SAND2022-9365C

Sandia National Laboratories

Using Manifold Learning to Enable Computationally Efficient Stochastic Inversion with High-dimensional Data

Tian Yu Yen and Tim Wildey

July 13, 2022

Virtual Talk for WCCM-APCOM 2022

Session 1707: Uncertainty Quantification for Data-Intensive Inverse Problems and Machine Learning

Acknowledgements: *This material is based upon work supported by the U.S. Department of Energy, Office of Science, ASCR, Early Career Research Program.*

Outline

1. **The Context:** Stochastic Inverse Problems (RLC Circuit Example)
2. **The Method:** Data-consistent Inversion (DCI)
3. **The Approach:** Tackling High-dimensional Data (Manifold Learning)
4. **Some Reflections:** Future Work and Analysis

The Context

Stochastic Inversion with RLC Circuits

Typical Inverse Problem

<u>Goal of a Typical Inverse Problem:</u>

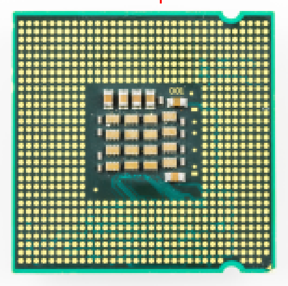
Given a model $Q(\lambda)$ and data $\{q_i\}$,

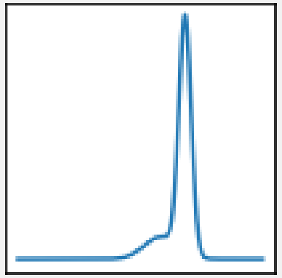
1. Find the best estimate of λ
2. Quantify uncertainty in estimate (e.g., using prob. distributions)

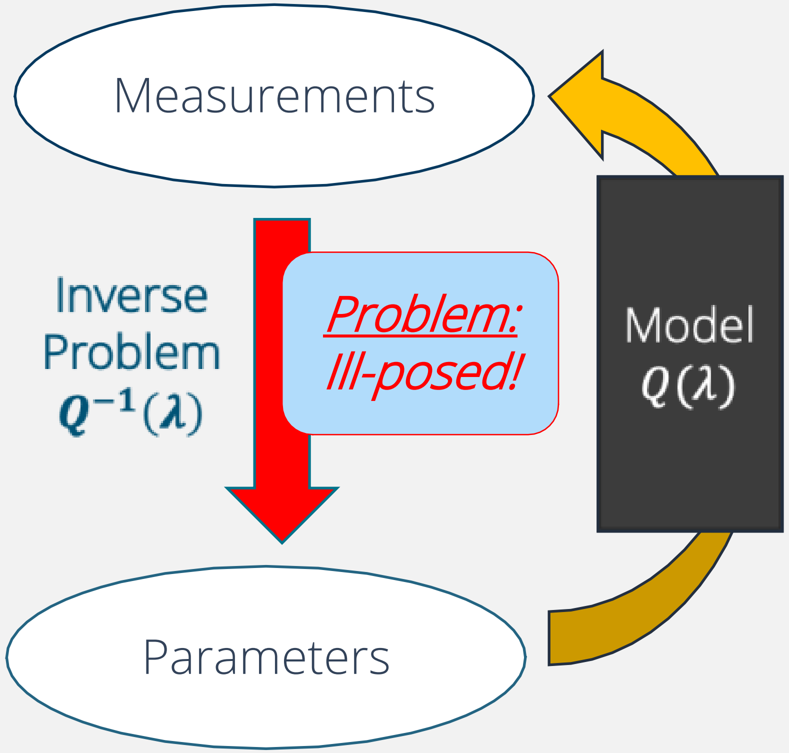
<u>Typical Solution</u>: Bayesian posterior!

What is the parameter for this component given uncertain measurements (*<u>epistemic</u>*)?

One Component







Stochastic Inverse Problem

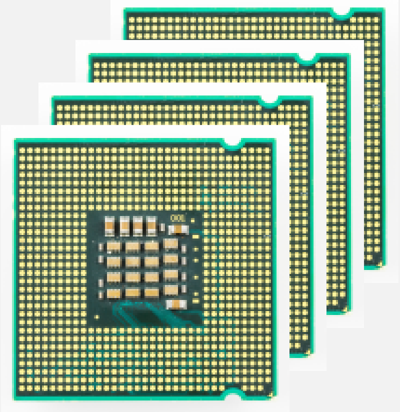
<u>Goal of a Stochastic Inverse Problem:</u>

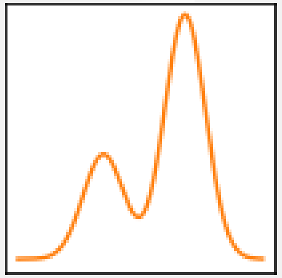
Given a model $Q(\lambda)$ and data $\{q_i\}$,

