* (Hint: $dG = -SdT + VdP + \mu dN$)

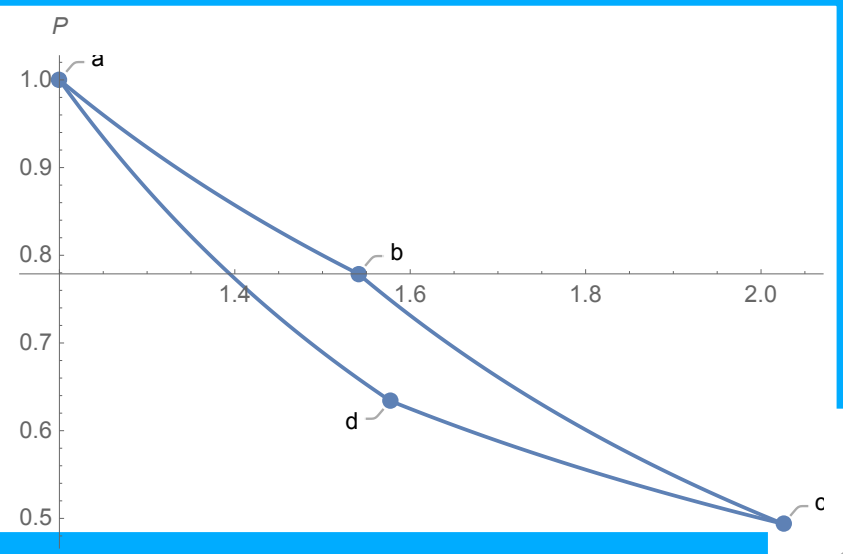
(d) *Compute the* 4×4 *matrix* $g^{\mathrm{p}\beta}$ *explicitly for the ideal gas.*

Control and Carnot Corrections

Ben Machta (not me)

$$\langle \Delta S_{\text{control}} \rangle = 4\sqrt{N}\log(P_a/P_b) + 2\sqrt{15N}\log(T_1/T_2)$$

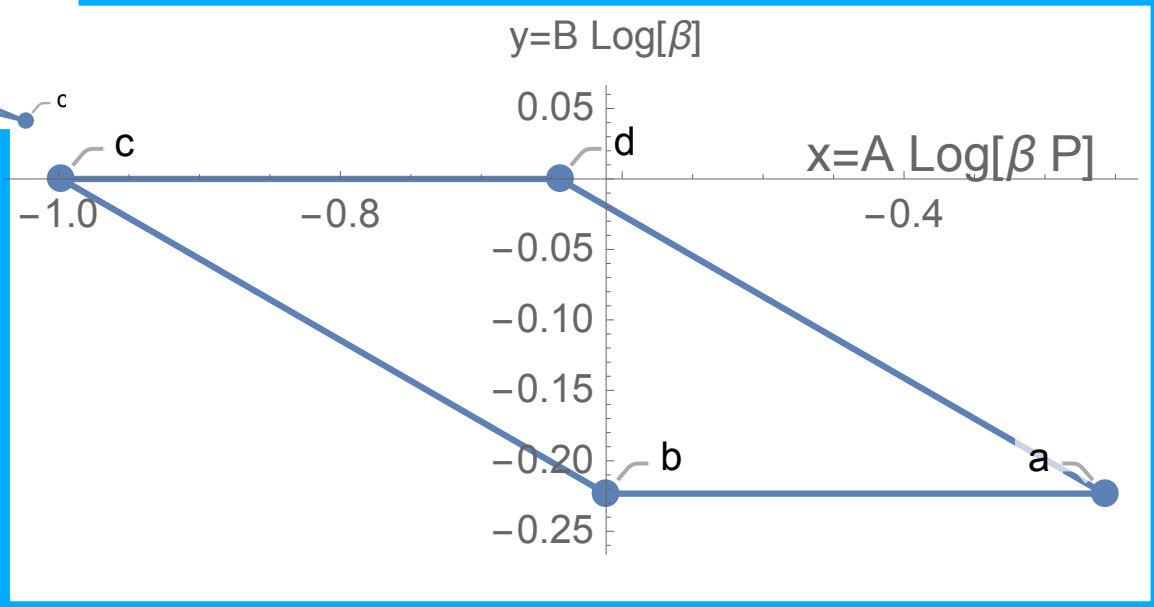
Subextensive:
Work ~N T_1 log(P_2/P_1)
Exponential family variables simpler:
p = β P = P/kT
Zero curvature to metric: Model manifold flat:
(x,y) = (A log(β P), B log(β))



$$g^{(P,\beta)}_{\mu\nu} = \begin{pmatrix} N/P^2 & N/\beta P \\ N/\beta P & 5N/2\beta^2 \end{pmatrix}$$

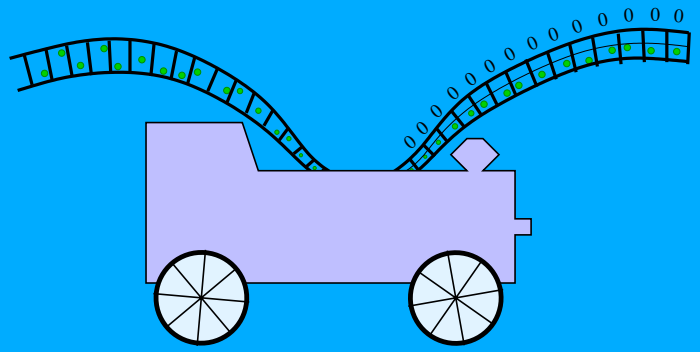
$$g^{(p,\beta)}_{\mu\nu} = \begin{pmatrix} N/p^2 & 0 \\ 0 & 3N/2\beta^2 \end{pmatrix}$$

$$g^{(x,y)}_{\mu\nu} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$



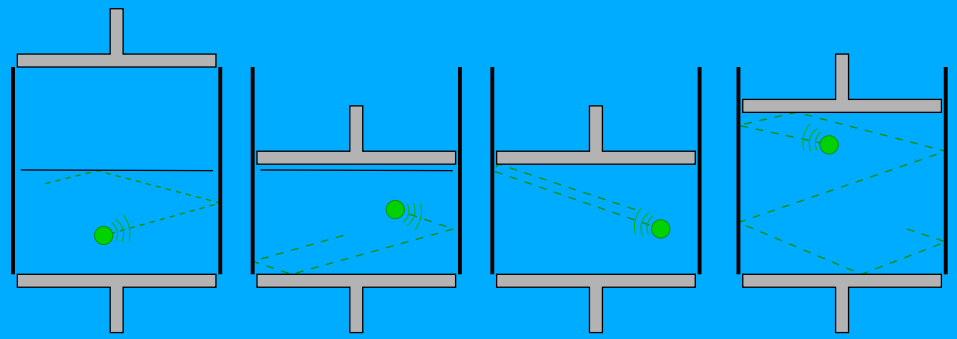
Control and Carnot Corrections

Ben Machta (not me)

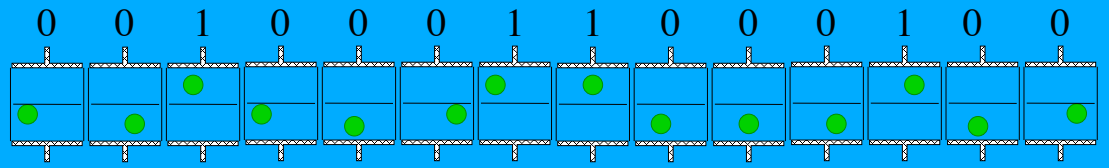


Szilard Engine
(Bennett, Feynman)

Work done by expanding piston is
kT log 2 = T ΔS



Szilard argued that information can be exchanged for work. We saw this in the ‘Burning information’ exercise, where a data tape stored bits by putting atoms on one side or another of a partitioned piston.



Machta argues that extracting this work will cost entropy 4 k log 2, a net loss!

$$\langle \Delta S_{\text{control}} \rangle = 4\sqrt{N} \log(P_a/P_b) \cdot$$

Big Questions

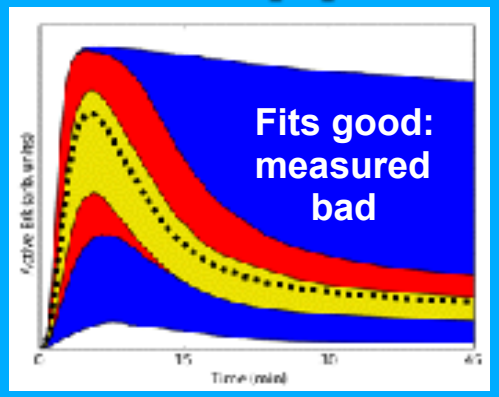
Consequences of Sloppiness

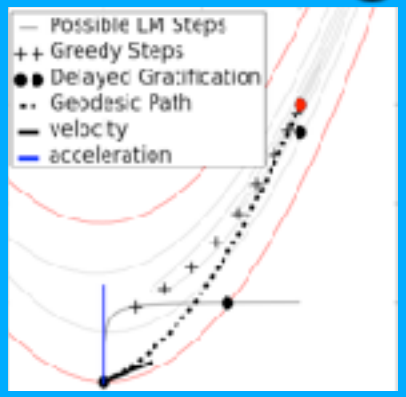
- Efficacy of principal component analysis
- Least squares fits: efficacy of Levenberg-Marquardt algorithm
- Biological evolution: sloppy or robust? Neutral spaces, evolvability
- Does pattern recognition work because of hyperribbons?
- Does big data work because of hyperribbons?
- Is science possible because of hyperribbons?
- Different models can describe the same behavior
- Why is the world comprehensible?

