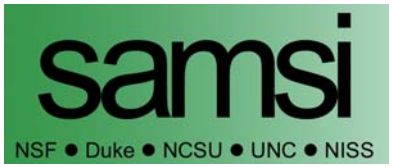
Modeling, Estimation and Control of HIV Dynamics

Research Team: B.M. Adams, H.T. Banks, M. Davidian, S. Grove, S. Hu, H-D Kwon, Y. Ma, E. Rosenberg, J.A. Toivanen, H.T. Tran

Center for Research in Scientific Computation

NC STATE UNIVERSITY



Outline: Efforts in Model Fitting

1. Overview of HIV infection
2. **Data:** clinical treatment interruption study
3. **Modeling goals** and ODE system
4. **Inverse problem** methods: single patient and distributed
5. **Computational results** with simulated data, use of **regularization**
6. **Confidence intervals** for estimates
7. Algorithm and preliminary results for **clinical data**
8. Summary and goals

Figure 1: Virions budding from T cell

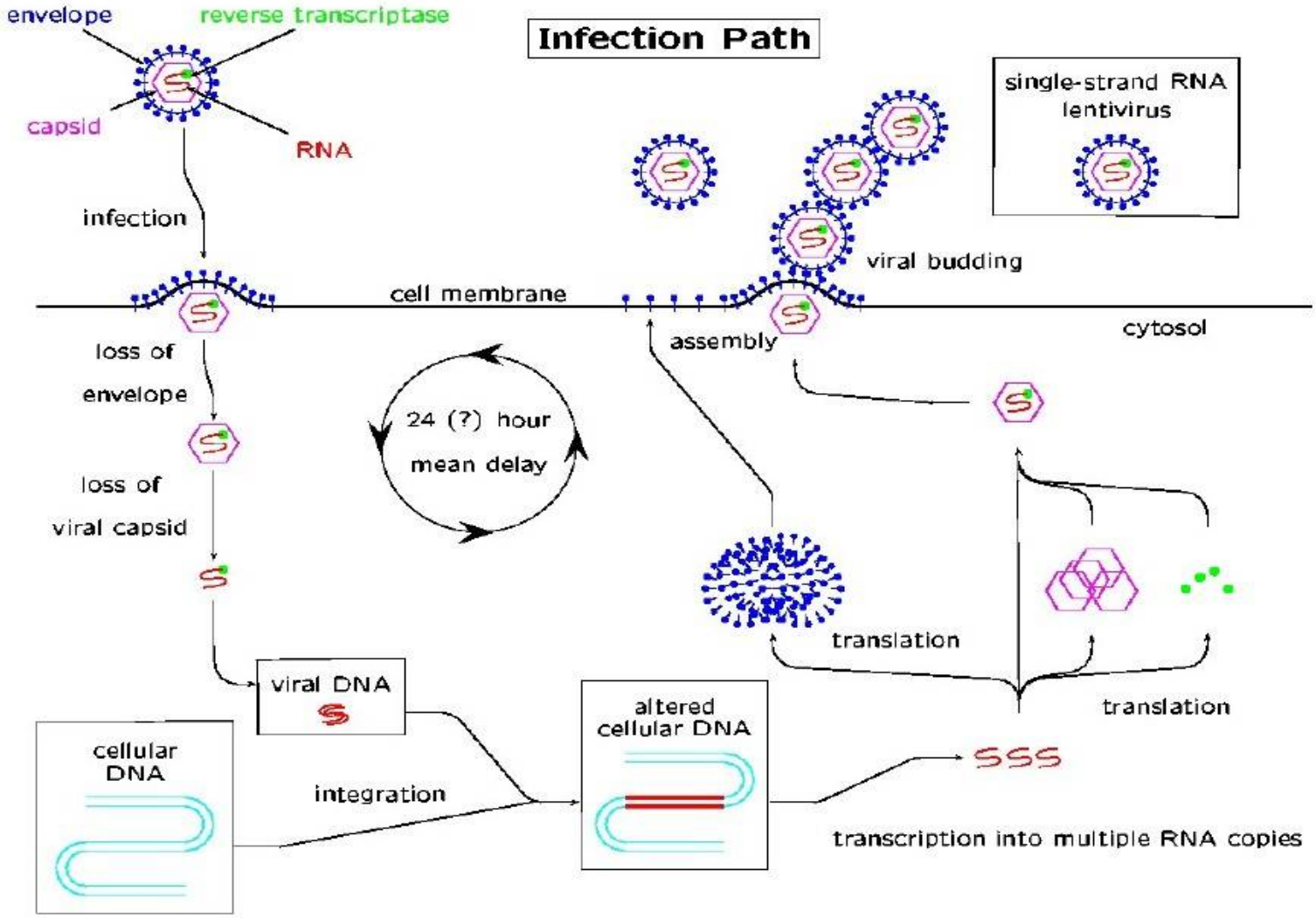


Figure 2: HIV retrovirus infection schematic

HIV and Treatment

- Human Immunodeficiency Virus is a retrovirus.
- Infects CD4 helper **T-cells** of the **immune system** to reproduce
- Typical HIV treatment (combination therapy) **suppresses viral infection and production**.

Difficulties with Continuous Therapy

- Serious side effects of long-term treatment
- Variable patient adherence; lack of availability / high cost of drugs
- Drug efficacy fades as virus mutates, becomes resistant
- Eradicating virus decimates immune system

Clinical Study: Structured Treatment Interruption (STI)

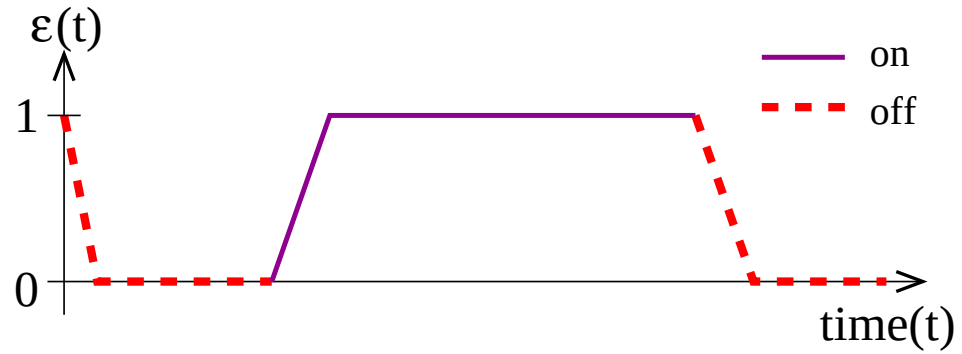
- Eric Rosenberg, M.D., Mass. General Hospital, Boston, **studies STI** – over 120 patients.
- Control drug via fixed schedule or feedback on virus or T-cell quantities

treatment protocol

(control input $\epsilon(t)$:

drug efficacy)

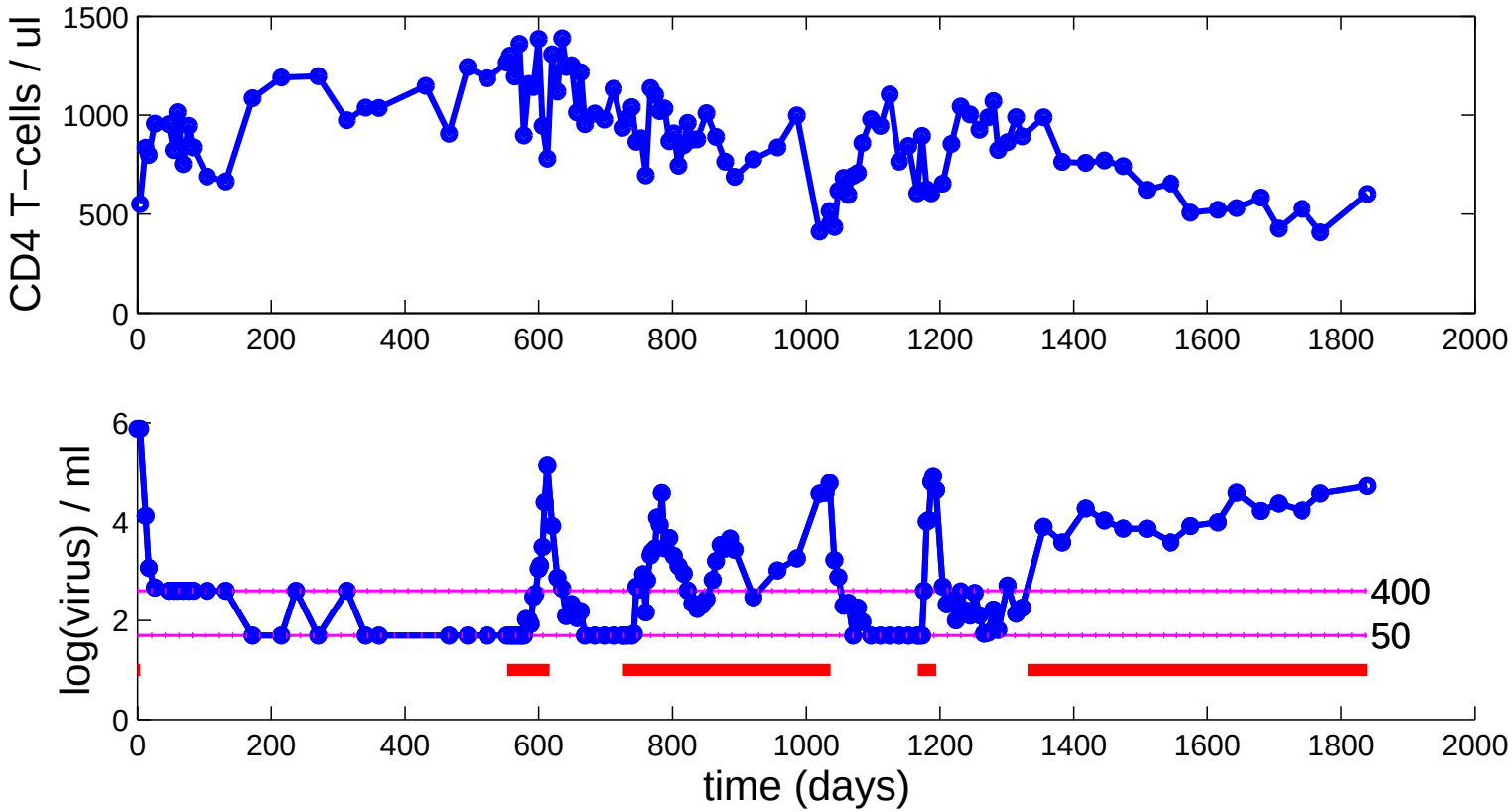
.



Why Interrupt?