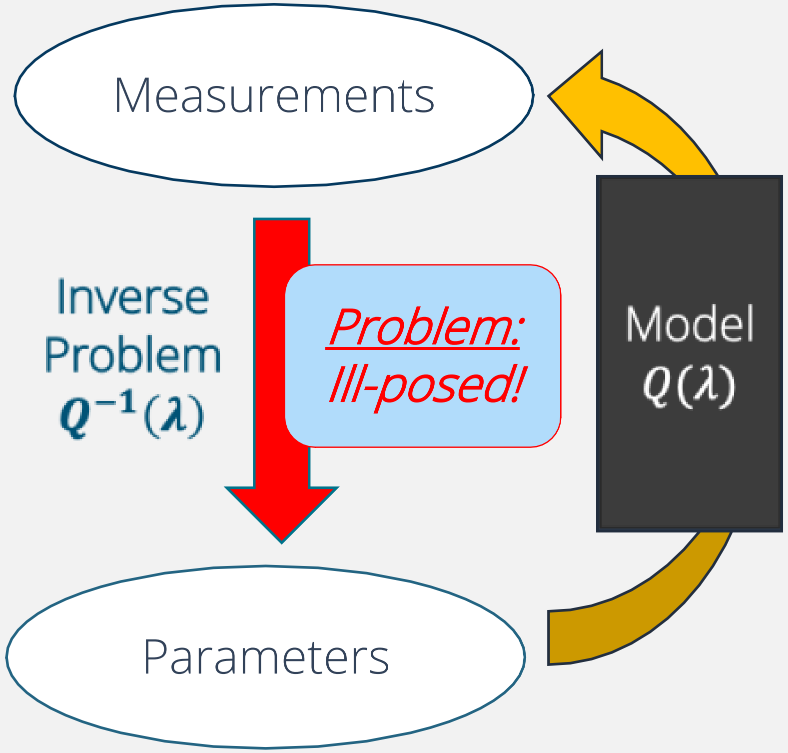
1. Characterize the variability of λ
2. Quantify uncertainty of λ using prob. distributions

What is the uncertainty in the parameter due to variability between components (*aleatoric*)?

Multiple Components







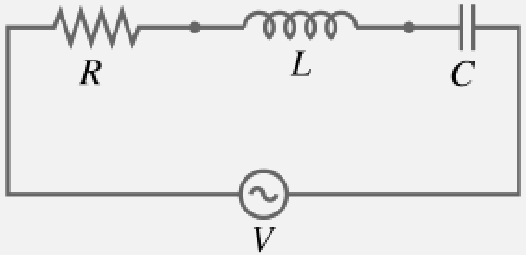
<u>Alt. Solution</u>: Data-consistent inversion!

Stochastic Inverse Problem: RLC Circuits Example

Parameters
$\lambda = (R, L, C)$

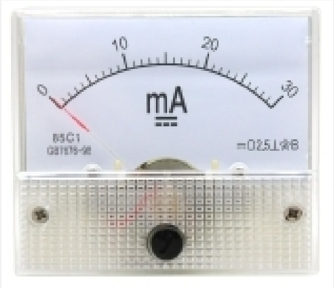
Model

$$L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{1}{C}q = f(t)$$



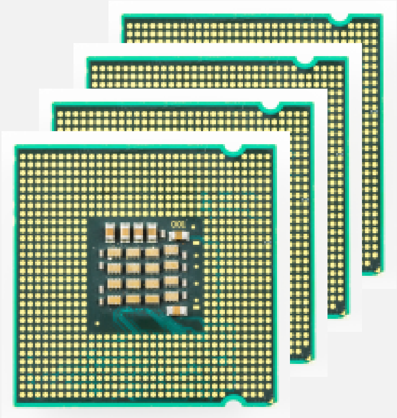
Quantity of Interest
$Q(\lambda) = (I_1(t), I_2(t))$

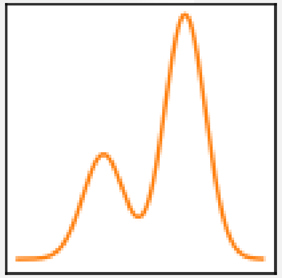
Time series of current with freq. $= (\omega_1, \omega_2)$