Sloppy Applications

Several applications emerge

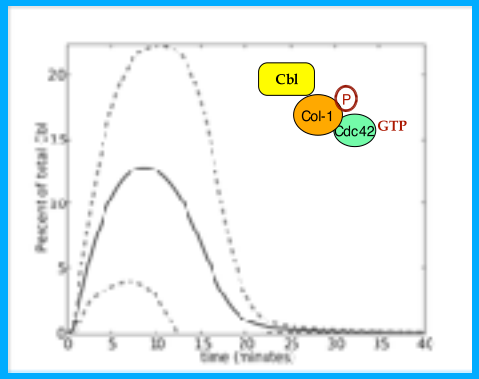
A. Fitting data vs. measuring parameters (Gutenkunst)



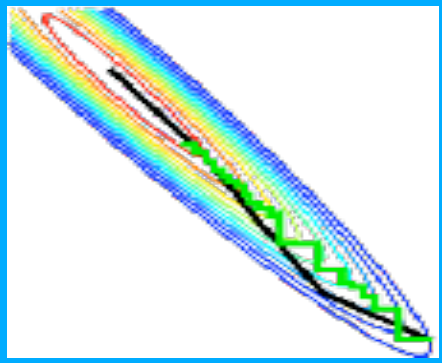


B. Finding best fits by geodesic acceleration (Transtrum)

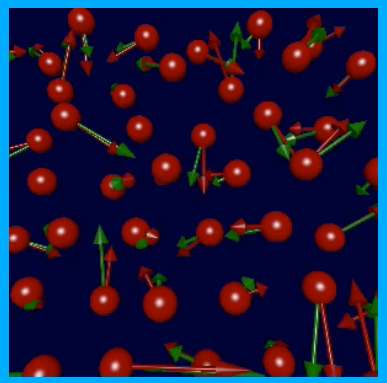
C. Optimal experimental design (Casey)



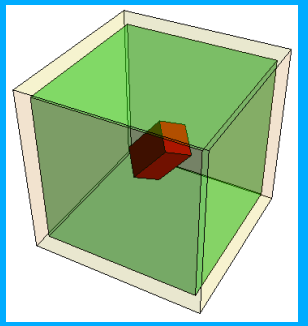
D. Sloppy fitness and evolution (Gutenkunst)



E. Estimating systematic errors: (Jacobsen et al.)

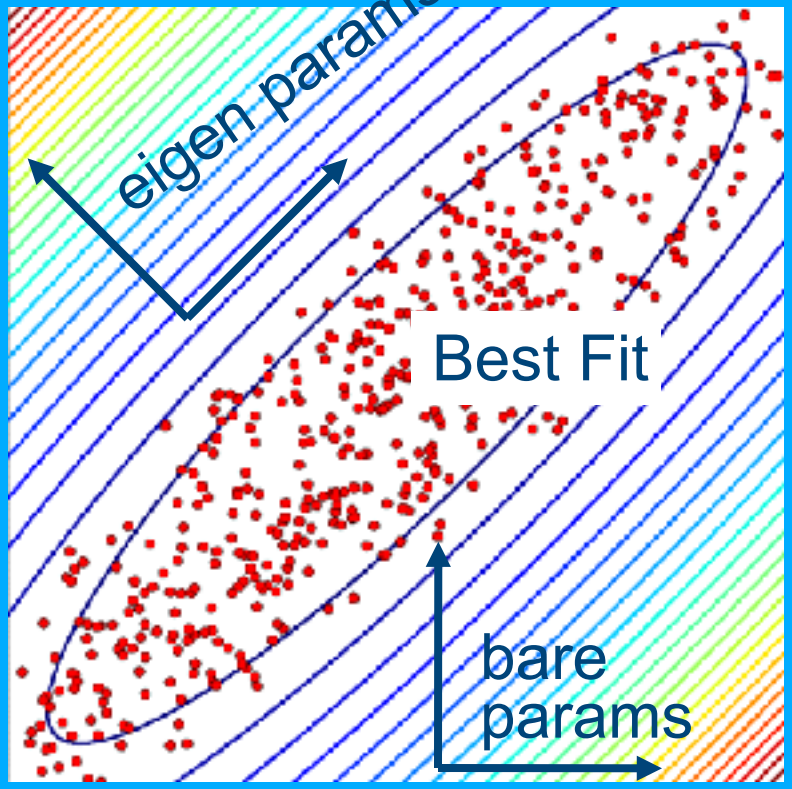


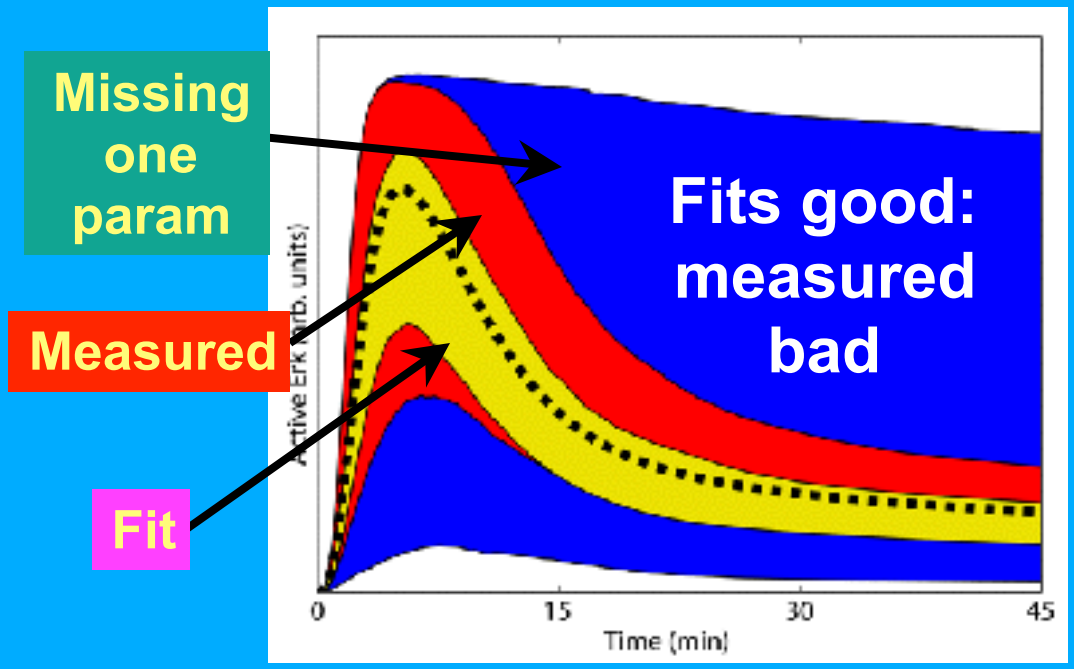
F. Sloppiness and Robustness



A. Are rate constants useful?

Fits vs. measurements

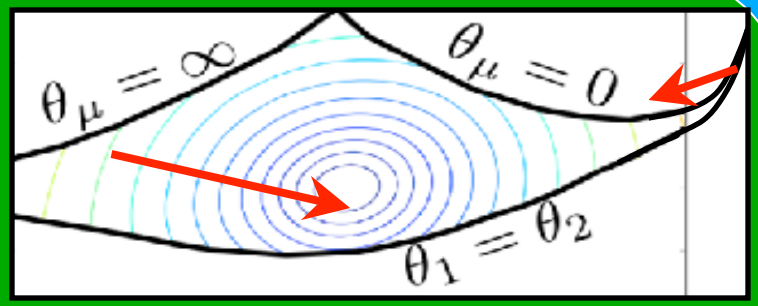




Monte Carlo (anharmonic)

- Easy to Fit (14 expts); Measuring huge job (48 params, 25%)
- One missing parameter measurement = No predictivity
- Sloppy Directions = Enormous Fluctuations in Parameters
- Sloppy Directions often do not impinge on predictivity

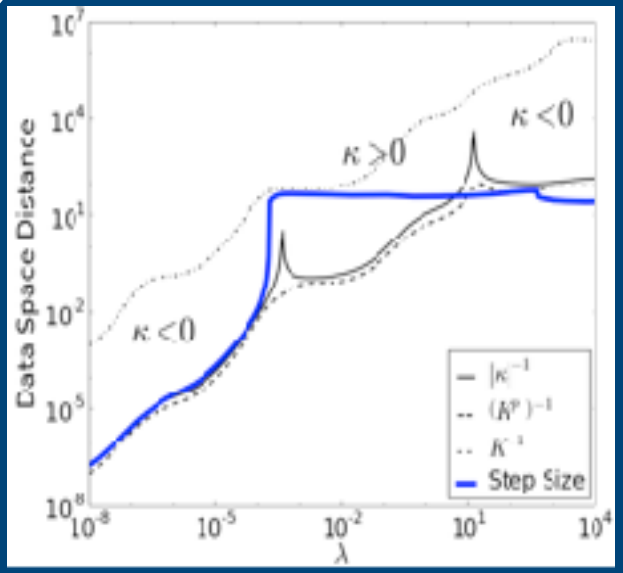
B. Finding best fits: Geodesic acceleration



