Ensemble Modeling

DAVID SWIGON

UNIVERSITY OF PITTSBURGH
Biomedical systems

GOAL: individualized medicine – designing patient-specific treatment

Model of the disease
- ODE
- PDE
- SDE

Calibration with patient data
- ML
- Bootstrap
- Bayesian

Risk assessment
- Cox

Treatment design
- Simulations

Adjustment
- Optimal control

Optimal control
Problems with biomedical model calibration

Model complexity
- Large number of variables and parameters, no symmetries
- Complex or unknown interactions

Data scarcity
- Unobserved variables
- Insufficient time resolution
- Small number of trajectories

Data inhomogeneity
- Variability of parameters across subjects
- Multiple subjects combined in one data set
- Data aggregation
Examples

INFLUENZA

Time courses for 16 human volunteers infected by Influenza A/Texas/91 (H1N1) [Hayden, JAMA, 96]

BACTERIAL PNEUMONIA

Data from MF1 mice subjects infected by D39 strain of pneumococcus, data are available for various other mice and pneumonia strains [Kadioglu et al., Inf. and Immun., 2000]:
INFLUENZA

Intranasal infection of 136 Balb/c mice by Influenza virus A/PR/8/34 (H1N1) [Toapanta, Ross, Morel]

20 variables, 97 parameters, ODE model
## Traditional modeling

### Dynamical model

\[
\begin{align*}
\dot{V} &= pI - cV \\
\dot{H} &= -\beta HV \\
\dot{I} &= \beta HV - \delta I
\end{align*}
\]

- \( V(0) = V_0 \)
- \( H(0) = 4 \times 10^8 \)
- \( I(0) = 0 \)

### Error minimization

\[
\min_{V_0, \beta, p, c, \delta} \sum_{k=1}^{8} \frac{(V(k) - V_k)^2}{s_k^2}
\]

### Parameter estimates

\((V_0, \beta, p, c, \delta)_{opt} = (0.0373, 0.0347, 1.33 \times 10^{-5}, 16.7, 0.806)\)

---

**Graphs:**
- Logarithmic plot of \( V(t) \) over days.
- Plots of \( V(t) \), \( H(t) \), and \( I(t) \) over days.
Questions

• Is there a parameter set/model that reproduces the data exactly?
• Is the optimal parameter set unique?
• How sensitive is the optimal parameter set to data, their uncertainty, and assumptions about error distribution?
• How well does the model fit the data? Is it even appropriate? Can the model be ruled out at the given data uncertainties?
• How large is the region in data space for which the model is appropriate?
• Since the data consist of multiple subjects, is the model appropriate and how do we know?

Answer

• Analyze more closely the relation between the system (parameters) and its solutions.
• Use probabilistic methods to describe the cohort of subjects or the parameter estimation uncertainty
Ensemble modeling

Deterministic model

Parameters
Mechanisms
Time points

Solution map
Trajectory data

Ensemble model

Distribution of
Parameters
Mechanisms
Time points

Distribution of data

Distribution describes

- Variability of parameters across cohort of subjects
- Uncertainty in determination of parameters from noisy data
ODE ensemble model

Initial value problem:
\[ \dot{x} = f(x, a), \quad x(0) = b \]

Solution:
\[ x = x(t; a, b) \]

Deterministic solution map:
\[ y = F(a) = (x(t_1, a), \ldots, x(t_n, a)) \]

Parameter distribution - random variable \( A \), density \( \rho(a) \)

Data distribution - random variable \( Y = F(A) \), density \( \eta(y) \)

Value of a variable at fixed time:
\[ X_t = x(t; A) \]
Influenza model

$\dot{V} = pI - cV$

$\dot{H} = -\beta HV$

$\dot{I} = \beta HV - \delta I$

$a = (V_0, \beta, p, c, \delta)$

$y = (V(1), ..., V(8))$

$(V_0, \beta, p, c, \delta)_{opt} = (0.0373, 0.0347, 1.33 \times 10^{-5}, 16.7, 0.806)$

$(V_0, \beta, p, c, \delta)_{base} = (0.25, 0.014, 2.7 \times 10^{-5}, 3.2, 3.2)$
Constructing ensemble models

(1) Inverse map of data sample
- Requires computation of $F^{-1}$, solution of the ODE inverse problem

\[ a_j = F^{-1}(y_j) \]

(2) Direct sampling of parameter distribution
- Requires formula for parameter density
ODE models

Initial value problem with matrix parameter $A$:

$$\dot{x} = f(x, A), \quad x(0) = b$$

**INVERSE PROBLEM:** Determine the function $f(\cdot, \cdot)$ and/or parameters $A, b$ from data about trajectory(ies) $x(\cdot; A, b)$.
Inverse problem

Mathematical analysis
- Existence, uniqueness/identifiability
- Robustness (existence/uniqueness on open sets)
- Construction of the inverse map (numerical algorithm)
- Maximal permissible uncertainty
- Prediction uncertainty
Existing results on ODE inverse problem