- Break from side effects, reduced drug treatment cost
- Boosts the immune system, may cause self-vaccination
- The **Berlin Patient** (Lisziewicz, Rosenberg, et al., 1999), others

Typical Study Data



- **Red bar** denotes off treatment periods – note viral rebound
- Viral load measurements have limit of detection: 400 or 50 copies/ml

Overview: Modeling and Control for HIV

GOAL: Use HIV infection models to help Rosenberg and clinicians understand patient data (e.g, what differentiates rapid progressors from long-term non-progressors) and suggest better treatment schemes.

Article: J. Comp. Appl. Math (CRSC–TR04–05) (invited-special issue on Math Applied to Immunology (2005))

Article: Math. Biosci. Engr. **1** (2004),223–241.

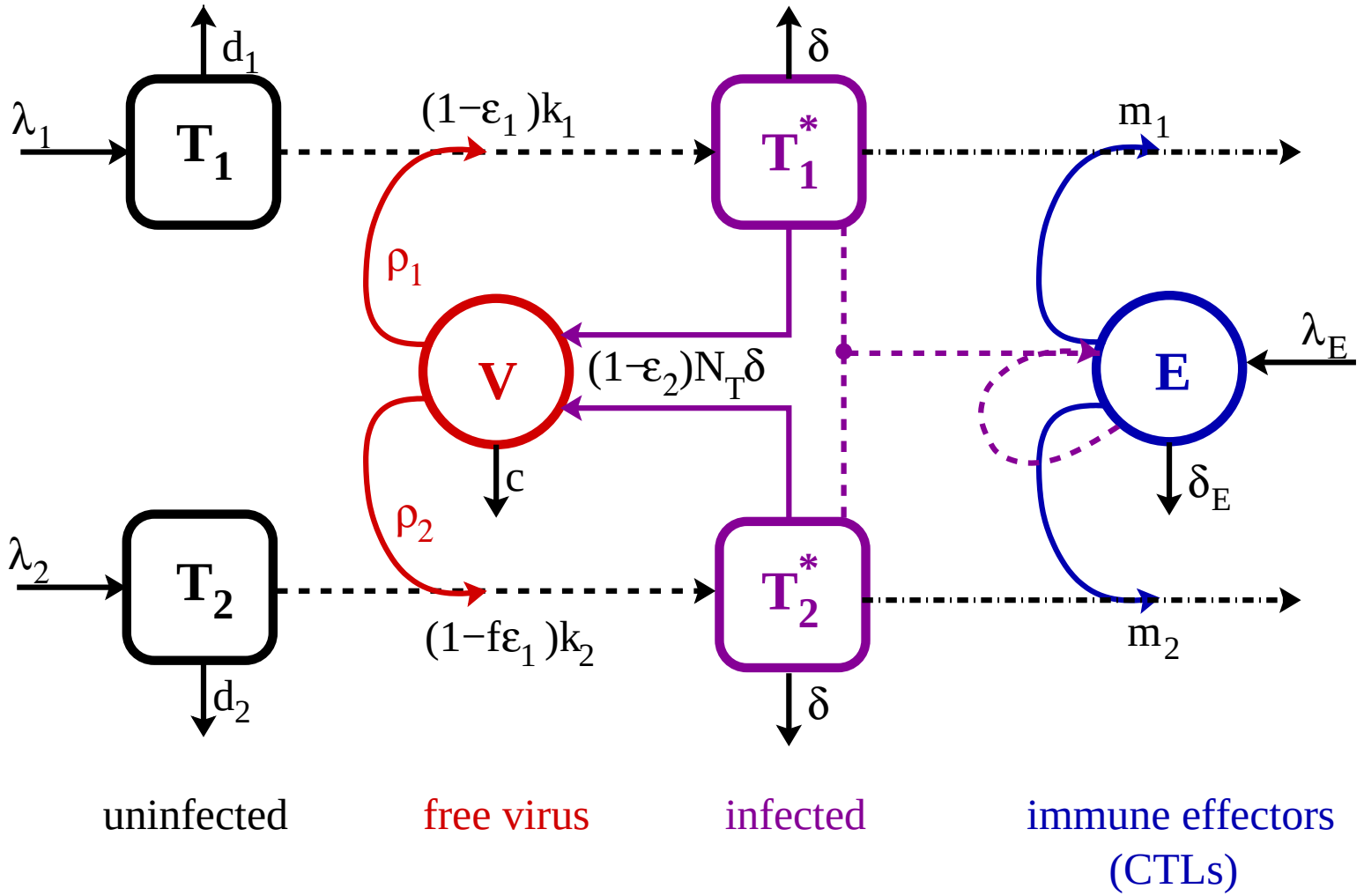
- Model survey, integration and development-study qualitative properties
- Open loop control theory using model to determine optimal treatment schedules
- Selection of patient data to fit based on analysis with POD (SVD, PCA)
- **Develop and apply inverse problem methods to fit model to patient data**

Desired Model Features

- Multiple stable steady states: **viral dominant**; **immune dominant**
- Ability to incorporate single or multi-drug therapy, appropriate sensitivity to drug treatment
- At minimum, model states (compartments) to reflect physiology and data:
  - ⋆ uninfected and **infected** T-cells
  - ⋆ free plasma **virus**
  - ⋆ **immune response**

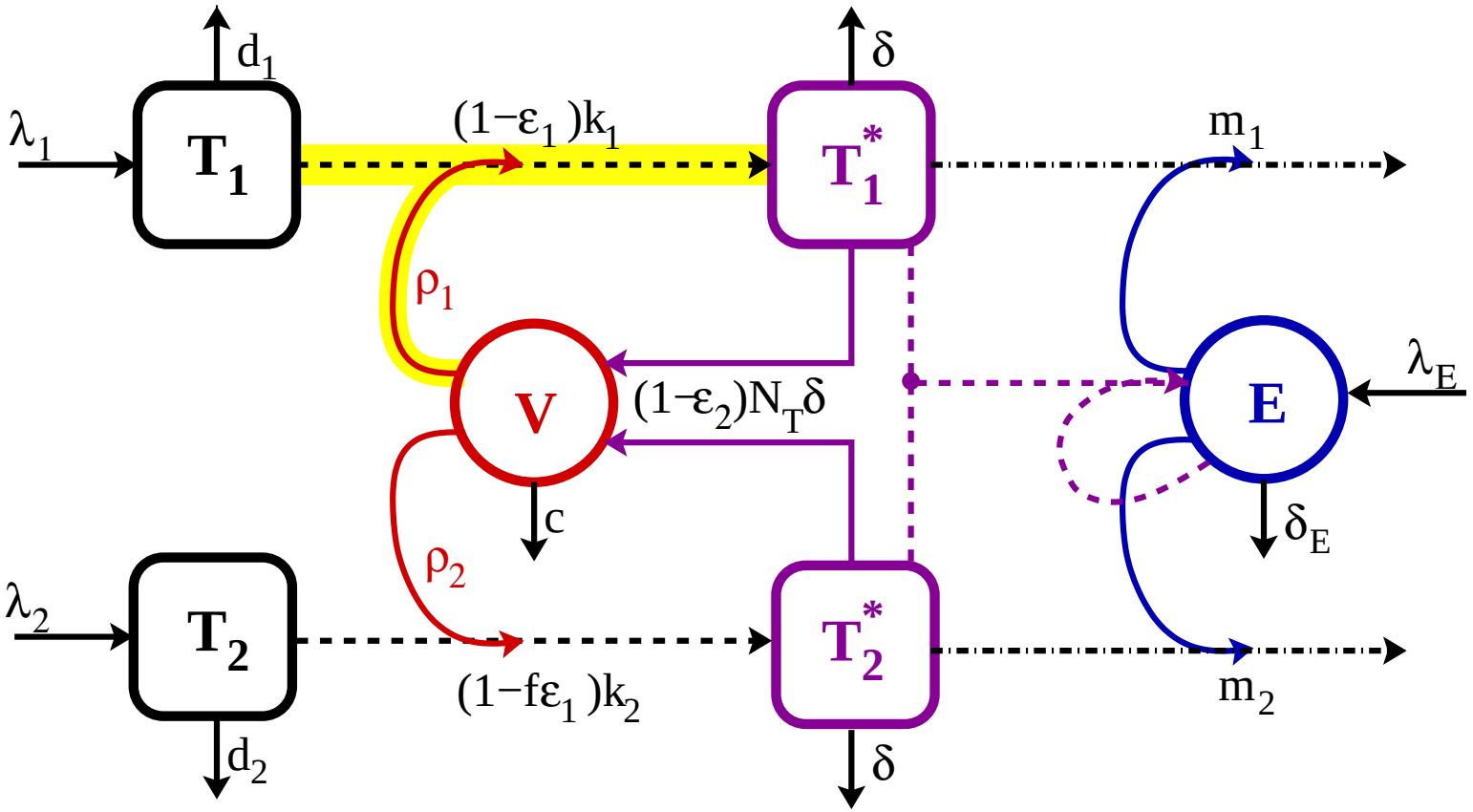
HIV Infection Dynamics Model

- Based on Callaway–Perelson (2001), Bonhoeffer, et. al. (2000) models
- Two co-circulating target cell populations T_1, T_2