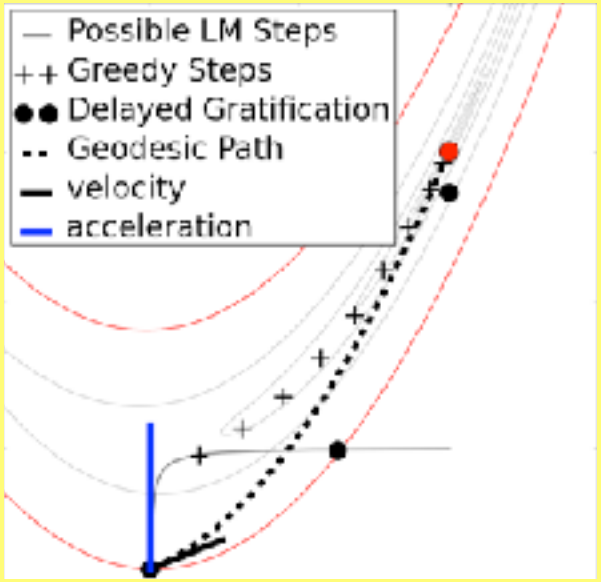
Geodesic Paths nearly circles
Follow local geodesic velocity?

$$\delta\theta^\mu = -g_{\mu\nu}\nabla_\nu C$$

➔ Gauss-Newton
➔ Hits manifold boundary



Model Graph
add weight λ of parameter metric yields Levenberg-Marquardt: Step size now limited by curvature



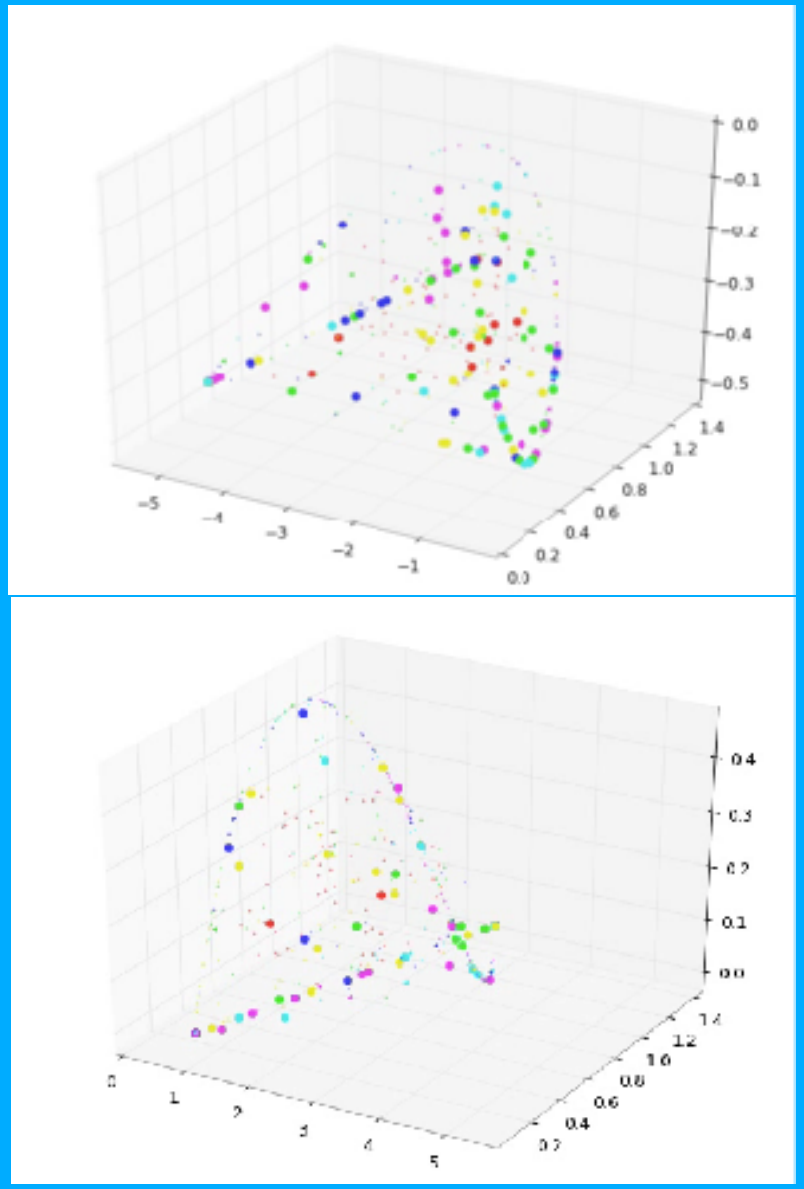
Algorithm	Success Rate	Mean njev	Mean nfev
Traditional LM + accel	65%	258	1494
Traditional LM	33%	2002	4003
Trust Region LM	12%	1517	1649
BFGS	8%	5363	5365

Follow parabola, ***geodesic acceleration***
Cheap to calculate; faster; more success

B. Finding best fits: Model manifold dynamics (Isabel Kloumann)

Dynamics on the model manifold: Searching for the best fit

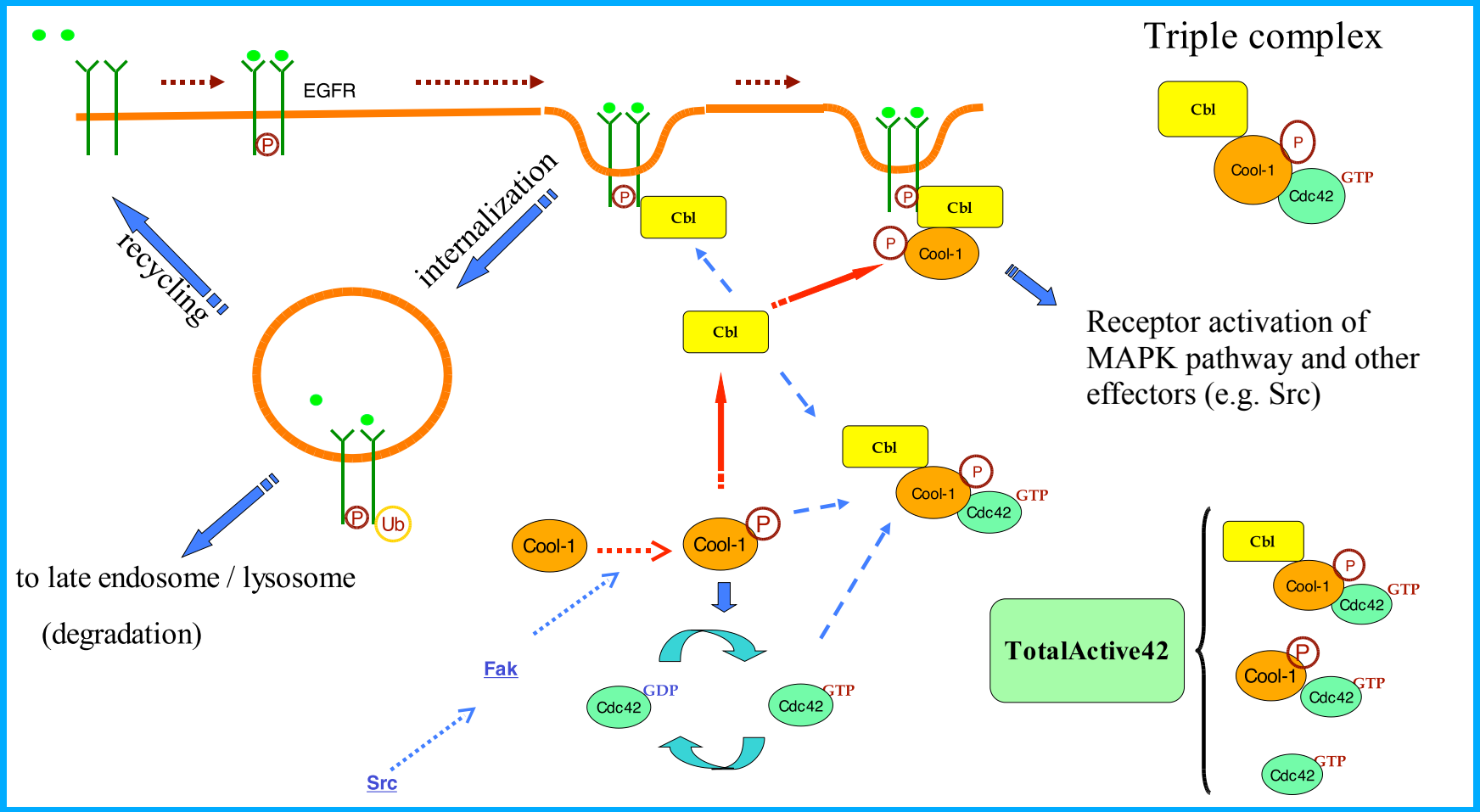
- Jeffrey's prior plus noise
- Big noise concentrates on manifold edges
- Note scales: flat
- Top: Levenberg-Marquardt
- Bottom: Geodesic acceleration
- Large points: Initial conditions which fail to converge to best fit



C. EGFR Trafficking Model

Fergal Casey, Cerione lab

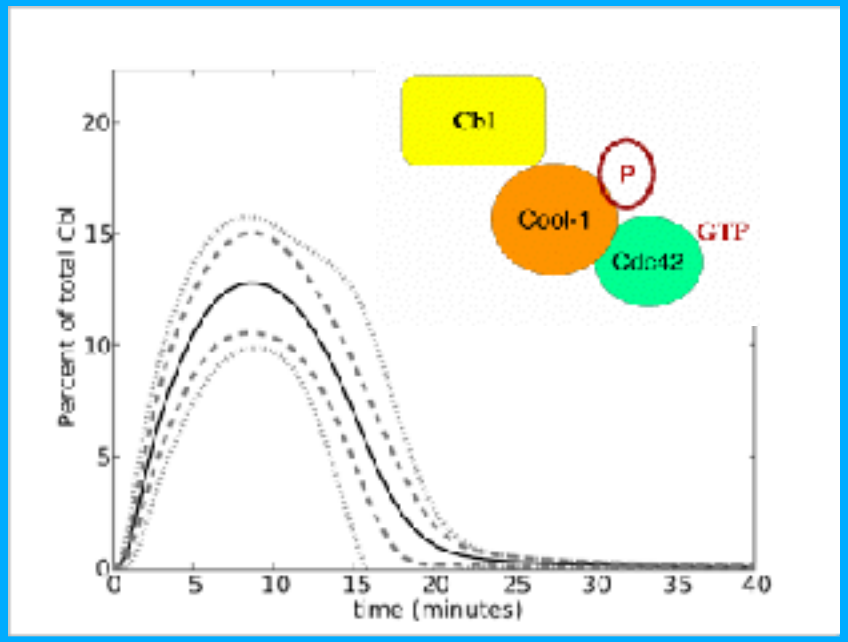
- Cerione lab: testing hypothesis, experimental design (Cool1≡β-PIX)
- 41 chemicals, 53 rate constants; only 11 of 41 species can be measured
- Does Cool-1 triple complex sequester Cbl, delay endocytosis in wild type NIH3T3 cells?