Observability theory
- Search for conditions that guarantee identifiability of the system state (initial conditions) from observed trajectory (can be extended to parameters)
- Kalman (1963) formulated observability conditions for linear dynamical systems
- Griffith & Kumar (1971), Kou, Elliot, & Tarn (1973) formulated sufficient conditions for observability of nonlinear dynamical systems
- Identification of analytic systems with $p$ parameters requires at most $2p + 1$ observations [Aeyels 1981; Sontag 2002]

Polynomial systems
- Reconstruction of polynomial fields from known algebraic attractors [Sverdlove 1980,81]
- Analysis of all possible topologically invariant planar quadratic dynamical systems [Perko textbook]
Existing results on inverse problem

**Algorithms for finding inverse**

- Numerical simulation combined with nonlinear least squares fitting
- Multiple shooting combined with Gauss-Newton minimization [Bock 1983; Lee]
- Collocation methods using basis function expansion, combined with NLS fitting [Varah 1982; Ramsey et al 2007]
- Contraction map and ‘collage method’ [Kunze et al. 2004]
- Parameter-free approach (using Takens map) [Sauer, Hamilton, ...]
- Compressed sensing

**Probabilistic approaches**

- Transformation of measure
- Numerical simulation combined with Bayesian inference
- Bayesian integration of ODEs [Chkrebtii et al.]
Identifiability from single complete trajectory

System: \( \dot{x} = f(x, A) \)

Initial condition: \( x(0) = b \)

Solution: \( x(t; A, b) \)
Definitions

System $\dot{x} = f(x, A)$ is

**Identifiable in** $\Omega \subseteq \mathbb{R}^{n \times m}$ *iff* for all $A, B \in \Omega$ with $A \neq B$ there exists $b \in \mathbb{R}^n$ such that $x(t; A, b) \neq x(t; B, b)$ for some $t > 0$.

**Identifiable in** $\Omega \subseteq \mathbb{R}^{n \times m}$ *from* $b$ *iff* for all $A, B \in \Omega$ with $A \neq B$ it holds that $x(t; A, b) \neq x(t; B, b)$ for some $t > 0$.

**Identifiable in** $\Omega \subseteq \mathbb{R}^{n \times m}$ *from* $x(t; A, b)$ *iff* there exists no $B \in \Omega$ with $A \neq B$ such that $x(t; A, b) = x(t; B, b)$ for all $t$.

**Unconditionally identifiable in** $\Omega \subseteq \mathbb{R}^{n \times m}$ *iff* for all $A, B \in \Omega$ with $A \neq B$ and all $b \in \mathbb{R}^n$ it holds that $x(t; A, b) \neq x(t; B, b)$ for some $t > 0$. 
Results for linear system

Theorem [SRS14]:
System $\dot{x} = Ax$ is identifiably in $\mathbb{R}^{n \times m}$ from $x(t; A, b)$ if and only if the orbit $\gamma(A, b) = \{x(t; A, b), t \in \mathbb{R}\}$ is not confined to a proper subspace of $\mathbb{R}^n$.

Proof sketch: By theorem of Astrom, identifiability from $b$ is equivalent to $\det[b|Ab| \ldots |A^{n-1}b|] \neq 0$. It suffices to show that this condition is equivalent to the orbit $\gamma(A, b)$ not being confined to a proper subspace of $\mathbb{R}^n$. 

Examples: Inverse problem solution is unique only for certain trajectories/data.

\[
\begin{bmatrix}
\dot{x} \\
\dot{y} \\
\dot{z}
\end{bmatrix} =
\begin{bmatrix}
-0.2 & -1 & 0 \\
1 & -0.2 & 0 \\
0 & 0 & -0.3
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
z
\end{bmatrix}
\]

\[
\begin{bmatrix}
\dot{x} \\
\dot{y} \\
\dot{z}
\end{bmatrix} =
\begin{bmatrix}
-1 & 1 & 0 \\
0 & -1 & 0 \\
0 & 0 & -1
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
z
\end{bmatrix}
\]

Red trajectories are confined to linear subspaces of the flux space => they do not identify the system

[SRS14]
Result for linear-in-parameter system

**Theorem [SRS14]:**
System $\dot{x} = Ag(x)$ is **identifiable in** $\mathbb{R}^{n \times m}$ from $x(t; A, b)$ if and only if the image $g(\gamma(A, b))$ of the orbit $\gamma(A, b)$ is not confined to a proper subspace of $\mathbb{R}^m$.

\[
\begin{bmatrix}
\dot{x} \\
\dot{y}
\end{bmatrix} =
\begin{bmatrix}
-2 & 1 & 0 \\
0 & 1 & -3
\end{bmatrix}
\begin{bmatrix}
x \\
x y \\
y
\end{bmatrix}
\quad b = \begin{bmatrix}
1 \\
1
\end{bmatrix}
\]