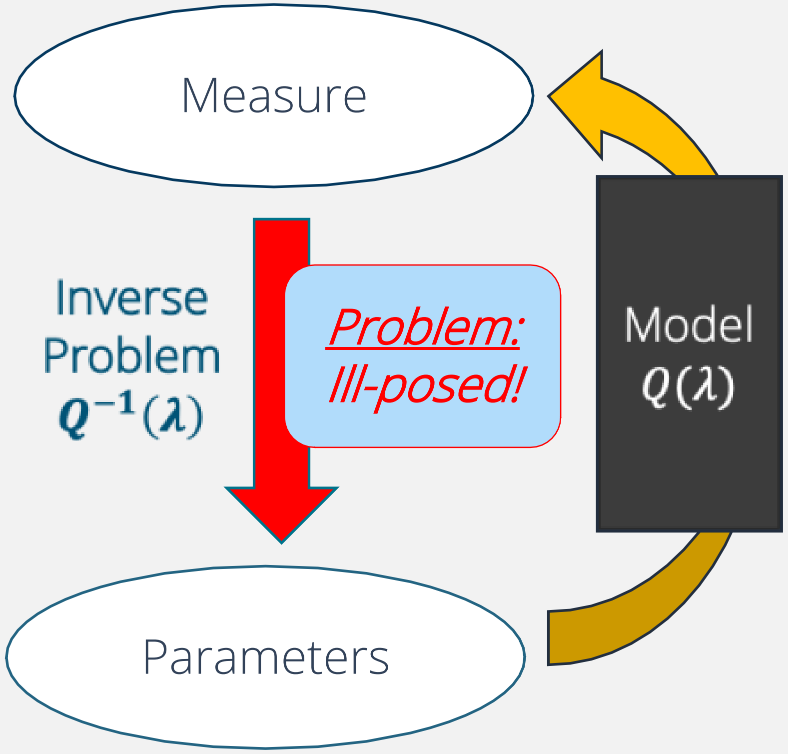


What is the uncertainty in the parameter due to variability between components (***aleatoric***)?

Multiple Components







Alt. Solution: Data-consistent inversion (DCI)!

The Method

Using Data-consistent Inversion to Solve Stochastic Inverse Problems

Data-consistent Inversion: What is the method?

A measure-theoretic approach… $E(r) = 1$

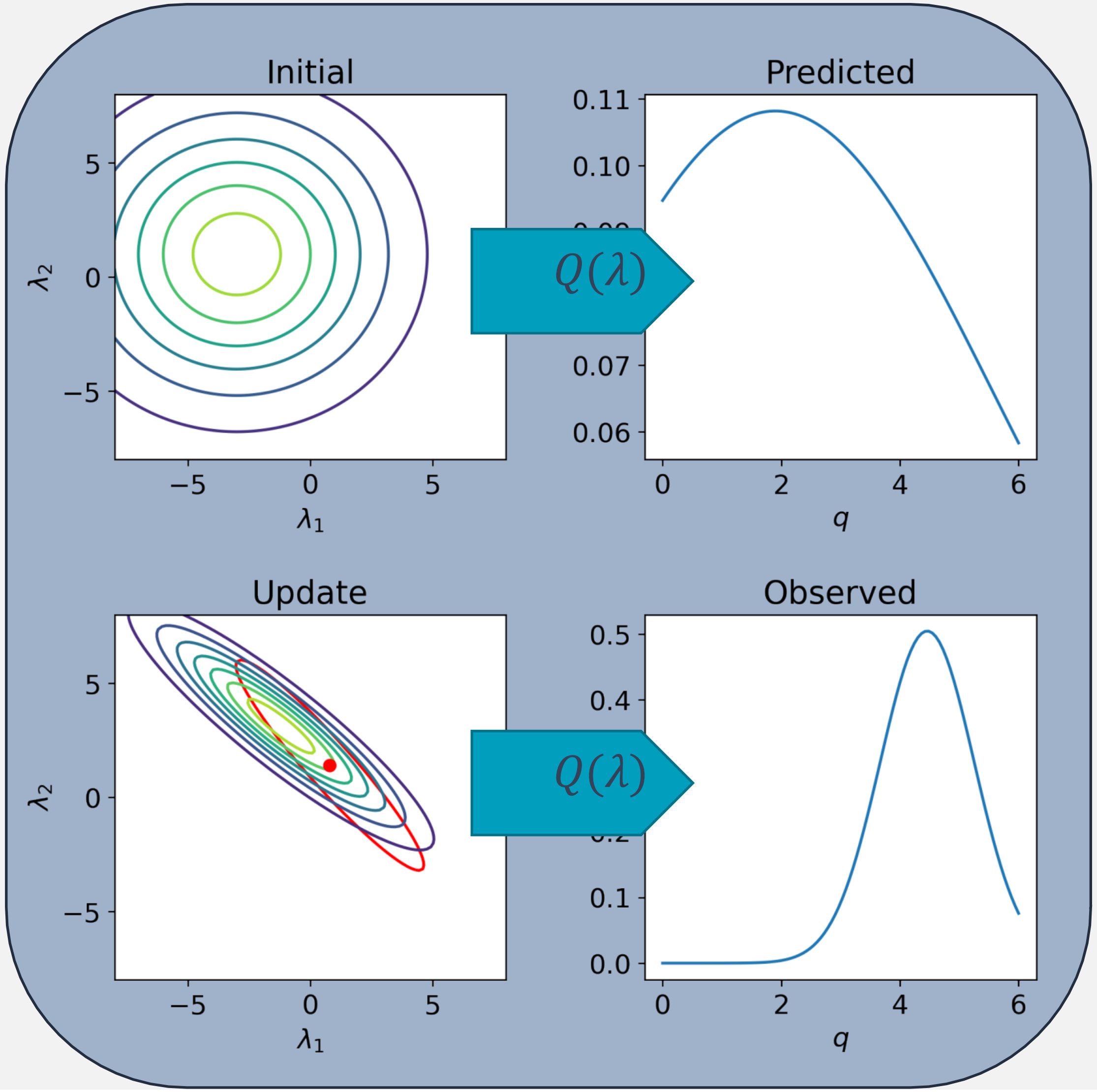
$$\pi_{update}(\lambda) = \pi_{init}(\lambda)\frac{\pi_{obs}(Q(\lambda))}{\pi_{predict}(Q(\lambda))}$$