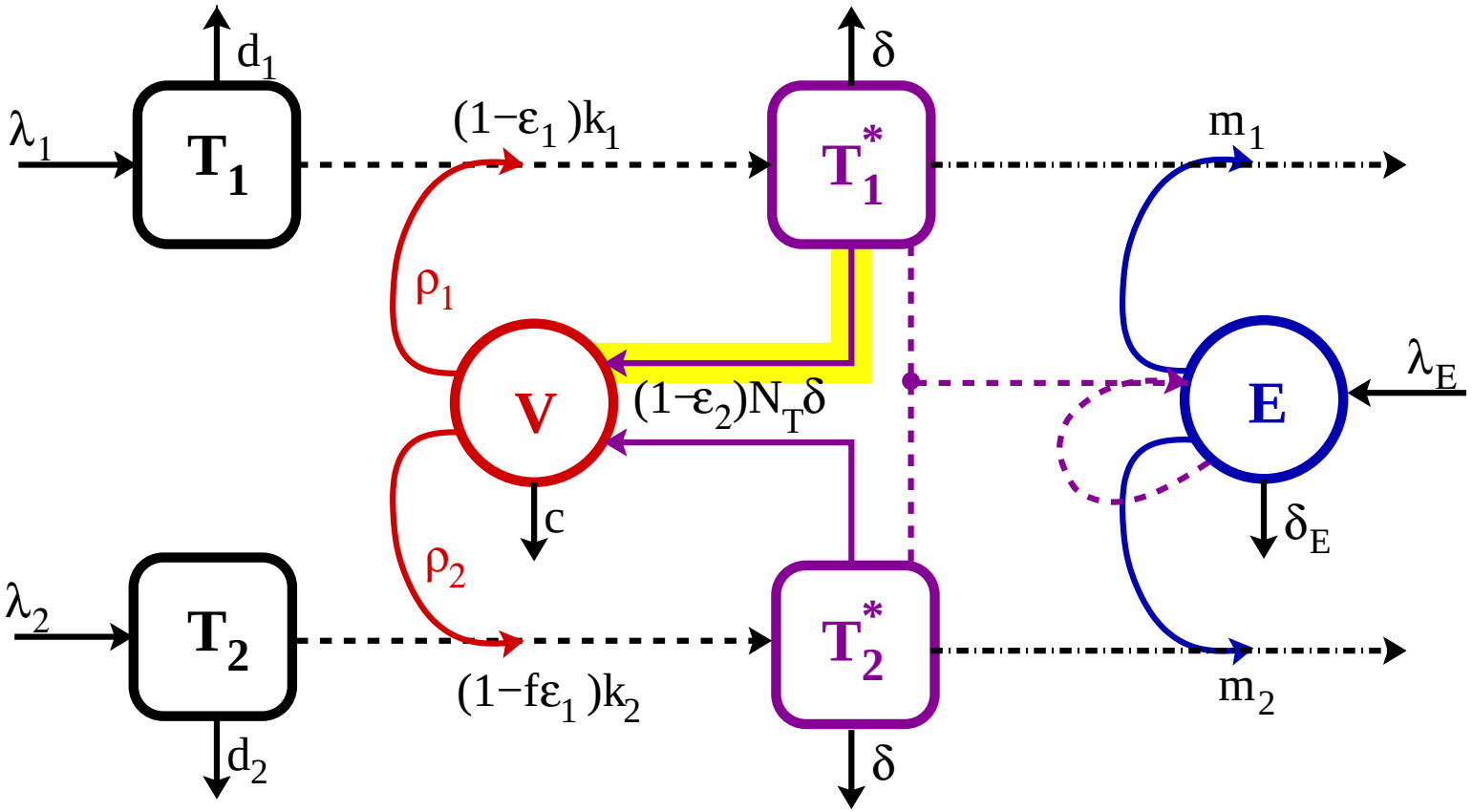
HIV Infection Dynamics Model

- Based on Callaway–Perelson (2001), Bonhoeffer, et. al. (2000) models
- Two co-circulating target cell populations T_1, T_2



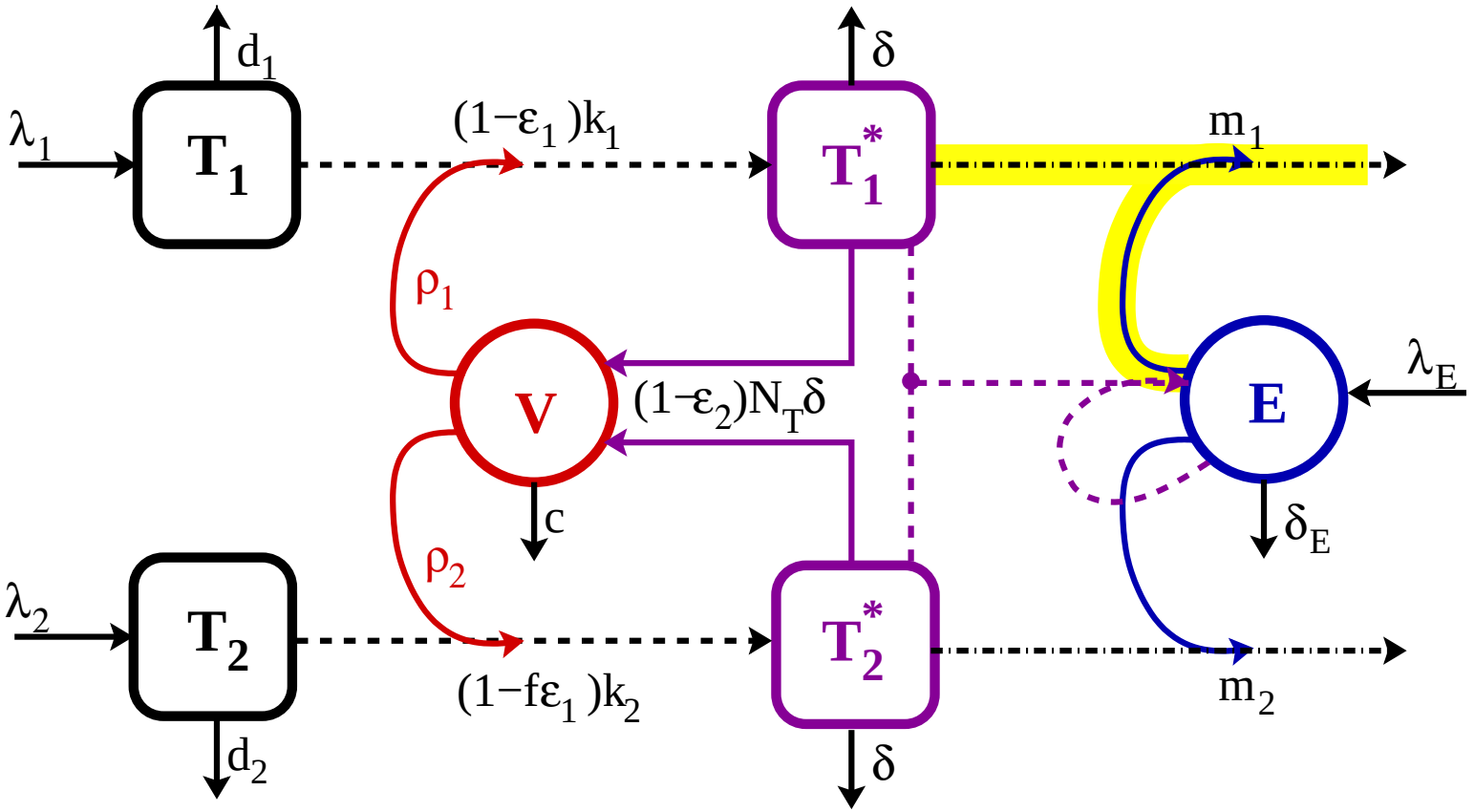
HIV Infection Dynamics Model

- Based on Callaway–Perelson (2001), Bonhoeffer, et. al. (2000) models
- Two co-circulating target cell populations T_1, T_2



HIV Infection Dynamics Model

- Based on Callaway–Perelson (2001), Bonhoeffer, et. al. (2000) models
- Two co-circulating target cell populations T_1, T_2



HIV Infection Dynamics Model

Uninfected type 1: $\dot{\mathbf{T_1}} = \lambda_1 - d_1\mathbf{T_1} - (1-\epsilon_1)k_1\mathbf{V}\mathbf{T_1}$

Uninfected type 2: $\dot{\mathbf{T_2}} = \lambda_2 - d_2\mathbf{T_2} - (1-f\epsilon_1)k_2\mathbf{V}\mathbf{T_2}$

Infected type 1: $\dot{\mathbf{T}}_1^* = (1-\epsilon_1)k_1\mathbf{V}\mathbf{T_1} - \delta\mathbf{T}_1^* - m_1\mathbf{E}\mathbf{T}_1^*$

Infected type 2: $\dot{\mathbf{T}}_2^* = (1-f\epsilon_1)k_2\mathbf{V}\mathbf{T_2} - \delta\mathbf{T}_2^* - m_2\mathbf{E}\mathbf{T}_2^*$

Free virions: $\dot{\mathbf{V}} = (1-\epsilon_2)N_T\delta(\mathbf{T}_1^* + \mathbf{T}_2^*) - c\mathbf{V} - [(1-\epsilon_1)\rho_1 k_1\mathbf{T_1} + (1-f\epsilon_1)\rho_2 k_2\mathbf{T_2}]\,\mathbf{V}$

Immune effectors: $\dot{\mathbf{E}} = \lambda_E + \frac{b_E(\mathbf{T}_1^*+\mathbf{T}_2^*)}{(\mathbf{T}_1^*+\mathbf{T}_2^*)+K_b}\mathbf{E} - \frac{d_E(\mathbf{T}_1^*+\mathbf{T}_2^*)}{(\mathbf{T}_1^*+\mathbf{T}_2^*)+K_d}\mathbf{E} - \delta_E\mathbf{E}$

- From now on, $\mathbf{q}$ denotes one or more model parameters (of interest), for example, $\mathbf{q} = [k_1, c, N_T]$.
- Solve ODE system with `LSODE`, LLNL's `CVODE`, or Matlab's `ODE15s`

Sample Model Equilibria

(Off treatment steady states)

	EQ_0	EQ_1	EQ_2
$\mathbf{T_1}$ (cells/ml)	1000000	163573	967839
$\mathbf{T_2}$ (cells/ml)	3198	5	621
$\mathbf{T_1^*}$ (cells/ml)	0	11945	76
$\mathbf{T_2^*}$ (cells/ml)	0	46	6
$\mathbf{V}$ (copies/ml)	0	63919	415
$\mathbf{E}$ (cells/ml)	10	24	353108
local stability	unstable	stable	stable
	uninfected	*viral dominant*	*immune dominant*

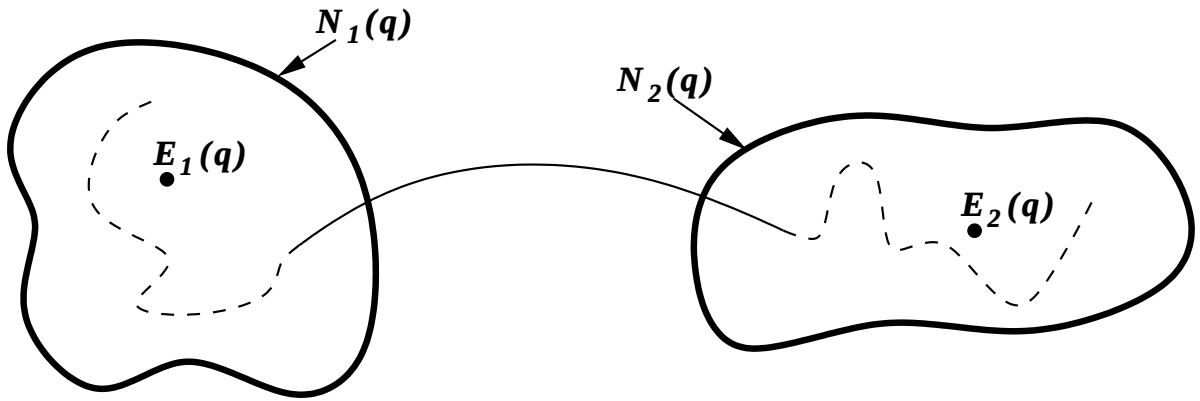


Figure 3: $E_1(q)$: "unhealthy" locally asymptotically stable equilibrium point with its domain of attraction $N_1(q)$; $E_2(q)$: "healthy" locally asymptotically stable equilibrium point with its domain of attraction $N_2(q)$; (- - -) uncontrolled trajectory; (—) controlled trajectory.