Trajectory is **confined** $\Rightarrow$ System is **not identifiable**!
System with the same trajectory:

\[
\begin{bmatrix}
\dot{x} \\
\dot{y}
\end{bmatrix} =
\begin{bmatrix}
-2 + \alpha & 1 & -\alpha \\
\beta & 1 & -3 - \beta
\end{bmatrix}
\begin{bmatrix}
x \\
x y \\
y
\end{bmatrix}
\]
Theorem [SRS14]:
Suppose that $V$ is a proper linear subspace $\mathbb{R}^n$ invariant under $A$. The following are equivalent

(i) $V$ is the minimal $A$-invariant subspace such that $b \in V$.
(ii) The orbit of $x(t; A, b)$ is not confined to a proper subspace of $V$.
(iii) There exists no $B \in \mathbb{R}^{n \times n}$ such that $B|_V \neq A|_V$ and $x(t; A, b) = x(t; B, b)$.

\[
A = \begin{bmatrix}
-1 & -3 & 2 \\
0 & -4 & 2 \\
0 & 0 & -2
\end{bmatrix}, \quad
b = \begin{bmatrix}
-1 \\
-1 \\
1
\end{bmatrix}, \quad
V = \text{span}\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \}, \quad
B = \begin{bmatrix}
1 & 0 & 0 \\
1 & 0 & 1 \\
0 & 1 & 0
\end{bmatrix}
\]
Numerical determination of confinement

Numerical errors (roundoff) can obscure the confinement of trajectory represented as a collection of data points

Use Singular Value Decomposition

\[ X = USV^T \]

- Dimension of confining subspace = \( rank(X) = \#\{ S_{jj} > 0 \} \)
- Numerical rank: \( rank_{num}(X) = \#\{ S_{jj} / \max_j S_{jj} > \varepsilon_{tol} \} \)
Identification from discrete data on a single trajectory

Discrete data: $X = (x_0, x_1, ..., x_n), \quad x_i = x(t_i; A, b)$

Identifiability from complete trajectory does not imply identifiability from discrete data.

$$A = \begin{bmatrix} -3/2 & -4 \\ 4 & -3/2 \end{bmatrix} + 2\pi r \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

$$e^A = \begin{bmatrix} 0.1458 & -0.1689 \\ 0.1689 & 0.1458 \end{bmatrix}$$
Identification from discrete data on a single trajectory

**GOAL:**

Partition the data space into domains in which

- the system is **identifiable** from data
- the system is **robustly identifiable** (on an open neighborhood)
- the system is **identifiable to within countable number** of alternatives
- the system has a **continuous family** of compatible parameters
- there is **no parameter set** corresponding to the data
- the system has a **specific dynamical behavior** (stable, saddle, ...)

Inverse map for linear systems

Discrete data (uniformly spaced): \( d = (x_0, x_1, \ldots, x_n) \)

Solution map: \( x_k = e^{Ak}b \)

**ALGORITHM**

*Step 1:* Let \( X_k = [x_k|x_{k+1}| \ldots |x_{n+k-1}], \ k = 0,1 \)

*Step 2:* Compute \( \Phi = X_1X_0^{-1} \)

*Step 3:* Solve \( e^A = \Phi \)

**Existence and uniqueness** of \( A \) derives from conditions for existence and uniqueness of real matrix logarithm [Culver 1966].
Robust existence & uniqueness
Robust existence & uniqueness

**Theorem [SRS17]:** There exists an open set $U \subset \mathbb{R}^{n \times n}$ containing $\Phi$ such that for any $\Psi \in U$ the equation $e^A = \Psi$ **has a solution** $A \in \mathbb{R}^{n \times n}$ iff

(a) \( \{x_0, x_1, \ldots, x_{n-1}\} \) are linearly independent and
(b) $\Phi$ has **only positive real or complex eigenvalues**.

**Theorem [SRS17]:** There exists an open set $U \subset \mathbb{R}^{n \times n}$ containing $\Phi$ such that for any $\Psi \in U$ the equation $e^A = \Psi$ **has a unique solution** $A \in \mathbb{R}^{n \times n}$ iff

(a) \( \{x_0, x_1, \ldots, x_{n-1}\} \) are linearly independent and
(b) $\Phi$ has $n$ **distinct positive real eigenvalues**.