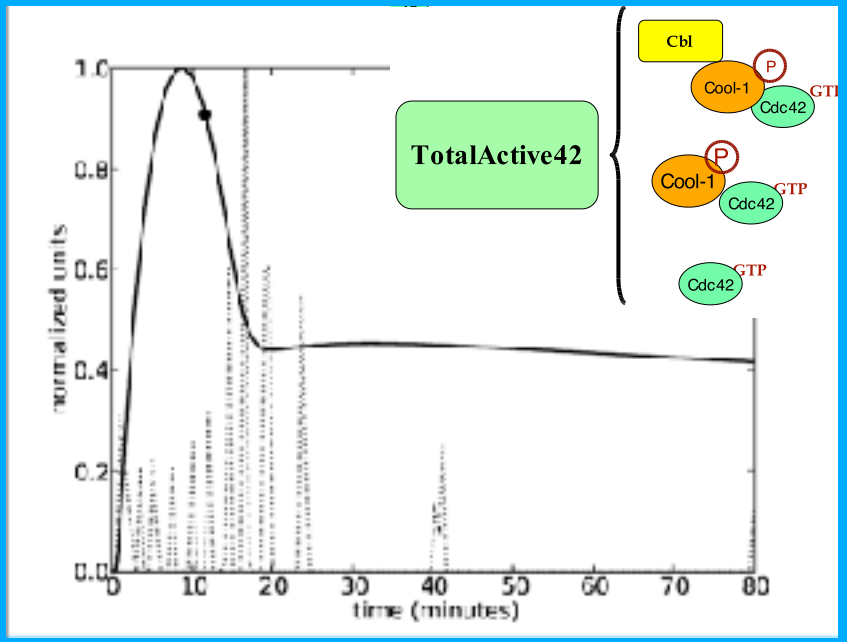


C. Trafficking: experimental design

Which experiment best reduces prediction uncertainty?

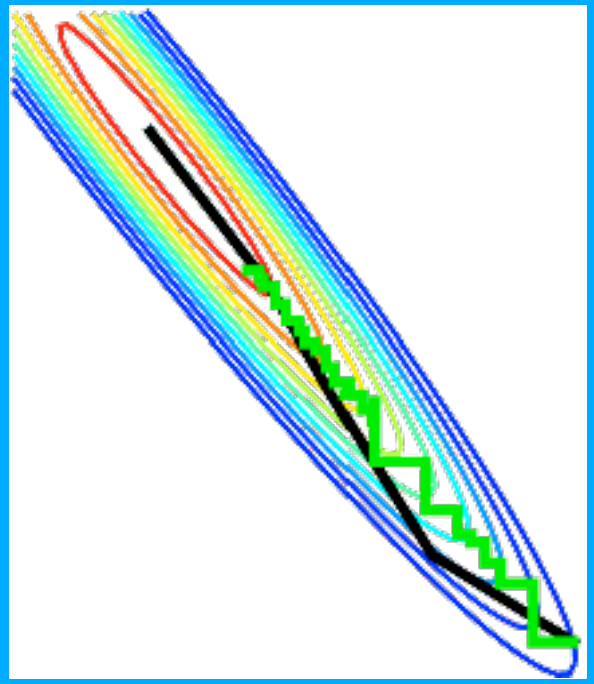
- Amount of triple complex was not well predicted
- V-optimal experimental design: single & multiple measurements
- Total active Cdc42 at 10 min.; Cerione independently concurs
- Experiment indicates significant sequestering in wild type
- Predictivity without decreasing parameter uncertainty

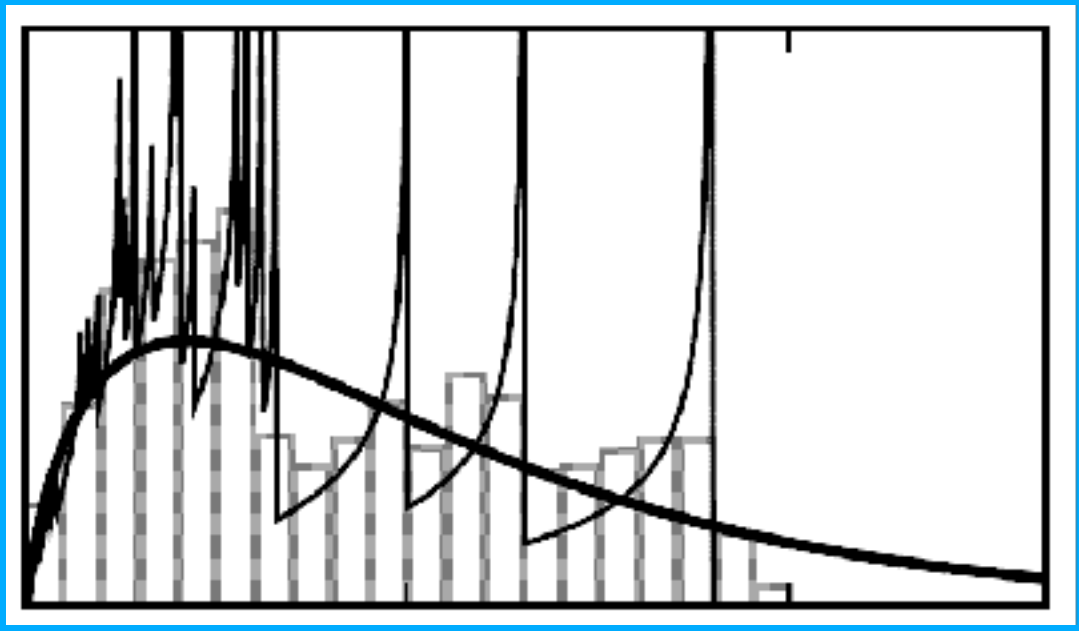




D. Evolution in Chemotype space

Implications of sloppiness?





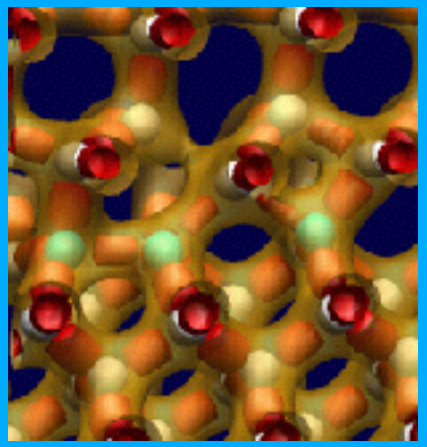
Fitness gain from first successful mutation

- Culture of identical bacteria, one mutation at a time
- Mutation changes one or two rate constants (no *pleiotropy*): orthogonal moves in rate constant (chemotype) space
- ***Cusps*** in first fitness gain (one for each rate constant, big gap)
- Multiple mutations get stuck on ridge in sloppy landscape

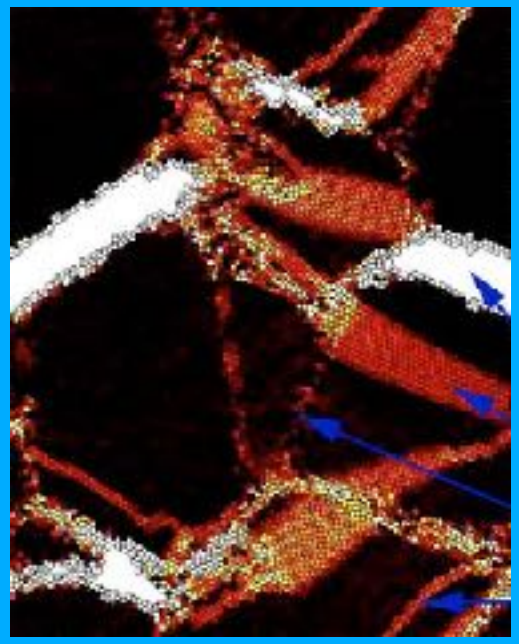
E. Bayesian Errors for Atoms

'Sloppy Model' Approach to Error Estimation of Interatomic Potentials

Søren Frederiksen, Karsten W. Jacobsen, Kevin Brown, JPS



Quantum Electronic Structure (Si) 90 atoms (Mo) (Arias)



Atomistic potential 820,000 Mo atoms (Jacobsen, Schiøtz)

Interatomic Potentials $V(\mathbf{r}_1,\mathbf{r}_2,\ldots)$

- Fast to compute
- Limit $m_e/M \rightarrow 0$ justified
- Guess functional form
  Pair potential $\sum V(\mathbf{r}_i-\mathbf{r}_j)$ poor
  Bond angle dependence
  Coordination dependence
- Fit to experiment (old)
- Fit to forces from electronic structure calculations (new)