Simulated Data Generation

- Have clinical data from patients corresponding to model states

$$\mathbf{x} = \begin{bmatrix} \mathbf{T_1} + \mathbf{T_1^*} \\ \mathbf{V} \\ \mathbf{E} \end{bmatrix} \begin{matrix} \text{(CD4 T-cells)} \\ \text{(free virions)} \\ \text{(CTL immune response)} \end{matrix}$$

- Verify methods on simulated data y_s^i generated for times $t^i, i = 1, \ldots, N$:

$$y_s^i = x_s(t^i; \mathbf{q}) + \epsilon_s^i.$$

Here, $s = 1, 2, 3$, indexes the components of the state $\mathbf{x}$ and errors ϵ_s^i are such that

⋆ **Mean response** given by ODE model: $\mathcal{E}(y_s^i) = x_s^i(\mathbf{q})$

⋆ **Variance model:** $\mathrm{Var}(y_s^i) = \sigma_s^2 \left\{x_s^i(\mathbf{q})\right\}^2$
(constant coefficient of variation or lognormal model – typical for blood draws)

Next: Fitting Model to Data

1. Overview of HIV infection
2. **Data:** clinical treatment interruption study
3. **Modeling goals** and ODE system
4. **Inverse problem** methods: single patient and distributed
5. **Computational results** with simulated data, use of **regularization**
6. **Confidence intervals** for estimates
7. Algorithm and preliminary results for **clinical data**
8. Summary and goals

Multiple Patient Inverse Problem

For each patient $j = 1 \ldots N_P$, we have (clinical or simulated) data pairs $(t^{ij}, \mathbf{y}^{ij})$ at times $t^{ij}, i = 1, \ldots, N_j$.

GOAL: Understand how one or more model parameters (e.g. k_1, infectivity) varies across the population – can we estimate distributions of parameters from patient data?

Multiple Patient Inverse Problem Approaches

- Fit ODE model to **each patient** j yielding parameters $\mathbf{q}_j$:

$$\mathbf{q}_j^* = \arg\min_{\mathbf{q}\in Q} J(\mathbf{q}) = \frac{1}{N_j}\sum_{i=1}^{N_j}\left|\mathbf{x}(t^i;\mathbf{q}) - \mathbf{y}^{ij}\right|^2$$

  (standard nonlinear least squares), then perform statistical analysis.

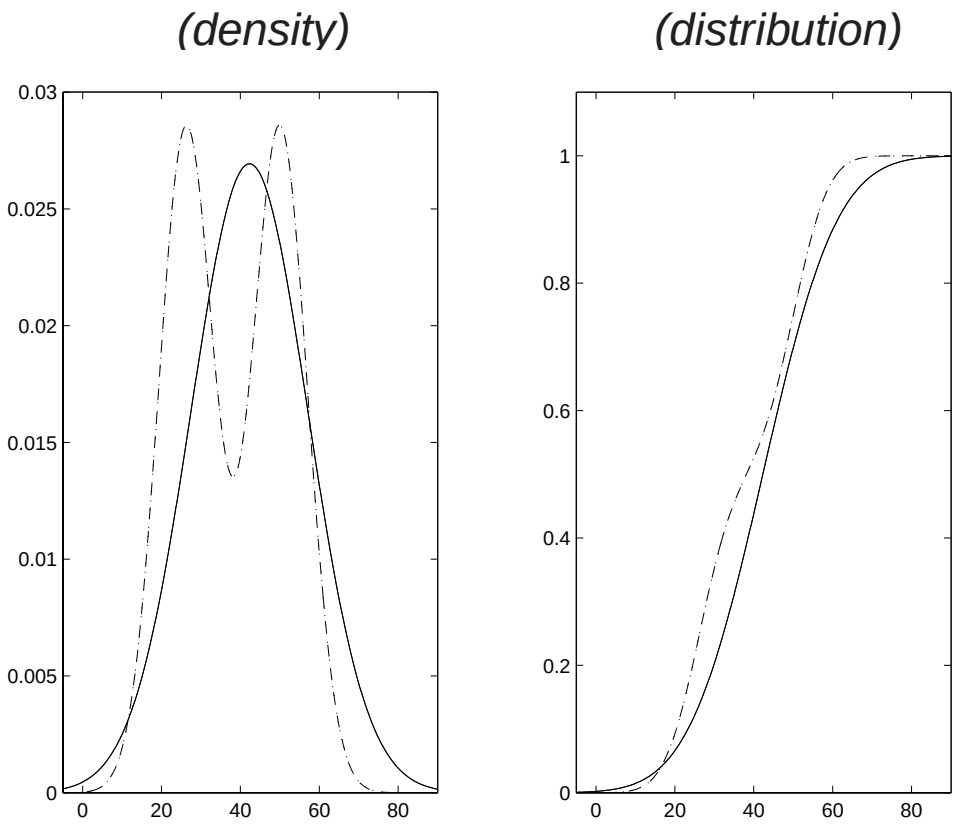
- Fit model to **all patients simultaneously**; each has a $\mathbf{q}_j$. Given a probability space $\mathcal{Q}$ in which the parameters of interest $\mathbf{q}$ live:

$$P^* = \arg\min_{P\in\mathcal{Q}} J(P) = \frac{1}{N_P}\sum_{j=1}^{N_P}\frac{1}{N_j}\sum_{i=1}^{N_j}\left|\mathcal{E}\left[\mathbf{x}(t^{ij};\mathbf{q})|P(\mathbf{q})\right] - \mathbf{y}^{ij}\right|^2$$

- Fully hierarchical – estimate distributions of parameters and errors for each patient as well, assuming a model for their overall distribution.

Why Estimate Distribution P?

- Estimate using data from multiple patients (**less costly**)
- Admits non-parametric distribution – can **avoid distribution misspecification** (e.g., seek parameters for normal when reality bimodal – example from Banks, Ma, Potter (2004))



Possible Spaces of Distributions

$$P^* = \arg\min_{P \in \mathcal{Q}} J(P) = \frac{1}{N_P} \sum_{j=1}^{N_P} \frac{1}{N_j} \sum_{i=1}^{N_j} \left| \mathcal{E}\left[\mathbf{x}(t^{ij}; \mathbf{q}) | P(\mathbf{q}) \right] - \mathbf{y}^{ij} \right|^2$$

- Problem assumes **model parameters** of interest $\mathbf{q}$ for each patient **are realizations of a random variable with probability distribution** P.
- P belongs to a probability space $\mathcal{Q}$ of distributions on the admissible parameter space Q. For example $\mathcal{Q}$ could be the set
  - ⋆ of normal distributions over Q, parameterized by $\mathcal{N}(\mu, \sigma^2)$
  - ⋆ $\mathcal{P}^M = \{P = \sum_{k=0}^{M} p_k \delta_{q_k}\}$, where $q_k \in Q, p_k \geq 0, \sum_{k=0}^{M} p_k = 1$ (point masses – nonparametric)

Inverse Problem: Theory and a Special Case

Inverse problem: Minimize $J(P)$ over $P \in \mathcal{Q} \subset \mathcal{P}(Q)$

- Banks/Bihari (2001): $P \to J(P)$ continuous in Prohorov ρ metric, with conditions on $\mathcal{Q}$ are sufficient to establish a minimizer's existence.

- Our special case: distributions P **characterized by their densities**: For $\mathcal{F} \subset L^2(Q)$, define

$$\mathcal{P}_{\mathcal{F}}(Q) := \{P \in \mathcal{P}(Q) | P' = f, f \in \mathcal{F}\}$$

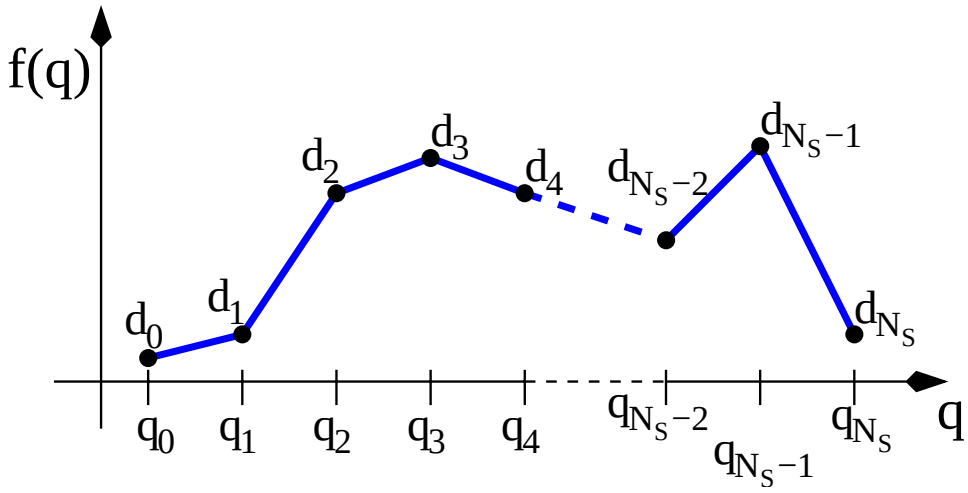
- Inverse problem is equivalent to minimizing $J\left(P(f)\right)$ over densities f, where

$$\mathcal{E}\left[\mathbf{x}(t^{ij}; q) | P(f)\right] = \int_Q \mathbf{x}(t^{ij}; q)\, f(q)\, dq$$

Approximating Probability Densities

Computationally: approximate densities $f(q)$ by **piecewise linear splines** defined on a finite dimensional set $\{q_k\}_{k=0}^{N_S} \subset Q$.

$$\left\{ f \approx \sum_{k=0}^{N_S} d_k \phi_k(q),\ d_k \geq 0,\ \sum_{k=0}^{N_S-1} \frac{\Delta q_k}{2} (d_k + d_{k+1}) = 1 \right\}$$



Banks with Pinter (2004), Potter (2003), Bihari (2001): such spline approximations yield well-posed inverse problems which converge in the Prohorov metric (hence in distribution).

Inverse Problem in this Setting: Quadratic Programming

$$
\begin{aligned}
J(P(f)) &= \frac{1}{N_P}\sum_{j=1}^{N_P}\frac{1}{N_j}\sum_{i=1}^{N_j}\left|\mathcal{E}\left[\mathbf{x}(t^{ij};q)|P(f(q))\right]-\mathbf{y}^{ij}\right|^2 \\
&= \frac{1}{N_P}\sum_{j=1}^{N_P}\frac{1}{N_j}\sum_{i=1}^{N_j}\left|\int_Q \mathbf{x}(t^{ij};q)f(q)dq-\mathbf{y}^{ij}\right|^2 \\
&\approx \frac{1}{N_P}\sum_{j=1}^{N_P}\frac{1}{N_j}\sum_{i=1}^{N_j}\left|\left\{\sum_{k=0}^{N_S-1}\frac{\Delta q_k}{2}\left(\mathbf{x}^{ij}(q_k)d_k+\mathbf{x}^{ij}(q_{k+1})d_{k+1}\right)\right\}-\mathbf{y}^{ij}\right|^2
\end{aligned}
$$

which is a **constrained quadratic programming problem** in the coefficients $d_k, k=0,\ldots,N_S$:

$$\boxed{J(\mathbf{d})=\mathbf{d}^T A\mathbf{d}+2\mathbf{b}^T\mathbf{d}+c}$$

where $A, \mathbf{b},$ and c are functions of the data $\mathbf{y}^{ij}$ and model solutions $\mathbf{x}^{ij}$ at fixed nodes q_k.