**Theorem [SRS17]:** There exists an open set $U \subset \mathbb{R}^{n \times n}$ containing $\Phi$ such that for any $\Psi \in U$ the equation $e^A = \Psi$ **does not have a solution** $A \in \mathbb{R}^{n \times n}$ iff

(a) \( \{x_0, x_1, \ldots, x_{n-1}\} \) are linearly independent and
(b) $\Phi$ has **at least one negative real eigenvalue of odd multiplicity**.
Region of unique inverses

Let

\[ \hat{\Phi} = X_0^{-1} \Phi X_0 = X_0^{-1}X_1 \]

Eigenvalues of \( \Phi \) are determined by \( y = X_0^{-1}x_n \)

Conditions on \( y \) that guarantee positive real eigenvalues have been worked out for arbitrary dimension [Gantmacher; Yang]

Example in 2D (regions of data which give unique inverse)
Regions of stability for 2x2 systems

\[
\begin{align*}
\det \Phi &= \det, \\
\text{tr} D_T &= \text{tr}, \\
1 - T + D &= 0,
\end{align*}
\]

\[D = 0, \quad D = 1\]

- DNE
- Stable Node
- Stable Spiral
- Unstable Spiral
- Unstable Node
- Saddle

\[T = \text{tr } \Phi, \quad D = \det \Phi\]
Numerical error of identification

**Exact formula**

\[ \Phi = X_1 X_0^{-1} \]

**Numerical sensitivity** of the inverse to errors in the data.

\[
\frac{\| \delta \Phi \|}{\| \Phi \|} \leq \kappa(X_0) \left( \frac{\| \delta X_0 \|}{\| X_0 \|} + \frac{\| \delta X_1 \|}{\| X_1 \|} \right)
\]

- Here \( \| \cdot \| \) is any matrix norm and \( \kappa(C) = \| C \| \| C^{-1} \| \) is the condition number of the matrix \( C \).
- Note that collinearity of vectors in \( X_0 \) leads to a loss of accuracy.
Identification from non-uniformly spaced data

Discrete data (non-uniformly spaced)

\[ d = \{ x_{j_0}, x_{j_1}, \ldots, x_{j_n} \} \quad x_{j_k} = x(j_k; A, b) = e^{A j_k} b \quad x_{k+1} = e^A x_k \]

**Lemma:** Let \( x_{j_k} = \Phi^{j_k} b \) where \( j_k \) are integers such that \( 0 = j_0 < j_1 < \cdots < j_n \). Let \( X_0 = [x_{j_0}, x_{j_1}, \ldots, x_{j_n-1}] \) be invertible. Let \( y = X_0^{-1} x_{j_n} \) be a vector with entries \( y_1, y_2, \ldots, y_n \). If \( \Phi \) is diagonalizable, then \( \lambda \) is an eigenvalue of \( \Phi \) only if it is a root of the polynomial

\[ p_y(\lambda) = \lambda^n - y_n \lambda^{n-1} - \cdots - y_2 \lambda^1 - y_1 \]
Identification from non-uniformly spaced data

**ALGORITHM**

*Step 1:* Get $y = X_0^{-1}x_j$

*Step 2:* Find all roots of $p_y(\lambda)$

*Step 3:* Choose a combination of $n$ distinct roots and form $\Phi$.

*Step 4:* Compute vectors $z_j, z_{j_1}, ..., z_{j_{n-1}}$ as $z_0 = e_1$ and $z_{j_{k+1}} = \Phi^{j_{k+1} - j_k}z_{j_k}$.

*Step 5:* Compute $\Phi = P^{-1}\hat{\Phi}P$ where $P$ is such that $Pz_{j_k} = z_{j_k}$ for $k = 0, 1, ..., n - 1$.

*Step 6:* Solve $e^A = \Phi$.

**Note:** The variety of choices for $\hat{\Phi}$ may lead to non-uniqueness of parameter estimates.
Uncertainty analysis

Uncertain data: \( \tilde{d} = (\tilde{x}_0, \tilde{x}_1, \ldots, \tilde{x}_n) \)

Bounds on uncertainty: \( C(d, \varepsilon) = \{ \tilde{d} : \max_{i,j} |\Delta x_{i,j}| < \varepsilon \} \)

Neighborhood \( C(d, \varepsilon) \) is called permissible for property P iff P is shared by all systems corresponding to the data \( \tilde{d} \in C(d, \varepsilon) \).

Value \( \varepsilon_P > 0 \) is maximal permissible uncertainty for property P iff \( C(d, \varepsilon) \) is permissible for P for all \( 0 < \varepsilon < \varepsilon_P \), and \( C(d, \tilde{\varepsilon}) \) is not permissible for P for any \( \tilde{\varepsilon} > \varepsilon_P \).

P can be
- Existence of inverse (unique or nonunique)
- Unique inverse
- Unique stable inverse
Results for linear systems