17 Parameter Fit

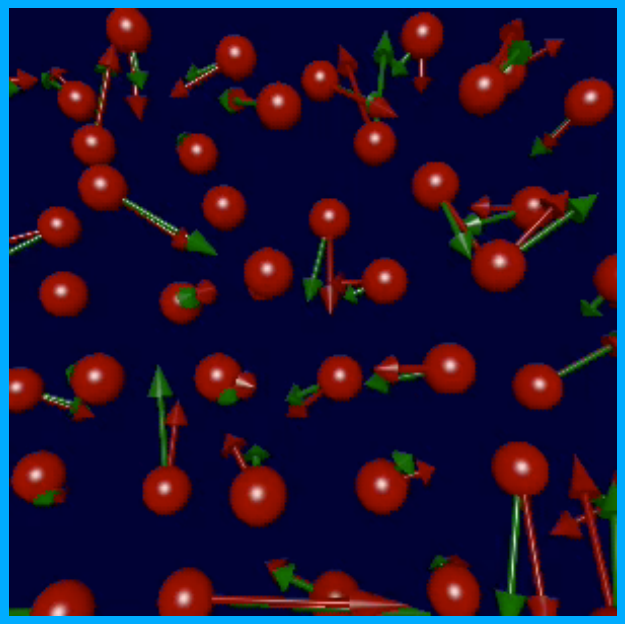
E. Interatomic Potential Error Bars

Ensemble of Acceptable Fits to Data

Not ***transferable***

Unknown errors

- 3% elastic constant
- 10% forces
- 100% fcc-bcc, dislocation core



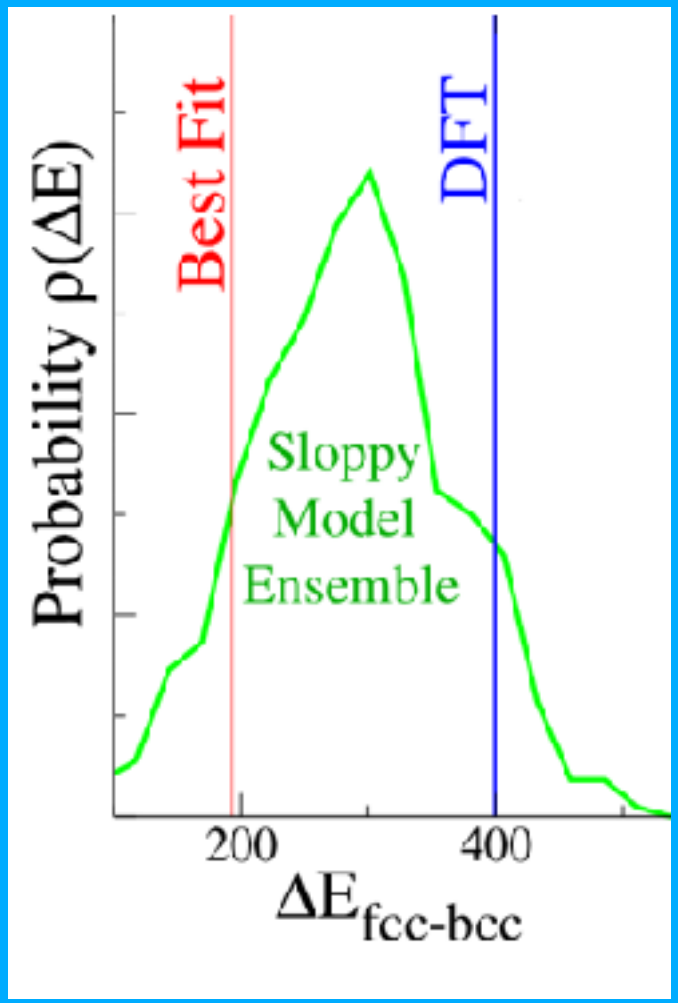
Green = DFT, Red = Fits

Best fit is ***sloppy***: ensemble of fits that aren't much worse than best fit.

Ensemble in Model Space!

T_0 set by equipartition energy = best cost

Error Bars from quality of best fit



Sloppy Molybdenum: Does it Work?

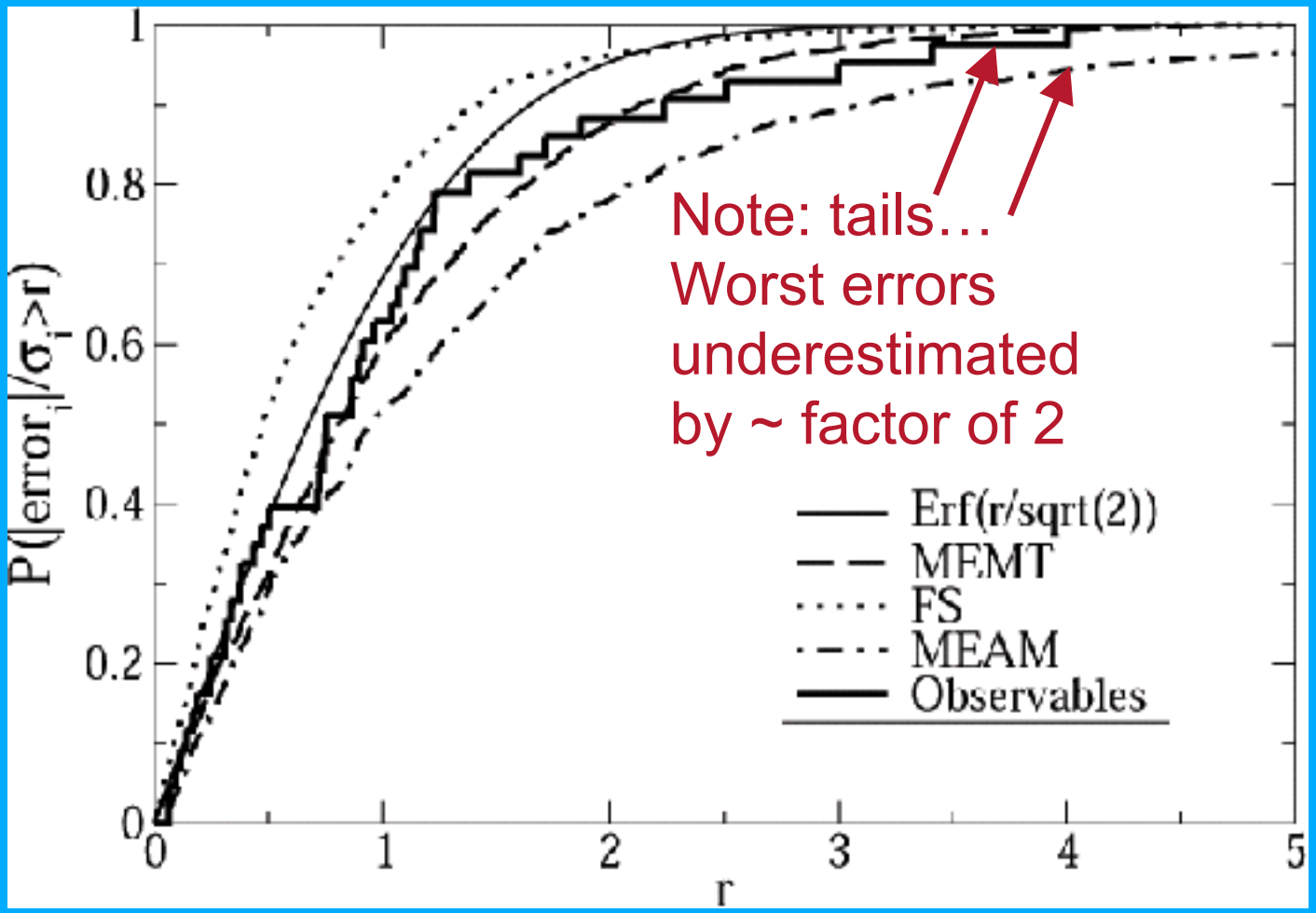
Estimating *Systematic* Errors

Bayesian error σ_i gives total error if ratio $r = \text{error}_i/\sigma_i$ distributed as a Gaussian: cumulative distribution $P(r)=Erf(r/\sqrt{2})$

Three potentials
- Force errors
- Elastic moduli
- Surfaces
- Structural
- Dislocation core
- $7\% < \sigma_i < 200\%$