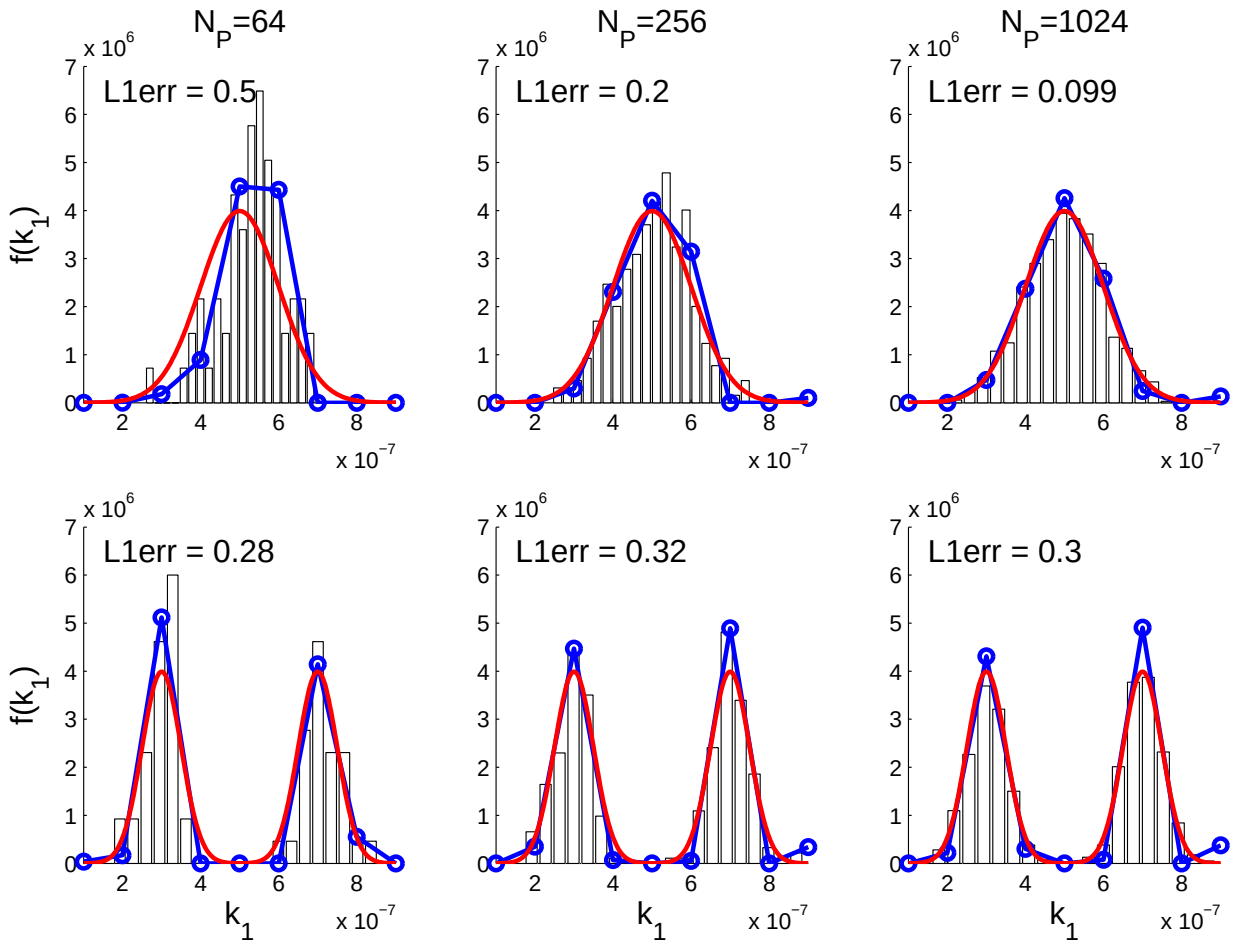
Computational Considerations

- Given nodes q_k, model solutions $\mathbf{x}(t; q_k)$ can be computed offline.
- Easily parallelizable, enabling adaptive node placement in reasonable time
- Matlab's `quadprog` solves $J(\mathbf{d}) = \mathbf{d}^T A\mathbf{d} + 2\mathbf{b}^T\mathbf{d} + c$ in *seconds*.

Next: Results for Density Estimation

1. Overview of HIV infection
2. **Data:** clinical treatment interruption study
3. **Modeling goals** and ODE system
4. **Inverse problem** methods: single patient and distributed
5. **Computational results** with simulated data, use of **regularization**
6. **Confidence intervals** for estimates
7. Algorithm and preliminary results for **clinical data**
8. Summary and goals

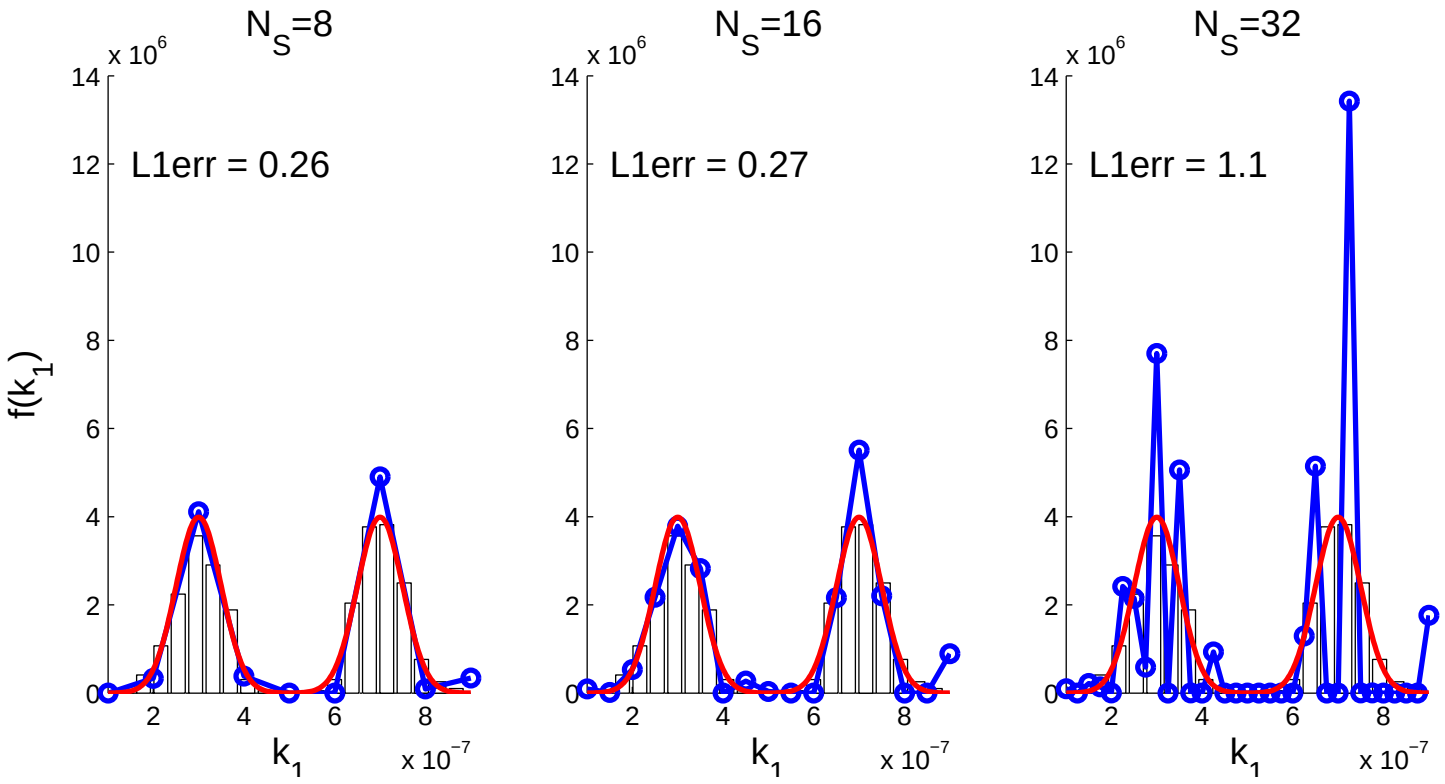
Results: Improvement with Larger Sample Size N_P



Results for virtual patients with infectivity k_1 sampled from a normal or bimodal distribution. Red: true density, bars: samples used for virtual patients, and blue: estimated density.

Improvement with More Splines?

Expect convergence as $N_S \longrightarrow \infty \ldots$



$\kappa_2(A) = 2.3 \times 10^6$ $\kappa_2(A) = 4.4 \times 10^{10}$ $\kappa_2(A) = 1.4 \times 10^{18}$

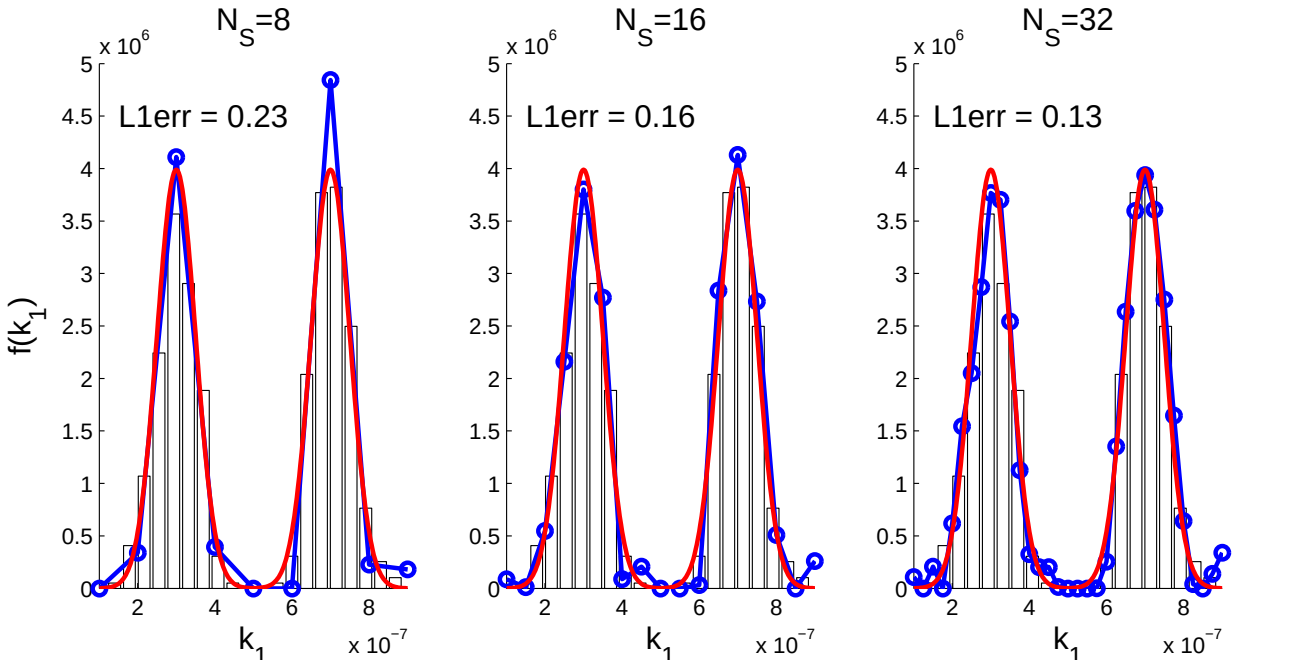
Problem condition worsens as number of nodes increases.