**Theorem (lower bound):** Let $d$ be such that $\Phi$ has $n$ distinct positive eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Let $m_1 = \frac{1}{2} \min_{i < j} |\lambda_i - \lambda_j|$, $m_1 = \min_i \lambda_i$, and $\delta_U = \min\{m_1, m_2\}$. If $0 < \varepsilon \leq \varepsilon_U$ where

$$
\varepsilon_U = \frac{\delta_U}{n(\delta_u + 1 + \|\Lambda\|)\|S^{-1}\|\|X_0^{-1}S\|}
$$

with $\Phi = S\Lambda S^{-1}$, then for any $\tilde{d} \in C(d, \varepsilon)$, $\tilde{\Phi}$ has $n$ distinct positive eigenvalues.
Theorem (upper bound): Let $d$ be such that $\Phi$ has $n$ distinct positive eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, and let $y$ be the last column vector of the companion matrix form $\tilde{\Phi}$ of $\Phi$. The smallest neighborhood $C(d, \varepsilon)$ that contains data $\tilde{\alpha}$ for which companion matrix $\tilde{\Phi}$ has last column $\tilde{y}$ has $\varepsilon = \bar{\varepsilon}_U$ where

$$\bar{\varepsilon}_U = \frac{\|X_0(y - \tilde{y})\|_\infty}{\|\tilde{y}\|_1 + 1}$$
**Permissible data uncertainty:** Amount of measurement error that can be tolerated without altering inference qualitatively

\[ \dot{x} = Ax, \quad x(0) = b \]

Upper and lower bounds for linear systems.

<table>
<thead>
<tr>
<th>Property (X)</th>
<th>SN</th>
<th>U</th>
<th>S</th>
<th>DNE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x_2)</td>
<td>(3.68, -3.46)</td>
<td>(7, -10)</td>
<td>(3.1, -5.5)</td>
<td>(3.6)</td>
</tr>
<tr>
<td>(\hat{c}_x) Numerical estimate</td>
<td>0.216</td>
<td>1.075</td>
<td>0.519</td>
<td>3.055</td>
</tr>
<tr>
<td>(c_x)</td>
<td>0.059</td>
<td>0.149</td>
<td>0.207</td>
<td>0.609</td>
</tr>
<tr>
<td>(\bar{c}_x)</td>
<td>0.217*</td>
<td>1.078†</td>
<td>0.519‡</td>
<td>3.059#</td>
</tr>
</tbody>
</table>

* \(\lambda_1 = \lambda_2, \quad \dagger \lambda_1 = 0, \quad \ddagger \lambda_2 = 1, \quad \# \lambda_2 \to \infty \)

\[ [SRS17] \]
**Prediction Uncertainty:** In high dimensions inverse problem solution is very sensitive to small changes in data.

- **Stable node**: $\|\Delta x_n\| = 0.0049$
- **Saddle**: $\|\Delta x_n\| = 0.0295$
- **Stable Spiral**: $\|\Delta x_n\| = 0.0005$
- **No Real Inverse**: $\|\Delta x_n\| = 0.0005$

[SRS17]
SUMMARY

Existence and identifiability analysis provides regions on which the inverse problem for ODEs is well posed

Additional solution attributes (e.g., stability, spirality) can be included

OPEN PROBLEMS

Existence of inverse for linear-in-parameter systems

Maximal permissible uncertainty for linear-in-parameter systems
ODE Models with random effects

Initial value problem: \[ \dot{x} = f(x, a), \quad x(0) = b \]

Solution: \[ x(t; a, b) \in C^1(\mathbb{R}^n) \]

Data vector: \[ y = H(x(t_0), x(t_1), \ldots, x(t_n)) \]

Solution map: \[ y = F(a) \quad \text{invertible for } y \in C(\bar{y}, \varepsilon) \]

Random parameter model (RPM):
\[ Y = F(A) \quad A \text{ is a r. v. with density } \rho(a) \]

Random measurement error model (REM):
\[ Y = F(a) + G \quad G \text{ is a r. v. with density } \gamma(g) \]
Random-parameter model

\[ Y = F(A) \]

**Inverse problem:** Find parameter density \( \rho(a) \) from the knowledge of data density \( \eta(y) \) and \( F \).

Change of variables formula: for any set \( \Gamma \)

\[
\int_{\Gamma} \rho(a) da = \int_{F(\Gamma)} \eta(y) dy = \int_{\Gamma} \eta(F(a)) J(a) da
\]

\[ J(a) = |\det D_a F(a)| \]

Parameter density:

\[ \rho(a) = \eta(F(a)) J(a) \]

**Distribution of parameters:** represents subject variability
Computational aspects

Jacobian is expensive to compute.

**Methods for computing \( J(a) \)**

- **Exact formula for linear models**
  
  For \( 2 \times 2 \) diagonalizable matrix \( A = WMW^{-1} \):
  
  \[
  J(a) = \frac{(b_2w_{11} - b_1w_{21})^2(b_2w_{12} - b_2w_{22})^2e^{\mu_1+\mu_2}(e^{\mu_1} - e^{\mu_2})^4}{(\det W)^2(\mu_1 - \mu_2)^2}
  \]

- **Numerical differentiation**
  
  \[
  \frac{\partial F}{\partial a_j} (\hat{a}) \approx \frac{1}{\varepsilon} (F(\hat{a} + \varepsilon e_j) - F(\hat{a}))
  \]

  Requires \( n \) additional integrations of the system in each step

[SSZR19]
Computational aspects

- Numerical integration of the first variational equations (sensitivity coefficients) for ODE models