"Sloppy model" systematic error most of total

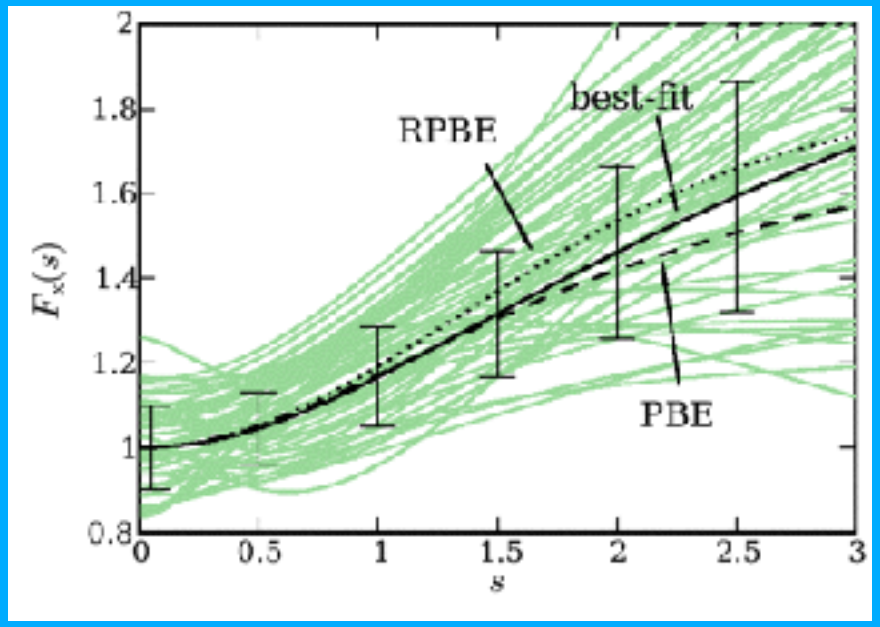
~2 << 200%/7%



Systematic Error Estimates for DFT

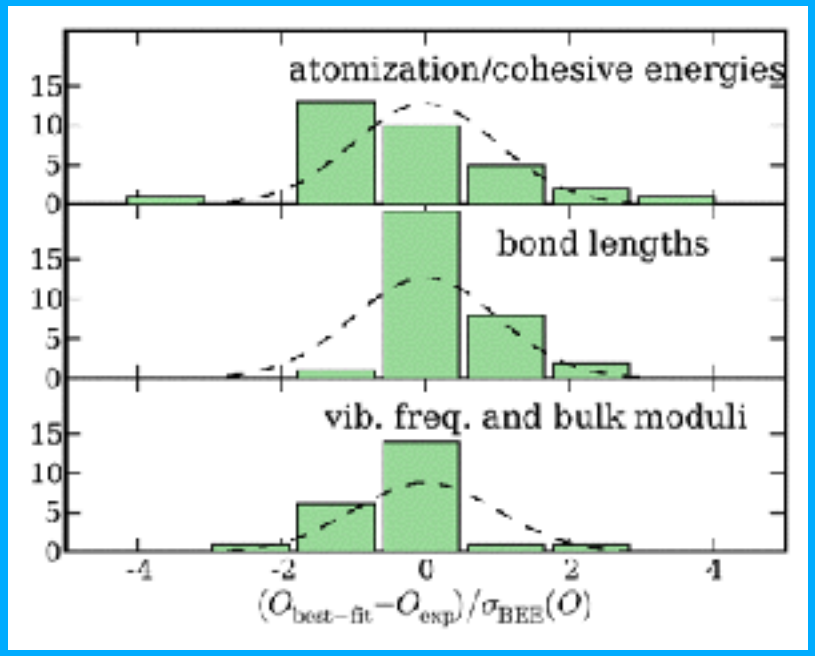
GGA-DFT as Multiparameter Fit?

J. J. Mortensen, K. Kaasbjerg, S. L. Frederiksen,
J. K. Nørskov, JPS, K. W. Jacobsen,
(Anja Tuftelund, Vivien Petzold, Thomas Bligaard)



Enhancement factor $F_x(s)$
in the exchange energy E_x

Large fluctuations



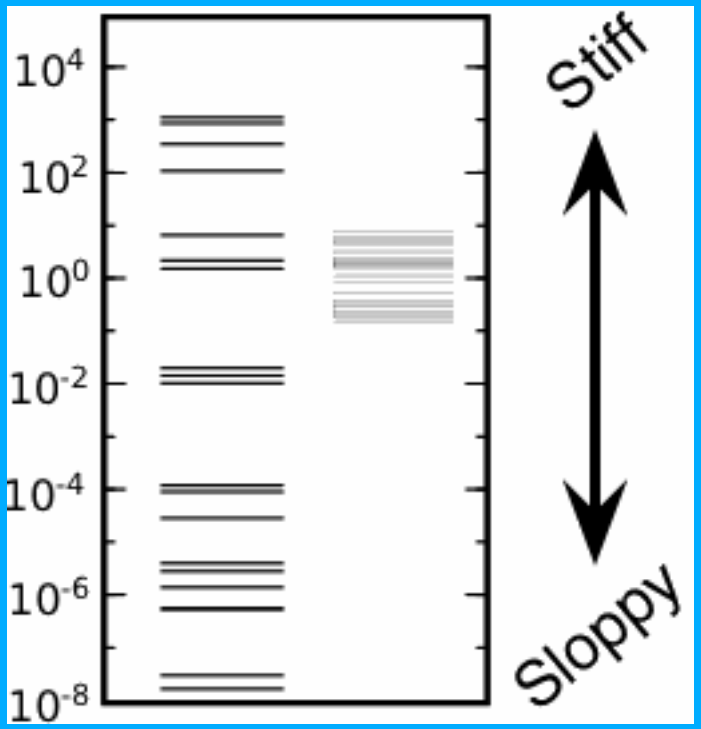
Actual error / predicted error

Deviation from experiment well described by ensemble!

F. Parameter robustness and sloppiness

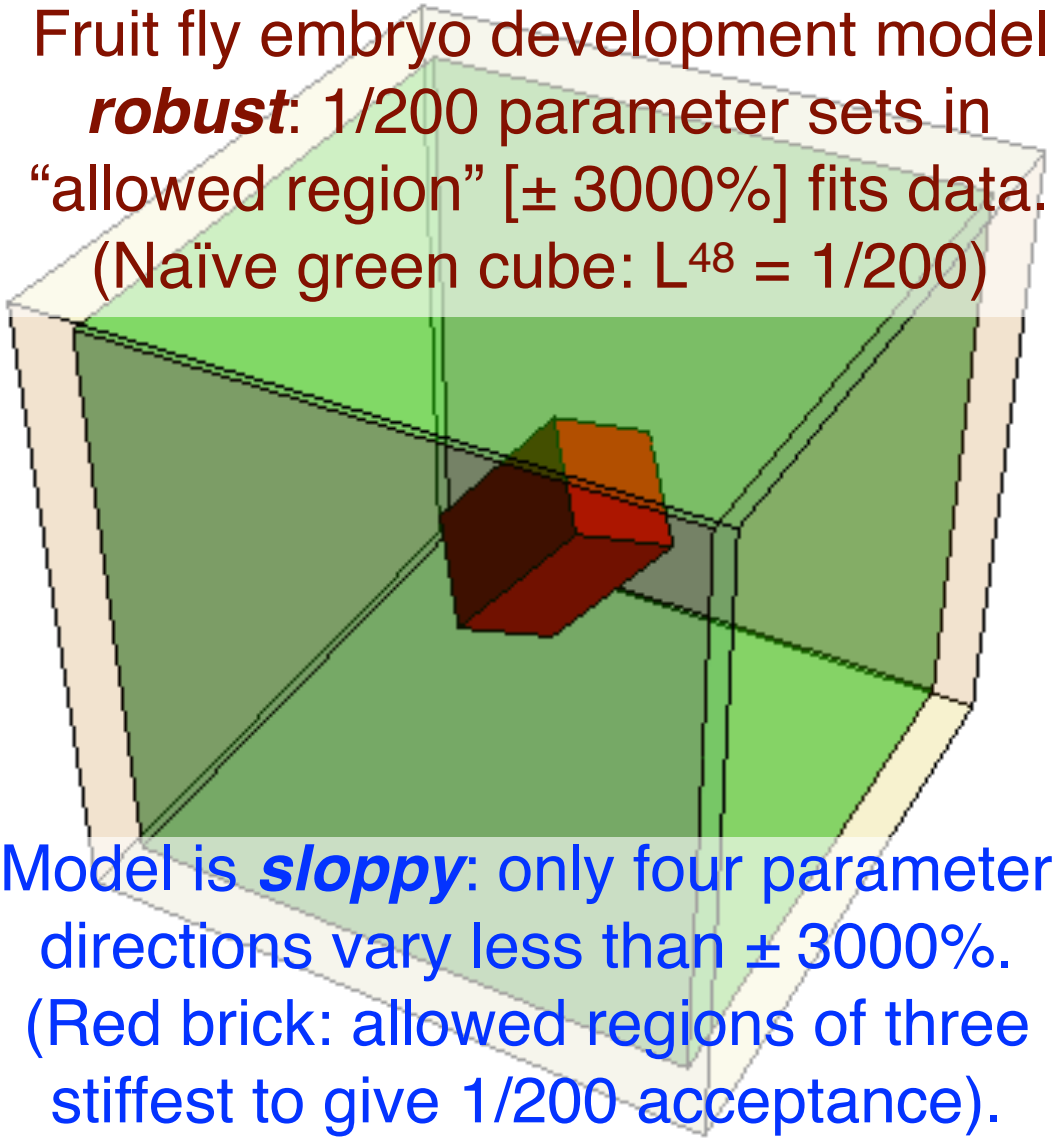
Do parameters matter at all?

Bryan Daniels, Yanjiun Chen, Ryan Gutenkunst, Chris Myers



Segment polarity model is sloppy (eigenvalues, left) and robust (PCA, phenotype-preserving in cube, right).