Help from Regularization

Modified cost criterion to penalize non-smoothness:

$$J(f) = \frac{1}{N_P} \sum_{j=1}^{N_P} \frac{1}{N_j} \sum_{i=1}^{N_j} \left| \mathcal{E}\left[\mathbf{x}(t^{ij}; q) | f \right] - \mathbf{y}^{ij} \right|^2 + \beta \left\| \frac{df}{dq}(q) \right\|_{L_2}^2$$

Approximate regularization term with trapezoid rule, build directly into quadratic programming problem A matrix.



$\kappa_2(A) = 1.6 \times 10^6$ $\kappa_2(A) = 1.7 \times 10^6$ $\kappa_2(A) = 1.2 \times 10^6$

Next: Uncertainty of Estimation Process

1. Overview of HIV infection
2. **Data:** clinical treatment interruption study
3. **Modeling goals** and ODE system
4. **Inverse problem** methods: single patient and distributed
5. **Computational results** with simulated data, use of **regularization**
6. **Confidence intervals** for estimates
7. Algorithm and preliminary results for **clinical data**
8. Summary and goals

Quantifying Uncertainty in Estimated Parameters

Goal: Quantify **variability of estimates** $f^*(q)$ this process yields.

First: Standard errors for NLSQ inverse problem for parameter vector $\mathbf{q}$:

$$\mathbf{q}^* = \arg\min_{\mathbf{q}} \sum_{i=1}^{N} \left| \mathbf{x}(t_i; \mathbf{q}) - \mathbf{y}^i \right|^2$$

If $X_{\mathbf{q}}(\mathbf{q}) = \frac{\partial \mathbf{x}}{\partial \mathbf{q}}$ is the Jacobian matrix for the ODE model responses w.r.t parameters, large sample theory dictates

$$\boxed{\hat{\mathbf{q}} \sim \mathcal{N}\left(\mathbf{q_0}, \boldsymbol{\Sigma}\right)} \text{ where } \Sigma = \left\{ X_{\mathbf{q}}(\mathbf{q}_0)^T G^{-1} X_{\mathbf{q}}(\mathbf{q}_0) \right\}^{-1}.$$

Here G is the diagonal weighting matrix of the variances:

$$\left(\sigma_1^2\{x_1^1\}^2, \ldots, \sigma_1^2\{x_1^N\}^2; \ldots; \sigma_3^2\{x_3^1\}^2, \ldots, \sigma_3^2\{x_3^N\}^2\right)$$

Yields Standard Errors and Confidence Intervals

Standard errors for parameter component k are given by $s_k = \sqrt{\Sigma_{kk}}$.

For large samples, a **95% confidence interval** can be constructed for each parameter component k:

$$[q_k^* - 2s_k, q_k^* + 2s_k]$$

- Given the estimation procedure, 95% of intervals constructed this way will include the true value of the parameter q_k.
- Requires sensitivity computations to determine $X_{\mathbf{q}}(t; \mathbf{q})$

Quantifying Uncertainty in Spline Coefficients

Rather than just $\mathbf{x}(t^i; \mathbf{q})$, the least squares cost is based on

$$M \equiv \mathcal{E}\left[x_s(t^i; q)|P\right] = \int_Q x_s(t^i; q) f(q) dq \approx \int_Q x_s(t^i; q) \sum_{k=0}^{N_S} d_k \phi_k(q) dq$$

(linear in d_k), so derivatives w.r.t parameters of interest (d_k) yield exact matrix entries

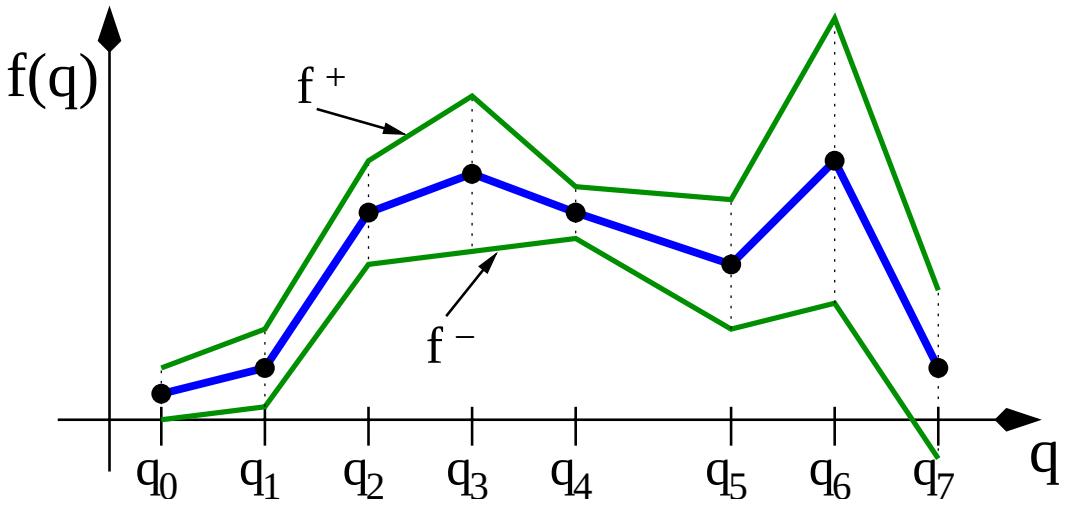
$$[X_q]_{ik} = \left[\frac{\partial M}{\partial d_k}\right] = \int_Q x_s(t^i; q) \phi_k(q) dq$$

- Can use same formulation from large sample theory above on the nodes d_k, but no need for sensitivity computations
- Unlike in single patient case, do not have good estimator for variance σ^2

Uncertainty in Estimating Spline Coefficients

Construct **piecewise linear standard error bands** using estimated coefficients d_k^* and corresponding standard errors s_k

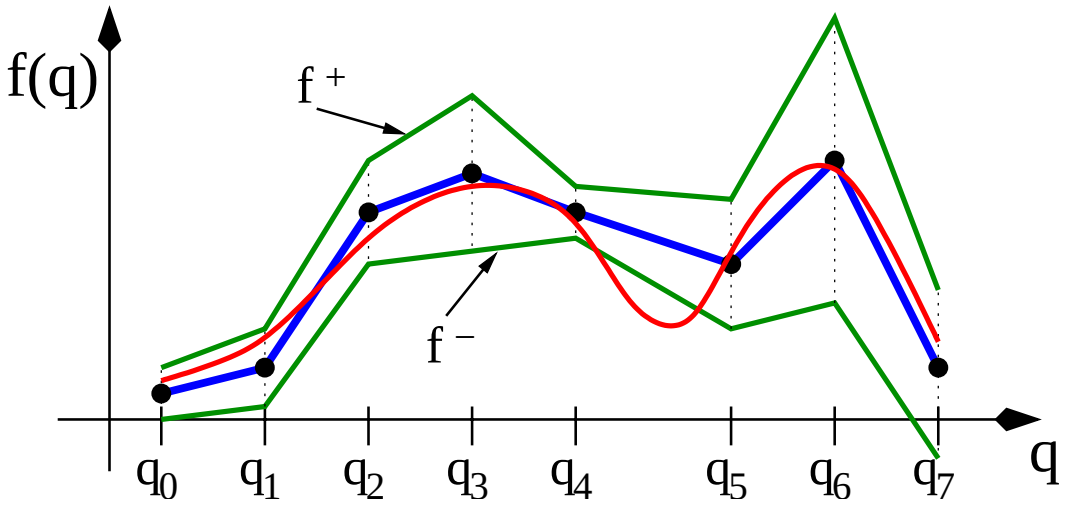
$$f^- = \sum_{k=0}^{N_S} (d_k^* - 2s_k)\phi_k(q) \leq f^* \leq \sum_{k=0}^{N_S} (d_k^* + 2s_k)\phi_k(q) = f^+$$