Let $s_{a_j}(t) = \frac{\partial x(t;a,b)}{\partial a_j}$, $s_{b_j}(t) = \frac{\partial x(t;a,b)}{\partial b_j}$

\[
\begin{align*}
\dot{x} &= f(x, a) & x(0) &= b \\
\dot{s}_{a_j}(t) &= \frac{\partial f(z,a)}{\partial z} \bigg|_{z=x(t)} s_{a_j}(t) + \frac{\partial f(z,a)}{\partial a_j} \bigg|_{z=x(t)} & s_{a_j}(0) &= 0 \\
\dot{s}_{b_j}(t) &= \frac{\partial f(z,a)}{\partial z} \bigg|_{z=x(t)} s_{b_j}(t) & s_{b_j}(0) &= e_j
\end{align*}
\]

Dimension of the ODE system increases from $n$ to $n(p + n + 1)$
Computational aspects

- Broyden’s method (rank-1 update)

\[
D_a F(\hat{a}) \approx D_a F(a^k) + \frac{F(\hat{a}) - F(a^k) - D_a F(a^k)\Delta a^k}{(\Delta a^k)^T \Delta a^k} (\Delta a^k)^T
\]

Only one computation of \(F(\hat{a})\) (integration of the system) needed at every step

Update every 10 computations keeps relative error of \(D_a F(\hat{a})\) below 4%

Accuracy of the posterior is lower than the accuracy of Jacobian

[SSZR19]
Examples

Common problem:
- Jacobian calculation is expensive, so can we replace $J(a)$ by a simpler distribution $\pi(a)$?
Linear dynamical system, uniform data distribution $U(\bar{x}_{ij} - \varepsilon, \bar{x}_{ij} + \varepsilon)$

\[
\dot{x} = \Lambda x \quad d = \begin{bmatrix} 10 & 6.2530 \\ 2 & -1.5624 \end{bmatrix} \begin{bmatrix} 5 \\ -2.3 \end{bmatrix}
\]

\[\varepsilon = 0.436\]

\[\Lambda = \begin{bmatrix} -0.5024 & -0.7491 \\ -0.4839 & -0.9519 \end{bmatrix} \quad a = (\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22})\]

$A$ is a sample of exact inverse density, $M_\pi$ is a sample of $\rho(a) = \eta(F(a))\pi(a)$, where $U(a) = 1, R(a) = \prod_i |a_i|^{-1}$
Linear dynamical system, uniform data distribution $U(\bar{x}_{ij} - \varepsilon, \bar{x}_{ij} + \varepsilon)$

Broyden’s method (rank-1 update)

$$D_a F(\hat{a}) \approx D_a F(a^k) + \frac{F(\hat{a}) - F(a^k) - D_a F(a^k)\Delta a^k}{(\Delta a^k)^T \Delta a^k} (\Delta a^k)^T$$

- $D_a F(\hat{a})$ recomputed exactly every N steps (101, 11, 2)
Jacobian alternatives sometimes work well

\[ \dot{x} = \Lambda x \]

\[ d = \begin{bmatrix} 10 & 5 & 4 \\ 2 & -2.3 & -2.2 \end{bmatrix} \]

\[ \varepsilon_U = 0.134 \]

\[ \Lambda = \begin{bmatrix} -1 & -1.5 \\ -1 & -2 \end{bmatrix} \]

Reciprocal works well

Uniform works well

\[ \dot{x} = \Lambda x \]

\[ d = \begin{bmatrix} 5 & 10 & 30 \\ 15 & 10 & -15 \end{bmatrix} \]

\[ \varepsilon_U = 1.272 \]

\[ \Lambda = \begin{bmatrix} 1.1347 & -0.2187 \\ -1.4213 & 0.3694 \end{bmatrix} \]

[SSZR19]
Nonlinear influenza dynamics, uniform data distribution \( U((1-\sigma)x_{ij}, (1+\sigma)x_{ij}) \)

\[
\begin{align*}
\dot{V} &= rI - cV \\
\dot{H} &= \beta HV \\
\dot{I} &= \beta HV - \delta I
\end{align*}
\]

\( \sigma = 0.1 \)

\( t = \{2, 3\} \)

\( a = (V_0, H_0, \beta, r, c, \delta) = (0.093, 4 \times 10^8, 2.7 \times 10^{-5}, 0.012, 3.0, 4.0) \)

Narrow data distribution, all \( \pi(a) \) work well

\( A \) is a sample of exact inverse density, \( M_\pi \) is a sample of \( \rho(a) = \eta(F(a))\pi(a) \), where \( U(a) = 1, R(a) = \prod_i |a_i|^{-1} \)
Examples

Common problem:
- Can $\eta(y)$ be approximated by aggregate Gaussian density $\tilde{\eta}(y)$?
Linear dynamical system, Gaussian parameter distribution, aggregate Gaussian data distribution.