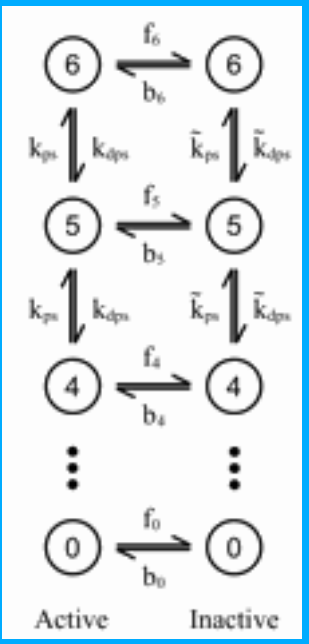
Fruit fly embryo development model ***robust***: 1/200 parameter sets in "allowed region" [± 3000%] fits data. (Naïve green cube: L^{48} = 1/200)

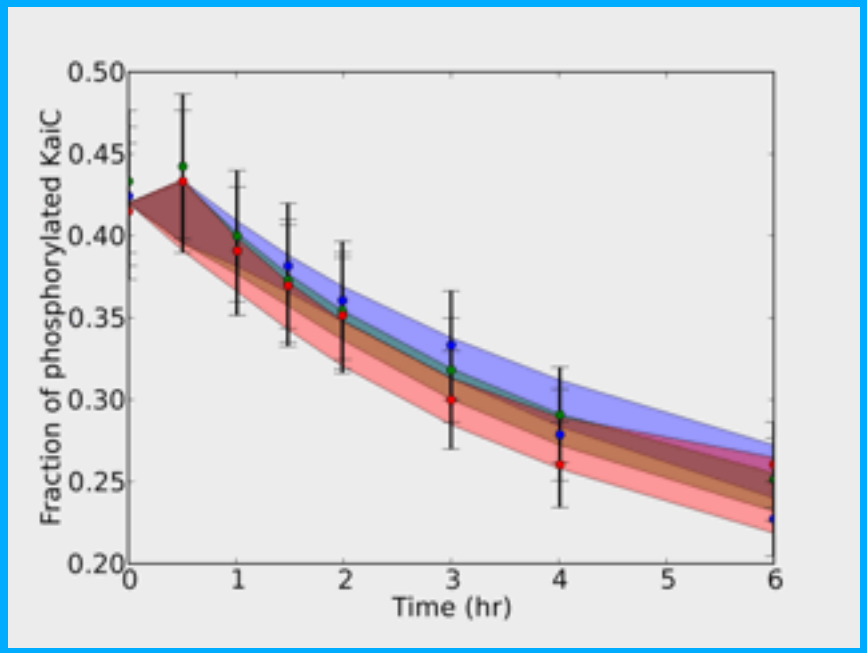


Model is ***sloppy***: only four parameter directions vary less than ± 3000%. (Red brick: allowed regions of three stiffest to give 1/200 acceptance).

F. Environmental robustness and sloppiness

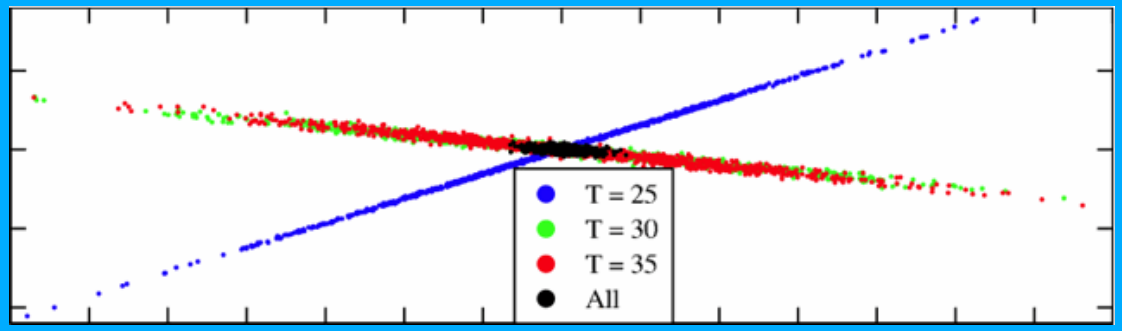
Circadian Rhythms





Bacteria know the time of day! How do they keep their clocks on time in the cold? All reaction rates exponentially dependent on temperature! Delicate cancellation?

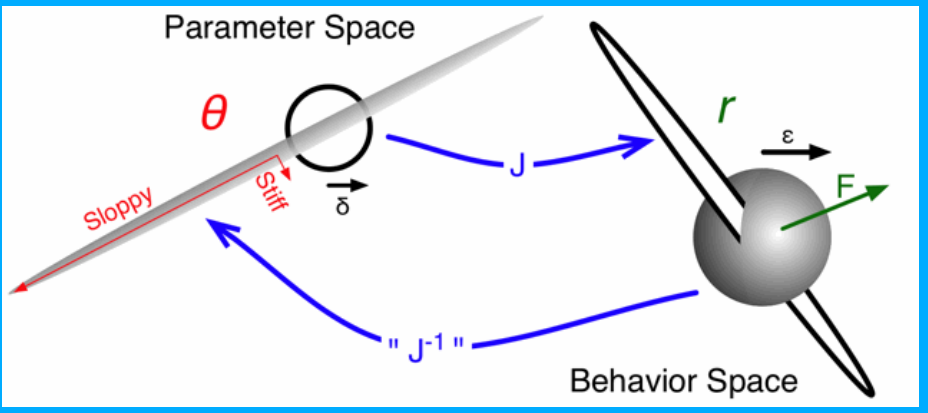
Sloppiness facilitates finding parameter values for robust response to environmental change.



Three sloppy directions, 18 rates exponentially dependent on temperature. Three stiff directions at each temperature? 18 – 3 (25°C) – 3 (30°C) – 3 (35°C) = 9 dimensional robust parameter space

F. Evolvability, robustness, and sloppiness

If it's robust, can it evolve?



How can evolution proceed if mutating the parameters doesn't matter? More robust, less evolvable!

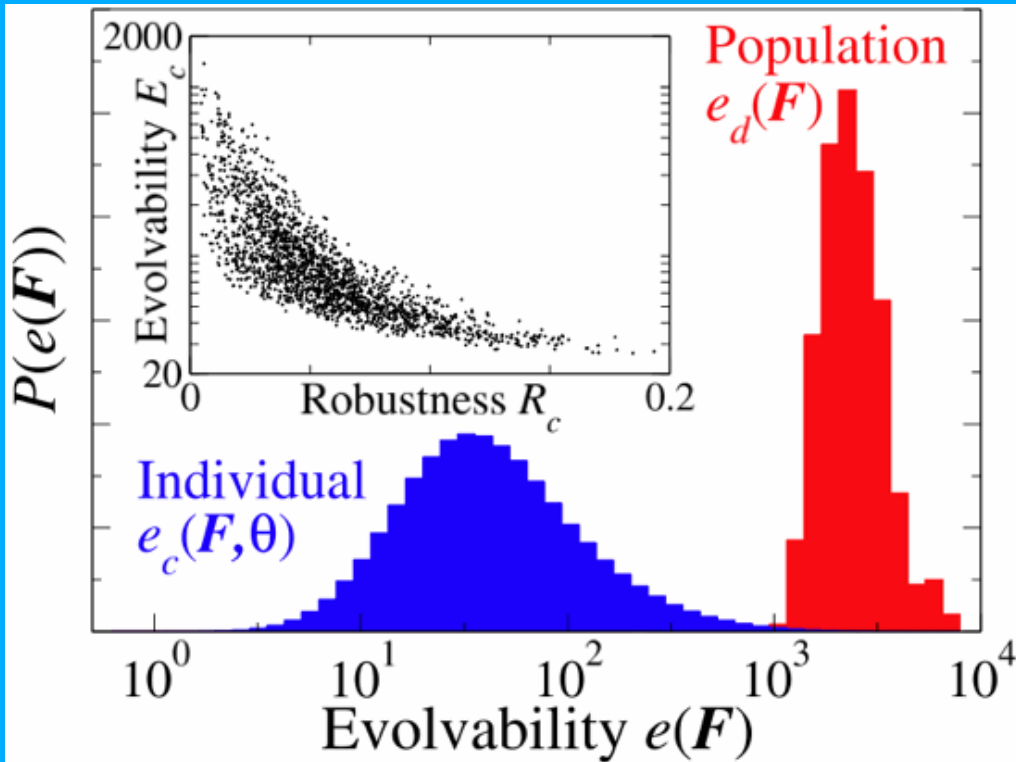
Evolutionary force F

Sloppy signal transduction model

Individual evolvabilities $e_c(F)$ within phenotype

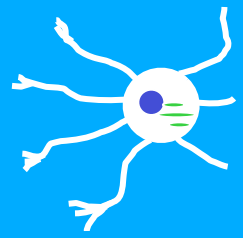
Population evolvability e_d within phenotype

[Sloppy neutral spaces allow species to explore large ranges of parameters]



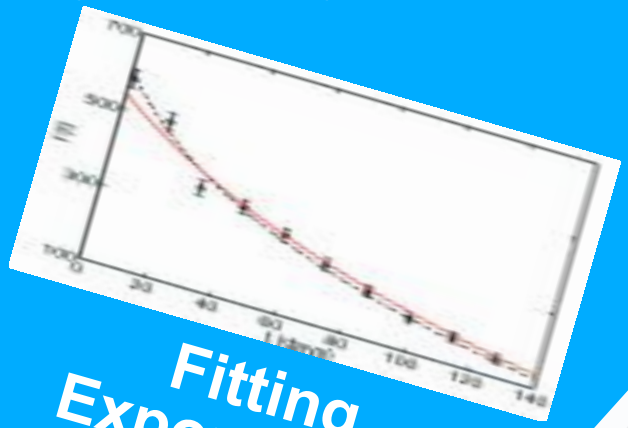
Sloppy Models, Differential geometry, and How science works

JPS, Katherine Quinn, Archishman Raju, Mark Transtrum, Ben Machta, Ricky Chachra, Kevin Brown, Ryan Gutenkunst, Josh Waterfall, Fergal Casey, Chris Myers, …