<u>Assumption</u>: *Predictability assumption*

- Given initial assumptions about λ, model $Q(\lambda)$ can predict the data

<u>Idea of Method</u>: *Update initial assumptions by*

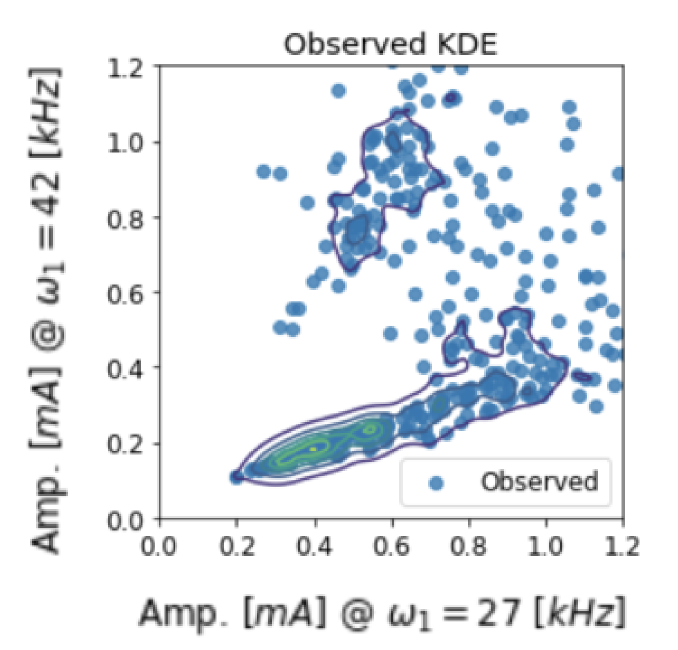
- Re-weighting initial with ratio of predicted (push-forward) to observed density

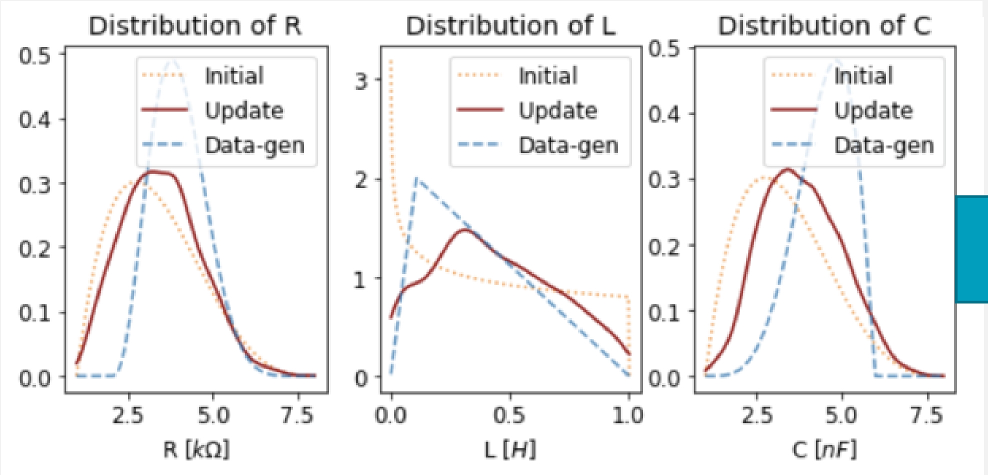


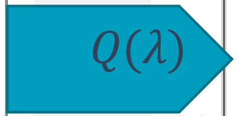
Data-consistent Inversion: A Consistent Solution