\[
\dot{x} = \Lambda x
\]

\[
\Lambda = \begin{bmatrix}
-1 & -1.5 \\
-1 & -2
\end{bmatrix}
\]

Only Jacobian produces localized distribution

\[
A \text{ is the exact source density, } M_\pi \text{ is a sample of } \rho(a) = \tilde{\eta}(F(a))\pi(a)
\]

[SSZR19]
Nonlinear influenza dynamics, Gaussian parameter distribution, aggregate Gaussian data distribution

\[
\begin{align*}
\dot{V} &= rI - cV \\
\dot{H} &= -\beta HV \\
\dot{I} &= \beta HV - \delta I
\end{align*}
\]

\[t = \{1, 2\}\]

\[
\bar{a} = (V_0, H_0, \beta, r, c, \delta) = (0.093, 4 \times 10^8, 2.7 \times 10^{-5}, 0.012, 3.0, 4.0)
\]

[SSZR19]
Domain deformation
Summary

• Accurate parameter inference of ensemble models requires determination of the domain of invertibility of the solution map and the computation of Jacobian

• Further work is required to design efficient computational procedures for Jacobian calculations
Random measurement error model

\[ Y = F(a) + G \]

**Inverse problem:** Find \( a \) from the knowledge of data \( D = (y_1, \ldots, y_N) \), \( F(a) \), and the error density \( \gamma(g) \).

Likelihood

\[ L(a|y) = \rho(y|a) = \gamma(g) = \gamma(y - F(a)) \]

Bayesian posterior

\[ \rho(a|y) \propto L(a|y)\pi(a) \]

- Prior \( \pi(a) \) contains all information about parameters known before data are taken into account

Distribution of parameters: represents uncertainty in parameter inference
Priors

Uniform prior: \( \pi(a) = 1 \)

Reciprocal prior: \( \pi(a) = \prod_i |a_i|^{-1} \)

Jeffreys invariant prior, based on the Fisher information matrix:
\[
\pi(a) = \sqrt{\det I(a)}
\]

\[
I(a)_{ij} = \int \left( \frac{\partial}{\partial a_i} \ln L(a|y) \right) \left( \frac{\partial}{\partial a_j} \ln L(a|y) \right) L(a|y)
\]

Jacobian prior: For model with likelihood \( L(a|y) = \gamma(y - F(a)) \),
\[
\pi(a) = J(a) \sqrt{\det W} \quad J(a) = | \det D_a F(a) |
\]
Relation between Bayesian posterior for REM and parameter density for RPM

Let $\bar{\eta}(y) = \bar{y}(y - \bar{y})$ be the aggregate density obtained from data sample $Y = \{y^1, y^2, \ldots, y^N\}$ and let $\bar{\rho}(a) = \bar{\eta}(F(a))J(a)$

Let $\bar{\sigma}(a|\bar{y})$ be the Bayesian posterior with likelihood $\bar{L}(a|\bar{y}) = \bar{y}(\bar{y} - F(a))$ and prior $\pi(a)$

**Theorem:** $\bar{\rho}(a) = \bar{\sigma}(a|\bar{y})$ whenever $\bar{y}(y)$ is symmetric and $\pi(a) = J(a)$

---

$M_\pi$ is a sample of Bayesian posterior $\sigma(a|\bar{y}) = \gamma(\bar{y} - F(a))\pi(a)$, where $U(a) = 1$, $R(a) = \prod_i |a_i|^{-1}$
Jacobian prior revisited

\[ Y = F(a) + G \]

Likelihood:
\[ L(a|y) = \gamma(y - F(a)) \]

Posterior:
\[ \sigma(a|y) = \frac{L(a|y)\pi(a)}{q(y)} \]

With Jacobian prior on parameters, the prior on data \( y \) is uniform

\[
q(y) = \int L(a|y)\pi(a)da \\
= \int \gamma(y - F(a))J(a)da = \int \gamma(y - z)dz = 1
\]

The Jacobian prior is noninformative in the original sense proposed by Bayes – it does not give preference to any observed data [Stigler, 1982].
Commonly used interpretation of noninformative prior

- All parameters are equally likely
- Observations can be anticipated
Bayes’ understanding of noninformative prior

- Any data vector is equally likely
- Observations cannot be anticipated
Relation to least squares

Objective function minimization

\[ a^* = \arg \min_a \sum_j \left( x_j - x(t_j; a) \right)^2 \]

Maximum likelihood estimate

\[ a_{MLE} = \arg \max_a L(a|y) = \arg \max_a \gamma(y - F(a)) = a^* \]

The data \( y \) is the most likely for model with the parameter \( a_{MLE} \)

Maximum a posteriori estimate

\[ a_{MAP} = \arg \max_a \rho(a|y) = \arg \max_a \gamma(y - F(a))J(a) \]

The model with parameter \( a_{MAP} \) is the most likely given observed data \( y \)
Applications to immunology