Strictly a Nodal Confidence Interval

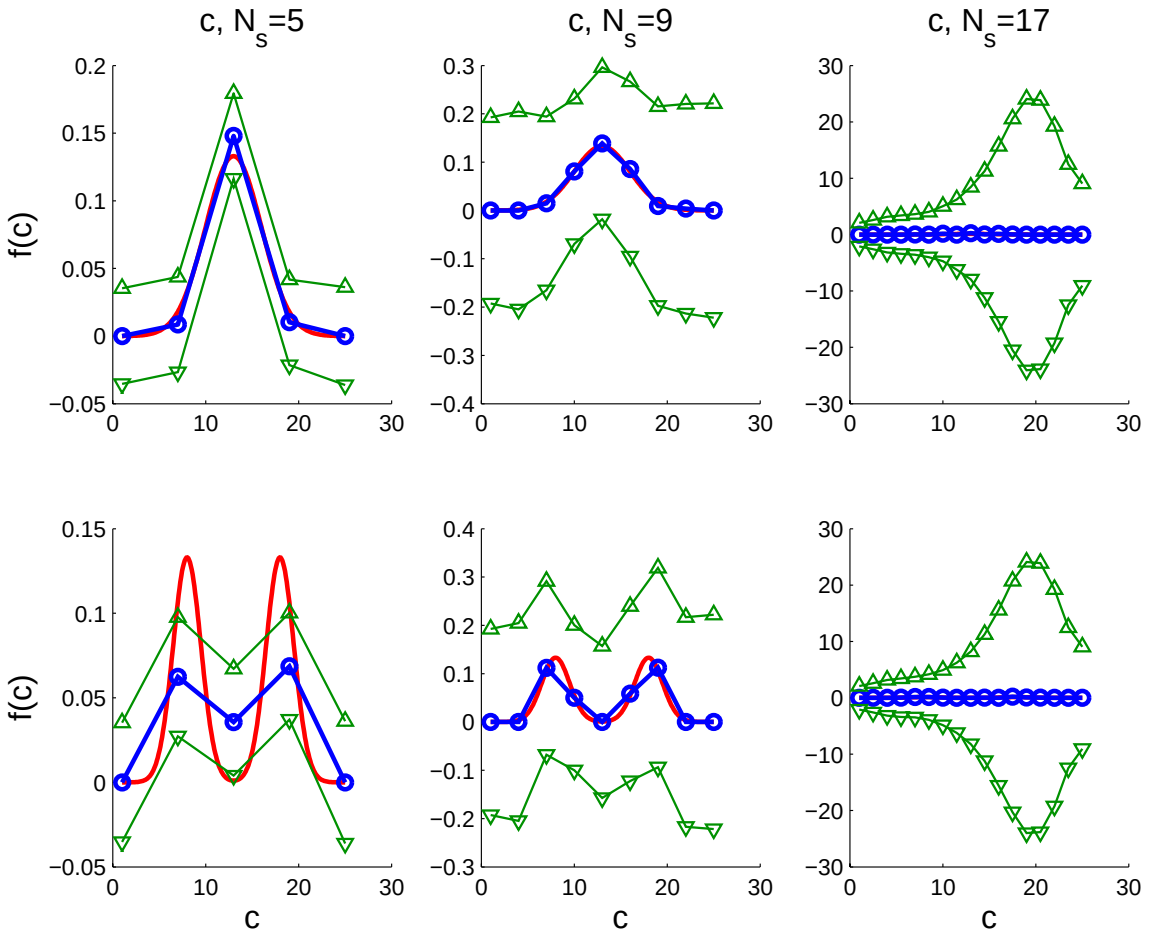
$$f^- = \sum_{k=0}^{N_S} (d_k^* - 2s_k)\phi_k(q) \leq f^* \leq \sum_{k=0}^{N_S} (d_k^* + 2s_k)\phi_k(q) = f^+$$

- **Not a functional confidence interval**, but a **nodal confidence interval**. 95% of intervals at nodes cover true nodal values – may or may not cover actual underlying function.



- Perhaps compare to or extend ideas of Wahba, et.al., on CI for smoothing splines to get true **functional confidence bands**.

Nodal Confidence Intervals on Splines for viral clearance c



Nodal CIs for c sampled from a normal or bimodal distribution. True density, estimated density, confidence intervals. *Note varying y-axis scale.*

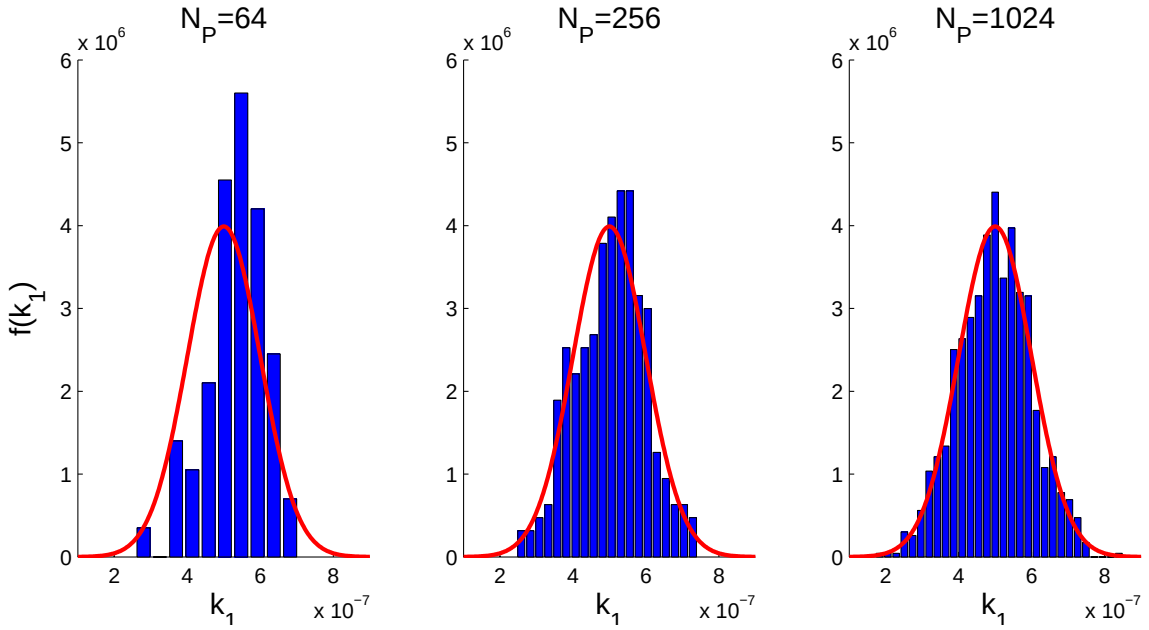
Estimating each Patient Individually

Solve for each patient j,

$$\mathbf{q}_j^* = \arg\min_{\mathbf{q}\in Q_{ad}} J(\mathbf{q}) = \frac{1}{N}\sum_{i=1}^{N} \left|\mathbf{x}(t^i;\mathbf{q}) - \mathbf{y}^{ij}\right|^2$$

- Used simulated data for 2048 patients; considered normal and bi-modal distributions various parameters. Inverse problem solved with Levenberg-Marquardt.
- **Computationally intensive:** several minutes *per patient* vs. several seconds for estimating density
- However, **gain understanding about each patient** as well as population

Results: Estimating Parameter k_1 per Patient



N_P	32	64	128	256	512	1024	2048
sample $\mu(\times 10^{-7})$	5.2991	5.3142	5.1789	5.0419	5.0260	4.9925	4.9973
sample $\sigma(\times 10^{-7})$	0.8346	0.8557	0.9492	0.9470	0.9990	1.0051	0.9838

Next: Results for Clinical Data

1. Overview of HIV infection
2. **Data:** clinical treatment interruption study
3. **Modeling goals** and ODE system
4. **Inverse problem** methods: single patient and distributed
5. **Computational results** with simulated data, use of **regularization**
6. **Confidence intervals** for estimates
7. Algorithm and preliminary results for **clinical data**
8. Summary and goals

Methodology for Censored Clinical Data

- Recall: viral load assays have **lower limit of quantification**: $L = 400$ or 50 cop/ml
- Need to **quantify uncertainty** about censored data, leveraging knowledge that they are below detection limit (in $[0, L]$)
- Still assume viral load V data y_2^{ij} arise from model $x_2^{ij}(q)$, but when below the limit of detection, assume data follow **truncated** distribution

---

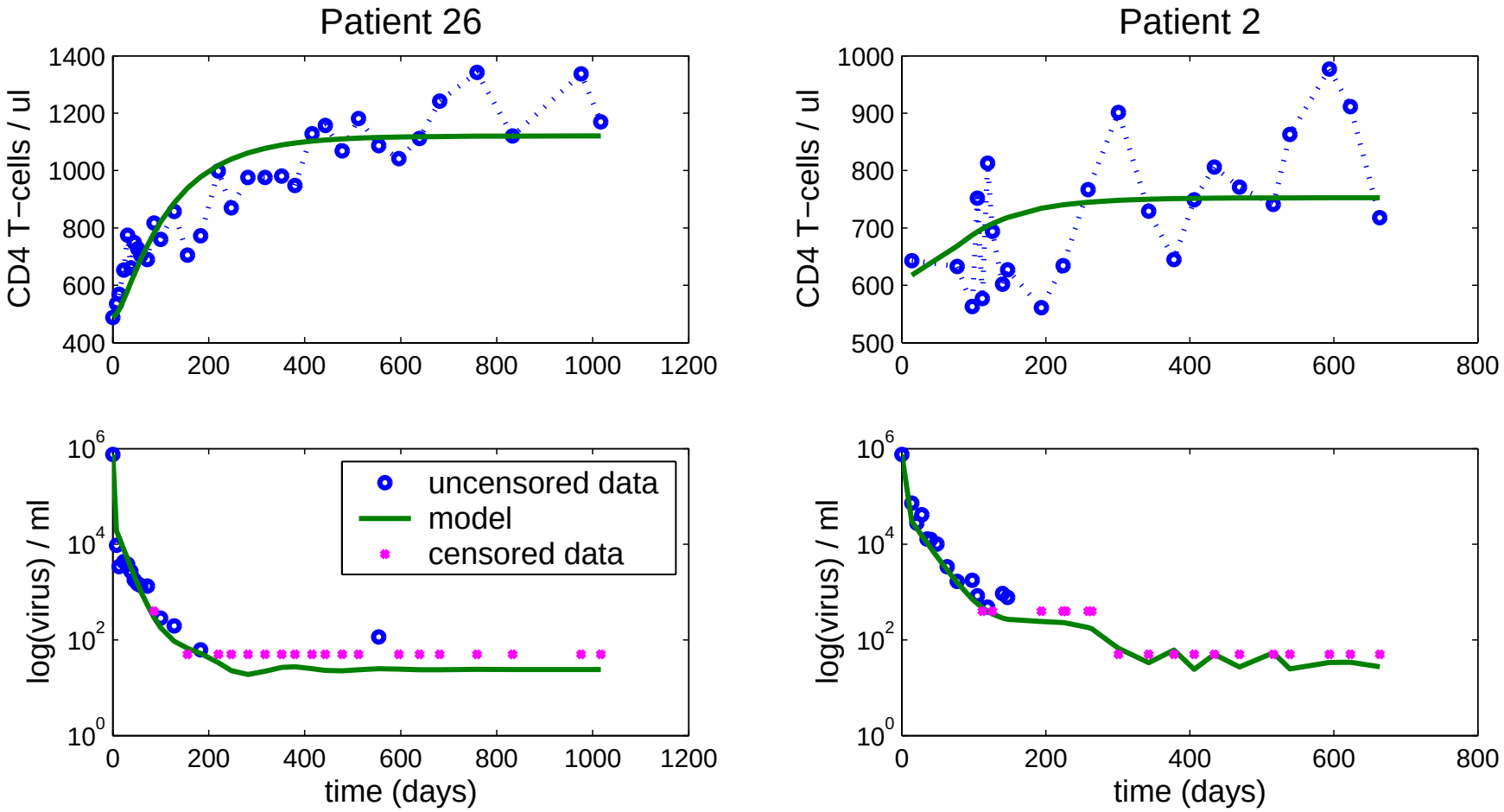
1. Given an estimate q^*, for censored data points, calculate

$$\mathcal{E}\left[y_2^{ij} | y_2^{ij} < L\right] \quad \text{and} \quad \mathcal{E}\left[\left(y_2^{ij}\right)^2 | y_2^{ij} < L\right]$$

   use the former to **replace censored data points** and both to update the estimator for variance $\hat{\sigma}^2$.