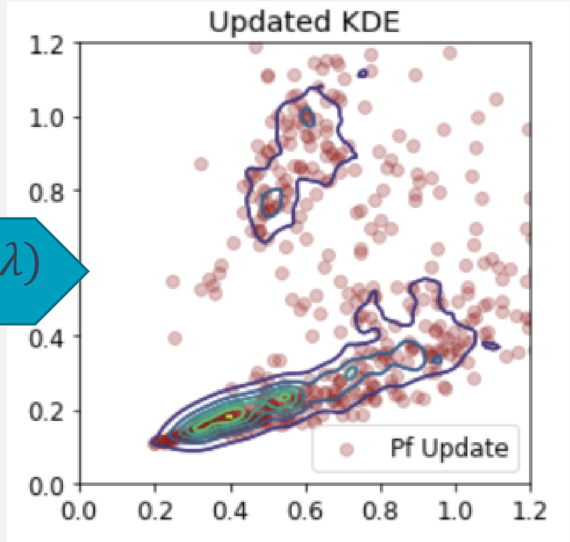
How does it work? Consider RLC Example...

$$\pi_{update}(\lambda) = \pi_{init}(\lambda)\frac{\pi_{obs}(Q(\lambda))}{\pi_{predict}(Q(\lambda))}$$









$\pi_{update}(\lambda)$ is a *consistent* solution to the stochastic inverse problem!

Data-consistent Inversion: Benefits and Drawbacks

Other benefits:

- Generally requires less model evaluations than hier. bayes
- Provides sanity check of predictability assumption ($E(r) = 1$)
- Density estimation in data space rather than parameter space

Some drawbacks:

- Density estimation in data space difficult when $\mathbf{dim}(D)$ is large

The Approach

Utilizing Nonlinear Dimension Reduction (e.g., Manifold Learning)

To Enable Density Estimation for Data-consistent Inversion

Is the data high-dimensional?

The manifold hypothesis states that the dimension of “high-dimensional” data is only superficially large…

- Data lie on a low dimensional manifold embedded in data space $\boldsymbol{D}$

In many cases of interest, a reasonable assumption!

- Multiple measurements made on physical systems likely to have structured correlation determined by physics laws…

Manifold Hypothesis: Consequences for DCI

Suppose there exists a manifold described by $z \in \mathbb{R}^m$, $m \ll \dim(D)$,

Let $f: Z \to D$ with $f(z) = q$,

$f(z)$ is <u>injective</u> (*man. hyp.*) $\Rightarrow$ $\pi_D(q) = \pi_Z\left(f^{-1}(q)\right) \cdot \det|J^T J|^{-1/2}$

$$\pi^{update}(\lambda) = \pi^{init}(\lambda) \cdot \frac{\pi_D^{obs}(Q(\lambda))}{\pi_D^{pred}(Q(\lambda))} = \pi^{init}(\lambda) \cdot \frac{\pi_Z^{obs}(f^{-1} \circ Q(\lambda))}{\pi_Z^{pred}(f^{-1} \circ Q(\lambda))}$$

Can we find a transformation of $f^{-1}: D \to Z$?

Manifold Hypothesis: Observations about f^{-1}

$$\pi^{update}(\lambda) = \pi^{init}(\lambda) \cdot \frac{\pi_Z^{obs}(f^{-1} \circ Q(\lambda))}{\pi_Z^{pred}(f^{-1} \circ Q(\lambda))}$$

Goal: find a transformation $f^{-1}: D \rightarrow Z$…

- $\dim Z \ll \dim D$
- Density estimation in Z is easier…
- Leverage *predicted* samples to learn manifold (n obs. can be small!)
- Computation of determinant-Jacobian of f^{-1} is not necessary!

Lots of Options: (*dimension reduction + density estimation*)

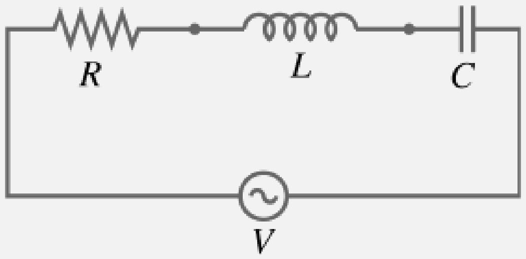
- Linear PCA + KDE
- Isomap (nonlinear) + Normalizing Flows