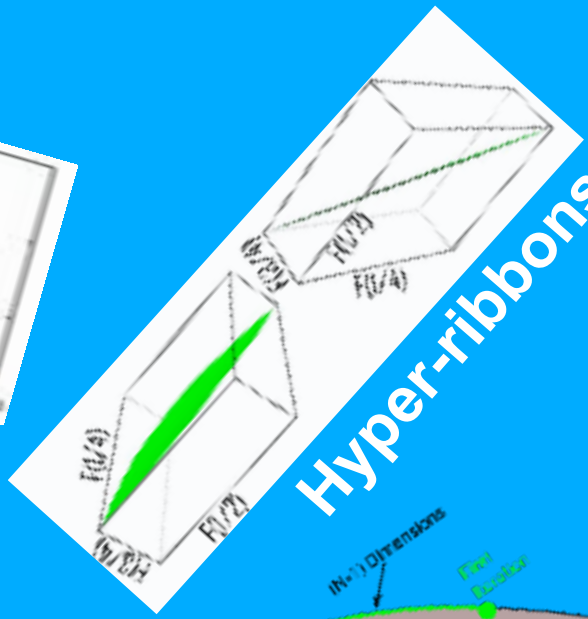


Cell Dynamics

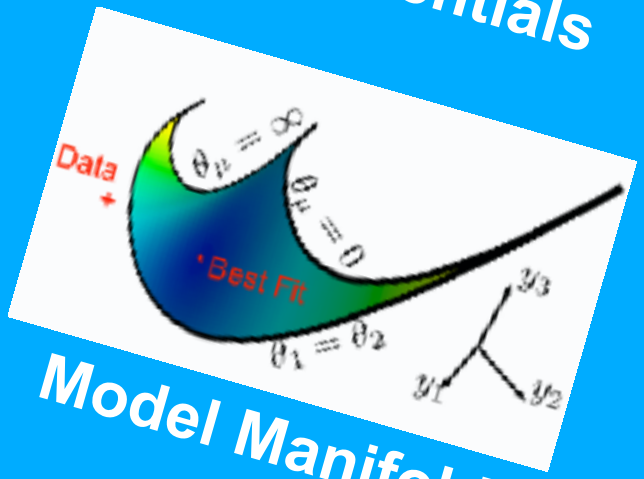
Sloppiness



Fitting Exponentials



Hyper-ribbons

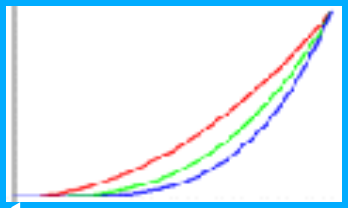


Model Manifold

Coarse-Grained Models

Proposed universal ensemble

Why are they sloppy?



Assumptions: (Not one experiment per parameter)

i. **Model predictions all depend on every parameter, *symmetrically*:** $y_i(\theta_1, \theta_2, \theta_3) = y_i(\theta_2, \theta_3, \theta_1)$

ii. **Parameters are nearly degenerate:** $\theta_i = \theta_0 + \varepsilon_i$

$$H = J^T J = V^T A^T A V$$

$$V = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_N \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_1^d & \varepsilon_2^d & \cdots & \varepsilon_N^d \end{bmatrix}$$

Vandermonde Matrix

$$\det(V) = \prod_{i<j} (\varepsilon_i - \varepsilon_j) \propto \varepsilon^{N(N-1)/2}$$

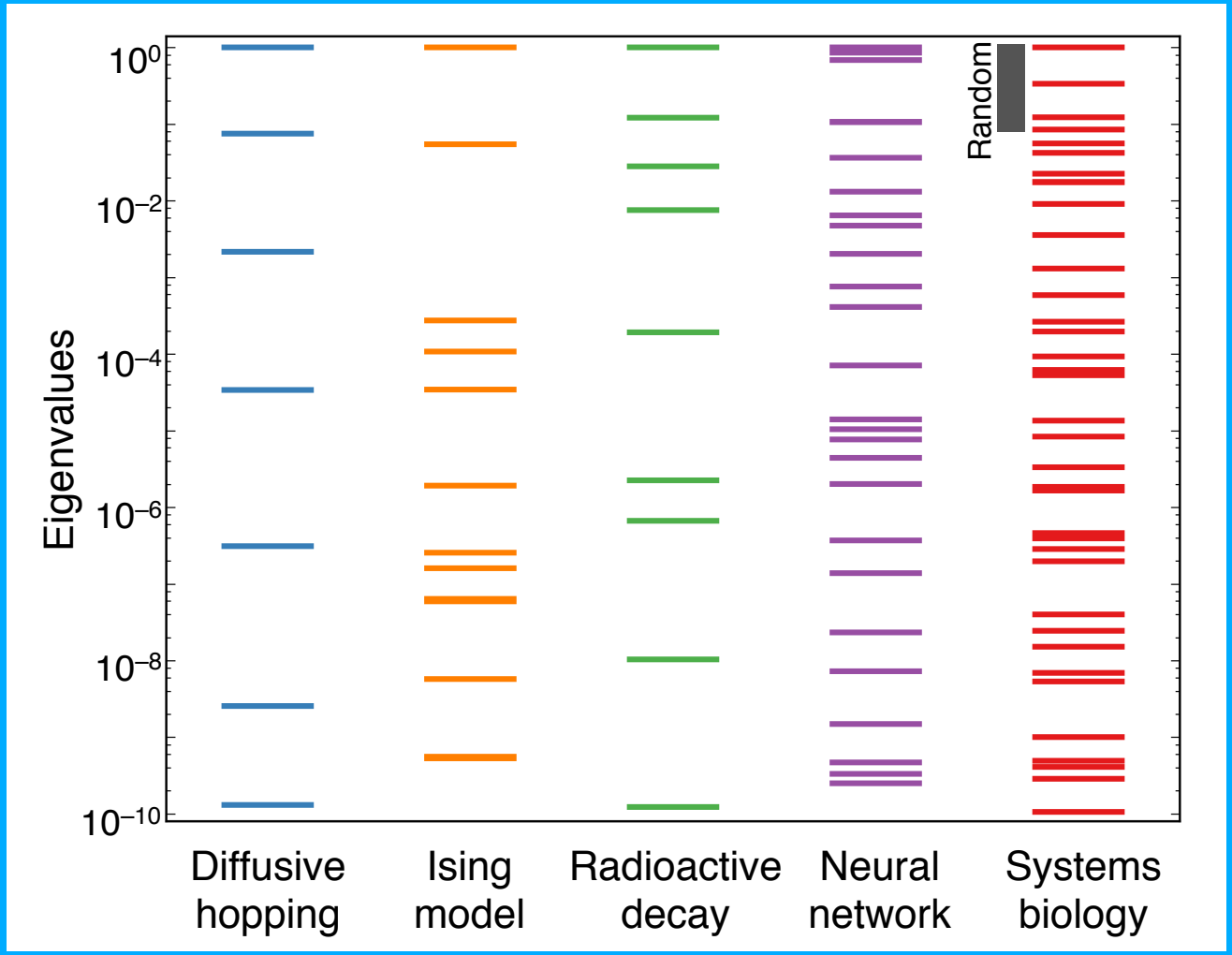
$d=N-1$

- Implies enormous range of eigenvalues
- Implies equal spacing of log eigenvalues
- Like universality for random matrices

Sloppy Models Outside Bio

Sloppy Systems

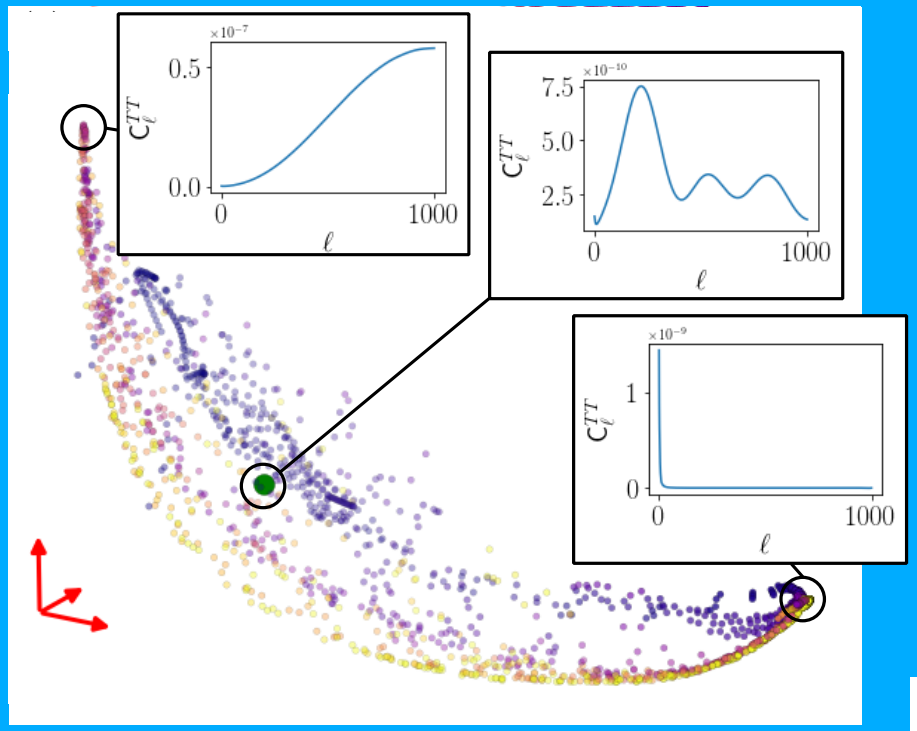
- Enormous range of eigenvalues
- Roughly equal density in log
- Observed in broad range of systems



From accelerator design to insect flight, multiparameter fits are sloppy

Intensive embeddings: Ising and CMB

Katherine Quinn



Space of all possible universes

Cosmic microwave background described by ΛCDM (cold dark matter) model

Probabilistic models

- $g_{\mu\nu}$ = Fisher Information Metric
- Isometric embedding?
- Orthogonality catastrophe — too much data
- Replica theory — extrapolate to zero data
- ‘Minkowski’ embedding in 0 D

Ising model (4x4)