Data on time courses of 16 volunteers infected by Influenza A/Texas/91 (H1N1) [Hayden, JAMA, 96]

\[
\begin{align*}
\dot{V} &= pI - cV \\
\dot{H} &= -\beta HV \\
\dot{i} &= \beta HV - \delta I
\end{align*}
\]

Classical parameter estimates

\[
(V_0, \beta, p, c, \delta)_{opt} = (0.0373, 0.0347, 1.33\times10^{-5}, 16.7, 0.806)
\]
Ensemble parameter distribution

\[(V_0, \beta, p, c, \delta)_{base} = (0.25, 0.014, 2.7 \times 10^{-5}, 3.2, 3.2)\]

Ensemble trajectory prediction
Parameter correlations – parameter distribution is organized in clusters
Optimal trajectories reflect bimodality of the posterior

\[
\begin{align*}
\dot{V} &= pI - cV \\
\dot{H} &= -\beta HV \\
\dot{I} &= \beta HV - \delta I
\end{align*}
\]
Uncertainty quantification of antiviral treatment

The effect of neuraminidase inhibitor GG167 is simulated by decreasing $p$ to 1/20th of its value.

Treatment starting at 26 hrs

Treatment starting at 50 hrs

Data from [Hayden, JAMA, 96]
Viral phenotype characterization

PB1-F2 protein occurring in 1918 pandemic H1N1 strain increases apoptosis in monocytes, alters viral polymerase activity in vitro, enhances inflammation and increases secondary pneumonia in vivo.

Mice infected intranasally with 100 TCID50 of influenza A virus PR8 (squares) or PR8-PB1-F2(1918) (triangles).

[Smith et al., PLoS ONE, 2011]
Comparison of model predictions and parameter distributions obtained using ensemble modeling.

**PR8**

**PR8-PB1-F2**

![Graphs showing model predictions and parameter distributions for PR8 and PR8-PB1-F2.](image-url)
Bacterial pneumonia

Data from MF1 mice subjects infected by D39 pneumococcus strain [Kadioglu et al., Inf. and Immun., 2000]:

\[
\frac{dP_L}{dt} = k_1P_L - \frac{v_P}{\mu + P_L} \left( \frac{\xi N_P}{1 + \xi P_L} \right) + f [(bD + a)P_B - bDP_L]
\]

\[
\frac{dP_B}{dt} = k_2P_B \left( \frac{P_B - \xi}{P_B + \xi} \right) \left( 1 - \frac{P_B}{K} \right) - \frac{\xi N_P}{1 + \xi P_B} - (bD + a)P_B + bDP_L.
\]

\[
\frac{dD}{dt} = qP_L (1 - D) - cD
\]

\[
\frac{dN}{dt} = \left( -N + \frac{hP_L}{n + P_L} \right) \left( \frac{1}{\tau_N} \right)
\]

Data from MF1 mice subjects infected by D39 pneumococcus strain [Kadioglu et al., Inf. and Immun., 2000]:

\[
\frac{dP_L}{dt} = k_1P_L - \frac{v_P}{\mu + P_L} \left( \frac{\xi N_P}{1 + \xi P_L} \right) + f [(bD + a)P_B - bDP_L]
\]

\[
\frac{dP_B}{dt} = k_2P_B \left( \frac{P_B - \xi}{P_B + \xi} \right) \left( 1 - \frac{P_B}{K} \right) - \frac{\xi N_P}{1 + \xi P_B} - (bD + a)P_B + bDP_L.
\]

\[
\frac{dD}{dt} = qP_L (1 - D) - cD
\]

\[
\frac{dN}{dt} = \left( -N + \frac{hP_L}{n + P_L} \right) \left( \frac{1}{\tau_N} \right)
\]
Bacterial pneumonia - strain dependence

Response to D39 pneumococcus infection in four mice strains
[Data from Kadioglu at al, I&I, 2000; Gingles et al., I&I, 2001; Kadioglu et al., JID, 2011]
Bacterial pneumonia - strain dependence

Response to D39 pneumococcus infection in four mice strains
[Data from Kadioglu et al, I&I, 2000; Gingles et al., I&I, 2001; Kadioglu et al., JID, 2011]

Strain dependent parameters $h, \nu, \xi_{nl}, \xi_{nb}$
References


Credits

Inverse problems
- Jon Rubin
- Shelby Stanhope
- Xiaoyu Duan

Influenza/pneumonia modeling
- Gilles Clermont
- Bard Ermentrout
- Ericka Mochan
- Sarah Lukens

Funding
- NIGMS
- NSF-RTG
Appendix
Posterior distributions

Uniform prior

Reciprocal prior

Jacobian prior

\[(V_0, \beta, p, c, \delta)_{base} = (0.25, 0.014, 2.7 \times 10^{-5}, 3.2, 3.2)\]
Probabilistic trajectory predictions

Uniform prior

Reciprocal prior

Jacobian prior