RLC Example: Time Series Data

Observe the current of a $n = 350$ different *simulated* RLC Circuits…

- At 2 different forcing frequencies
- 50 fixed time points

Parameters
$\lambda = (R, L, C)$

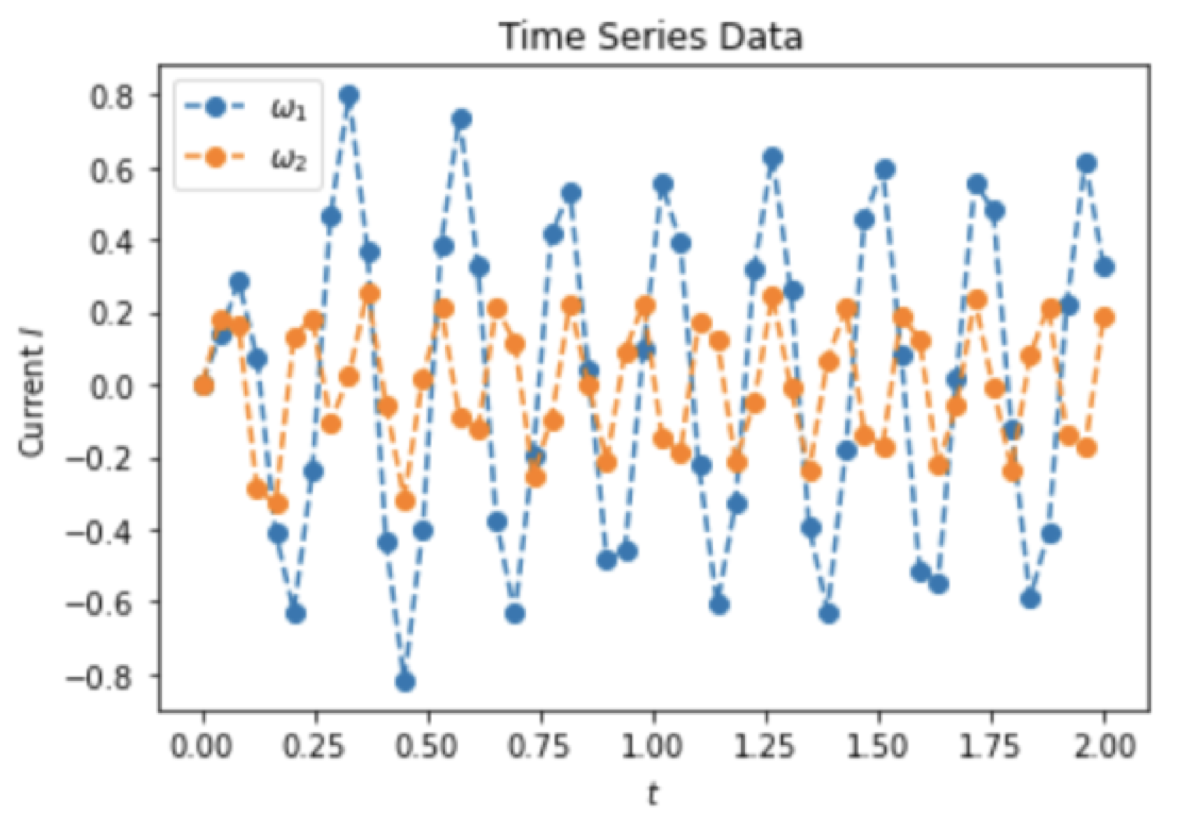


Model

$$L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{1}{C}q = f(t)$$

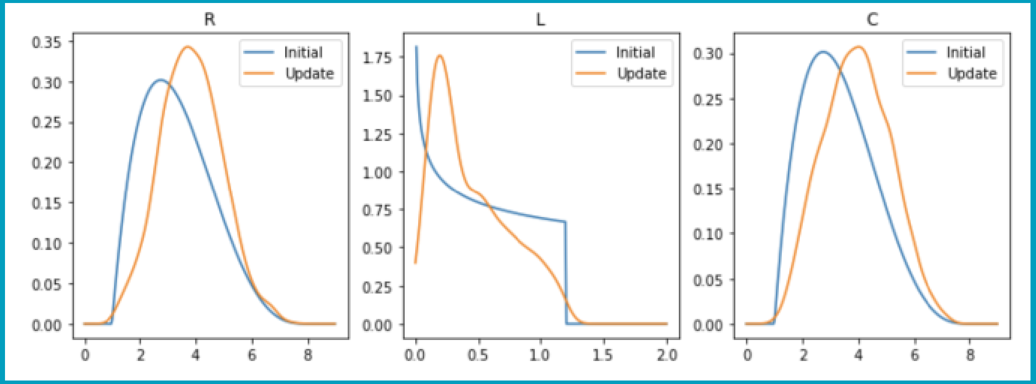
Quantity of Interest
$Q(\lambda) = (I_1(t), I_2(t))$

Time series of current with freq. $= (\omega_1, \omega_2)$



General Idea with Linear PCA + KDE

1. Sample initial distributions of RLC
2. Compute predicted time series data
3. Perform Linear PCA on predicted data
4. Transform observed data to PC-space
5. Compute KDEs on both observed and predicted PCA data
6. Apply DCI to obtain solution

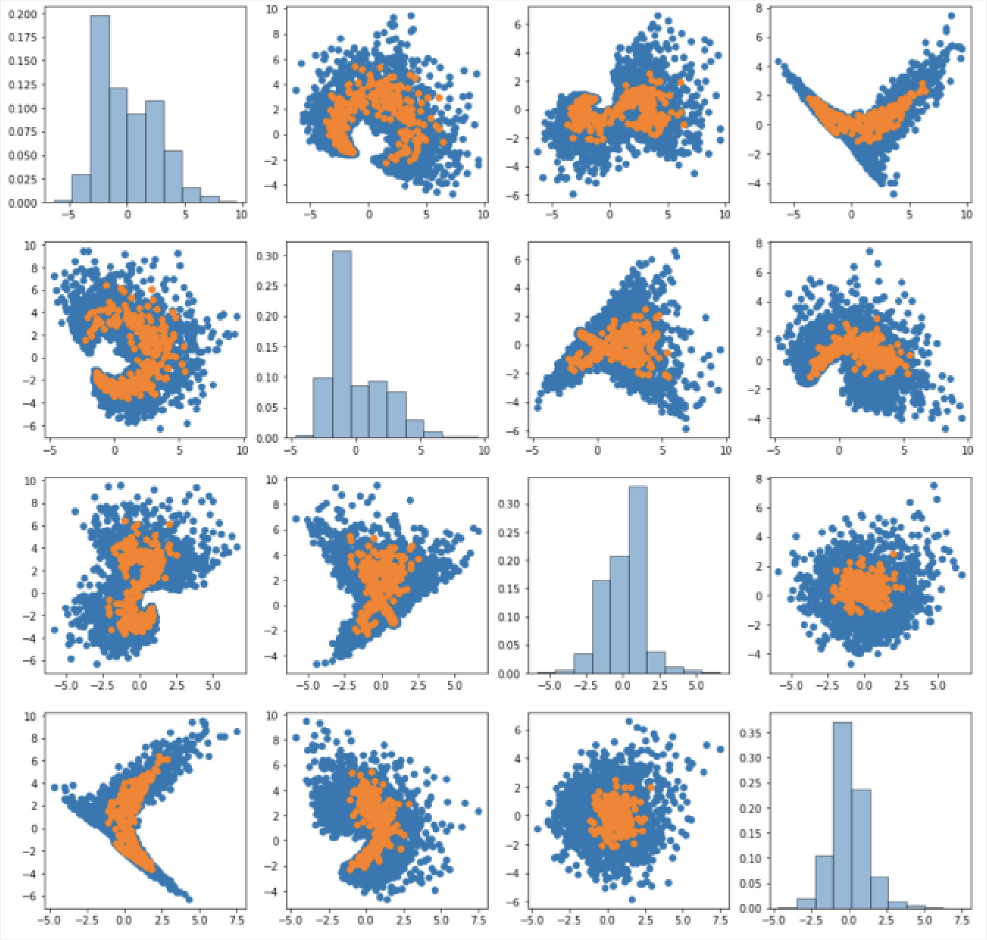


0)

Problems: KDE in Linear PCA space

Data-consistent solution looks nice but…

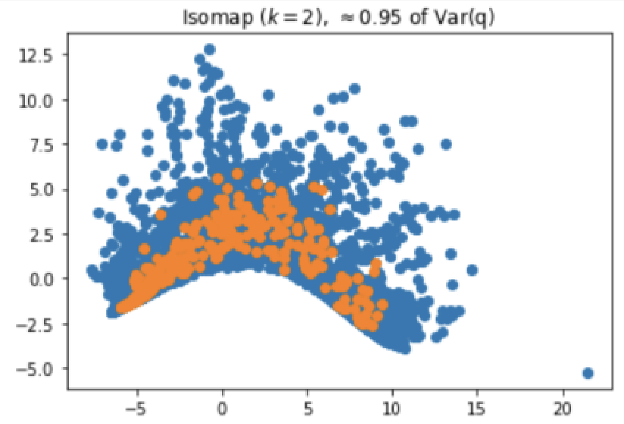
- *GKDE violates predictability assumption!*
  (though the assump. not violated by data)



Issues:

- GKDE struggles with this problem
- Choosing a bandwidth challenging

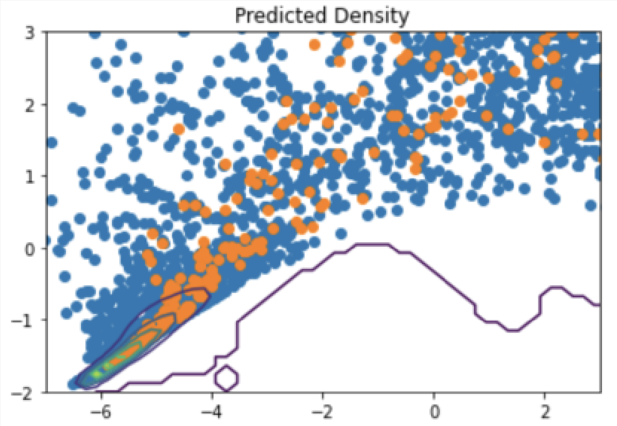
Isomap + Normalizing Flow

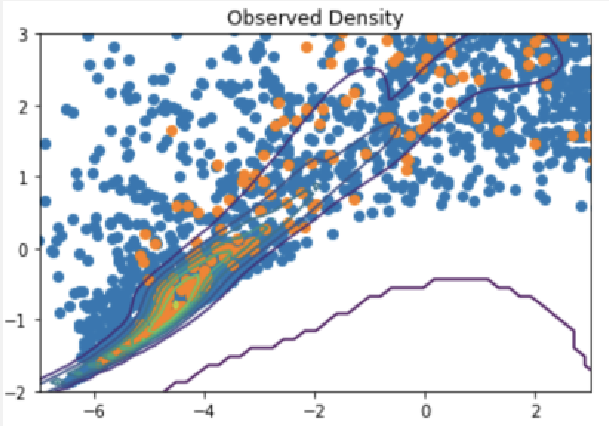


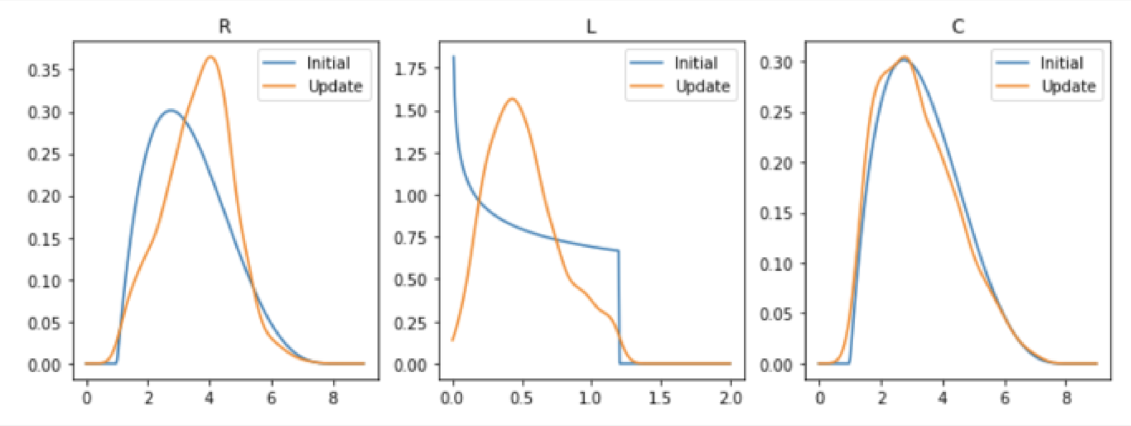
<u>Isomap:</u> nonlinear dimension reduction technique to find low-dim. embedding…

<u>Normalizing Flow:</u> neural network approach to density estimation

Similar qualitative results with only two components (95% of Var)!







Problems: Isomap + Normalizing Flows

Obtaining reliable density estimates using these ML techniques…

- Requires parameter tuning
- Dependent on the network architecture
- Stochastic optimizers stuck in local minima

Takeaway:

- In theory, any (*dimension reduction* + *density estimation*) can be used in conjunction with data-consistent inversion to find a solution
- In practice, finding a f^{-1} such that the density estimation problem is consistently tractable is difficult

Some Reflections

Future Work and Analysis

Conclusions

1. Data-consistent Inversion can *efficiently* solve stochastic inverse problems with high-dimensional data ($\mathbf{dim(D)}$ large)…
   a) *When there exists low-dimensional manifold…*
   b) *When we can find a reasonable manifold (**dimension reduction**)…*
   c) *When we can approximate the density (**density estimation**) on the manifold…*

2. Many new cutting-edge techniques for tackling *b)* and *c)*, which should we choose and when?