2. Solve the optimization problem to update q^*, return to 1. and iterate until convergence.

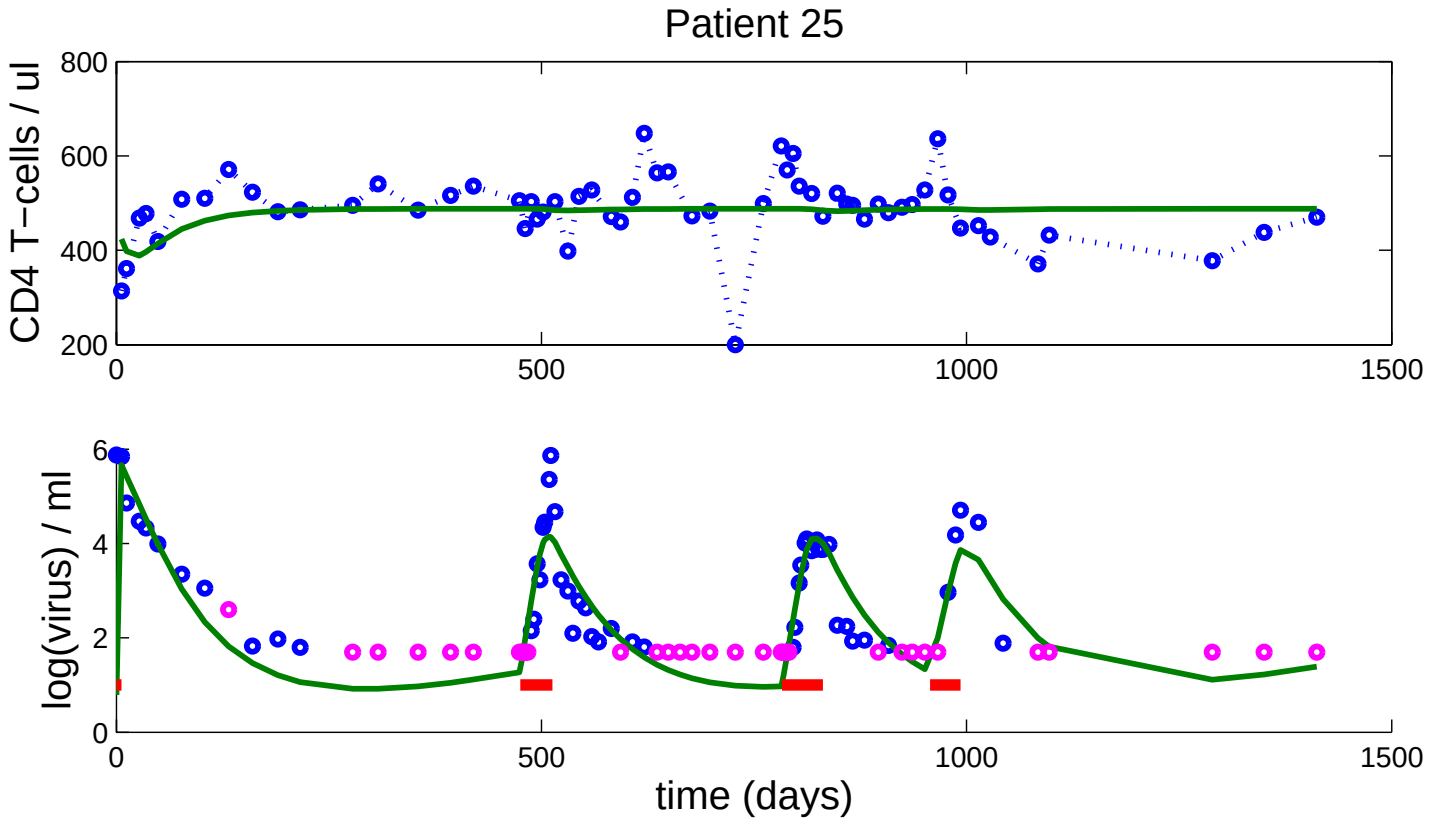
Sample Model Fits: On Treatment Data



Estimate parameters $d_1, k_1, d_2, k_2, \delta, N_T, c$, using DIRECT algorithm as implemented by Dan Finkel (NCSU).

Sample Model Fit: STI (On/Off Treatment) Data

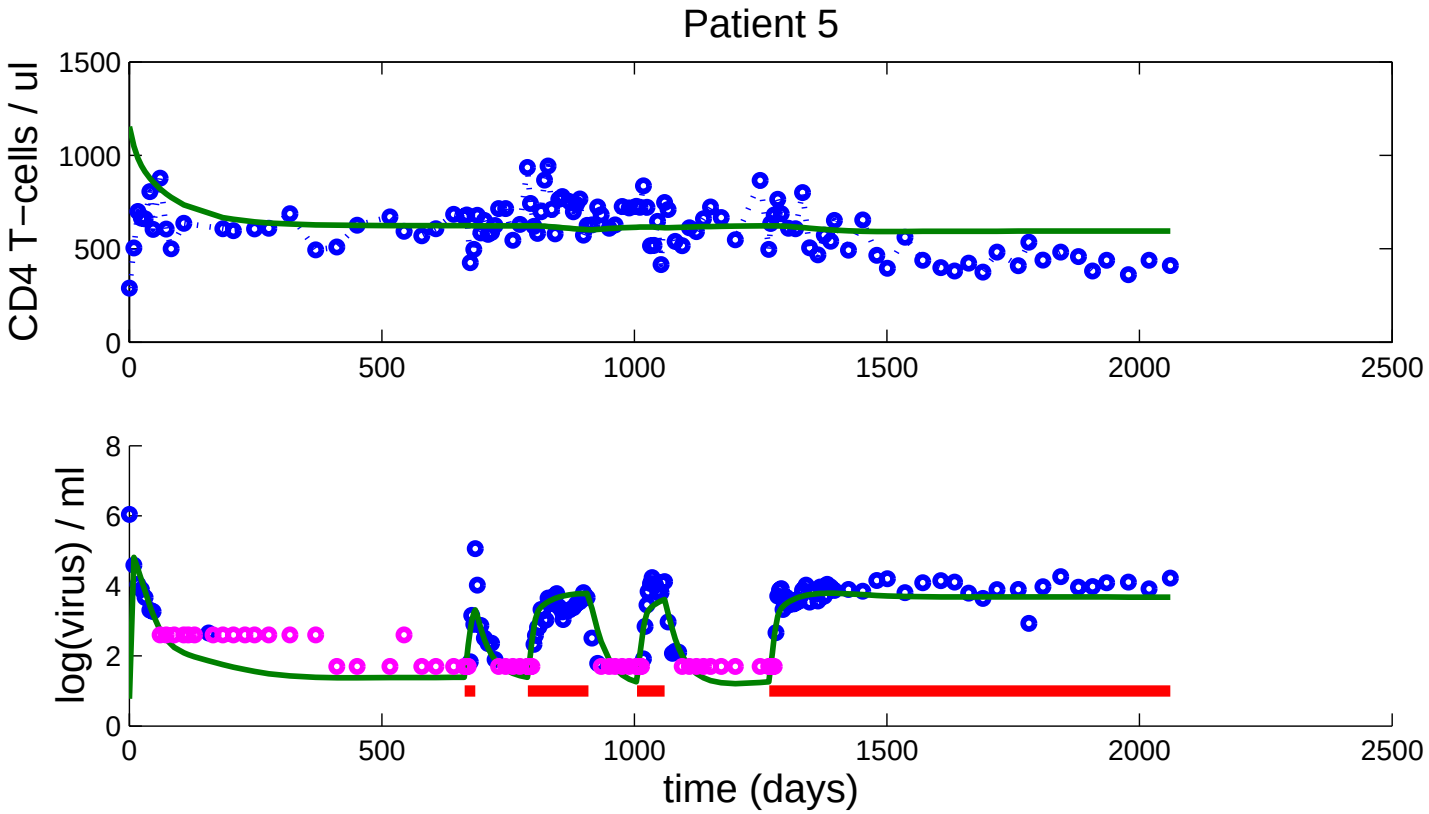
Estimate all parameters and initial conditions in two passes using `DIRECT` algorithm, fine tuned with Matlab's `lsqnonlin`



(red denotes off treatment periods)

Sample Model Fit: STI (On/Off Treatment) Data

Estimate all parameters and initial conditions in two passes using `DIRECT` algorithm, fine tuned with Matlab's `lsqnonlin`



(red denotes off treatment periods)

Summary and Goals

- Developed differential equation model with desired features to represent data
- Fit data from study patients (on therapy and on STI) using censored data algorithm; need to investigate other T-cell dynamics models, obtain better fits to data
- Verified distribution estimation procedure on simulated data – can detect different distributions – need to apply to clinical data sets
- Can construct nodal confidence bands; want to quantify overparametrization and complete theory for functional CIs
- In context of density estimation, need means to handle: censored data, estimate of error model variance

Publications

- B.M. Adams, H.T. Banks, H.T. Tran, and H. Kwon, Dynamic Multidrug Therapies for HIV: Optimal and STI Control Approaches, *CRSC Tech. Rpt. CRSC-TR04-18,* NC State University, April 2004; *Mathematical Biosciences and Engineering* 1 (2004), 223-241.

- B.M. Adams, H.T. Banks, M. Davidian, et. al., HIV Dynamics: Modeling, Data Analysis, and Optimal Treatment Protocols, *CRSC Tech. Rpt. CRSC-TR04-05,* NC State University, February 2004; *Journal of Computational and Applied Mathematics,* to appear, 2005.

- B.M. Adams, H.T. Banks, J.E. Banks, and J.D. Stark, Population Dynamics Models in Plant - Insect Herbivore - Pesticide Interactions, *CRSC Tech. Rpt. CRSC-TR03-12,* NC State University, March 2003; *Mathematical Biosciences,* to appear.

Other References

H.T. Banks and K.L. Bihari, Modelling and estimating uncertainty in parameter estimation, *Inv. Probs.*, **17** (2001), 95-111.

H.T. Banks, Y. Ma, and L.K. Potter, A simulation-based comparison between parametric and nonparametric estimation methods in PBPK models, *CRSC Tech. Rpt. CRSC-TR04-25,* NCSU, Raleigh, June, 2004.; *J. Inverse and Ill–Posed Problems,* **13** (2005), 1–26.

H.T. Banks and G.A. Pinter, A probabilistic multiscale approach to hysteresis in shear wave propagation in biotissue, *CRSC Tech. Rpt. CRSC-TR04-03,* NCSU, Raleigh, January, 2004; *SIAM J. Multiscale Modeling and Simulation*, **3** (2005), 395–412.

H.T. Banks and L.K. Potter, Probabilistic methods for addressing uncertainty and variability in biological models: Application to a toxicokinetic model, *CRSC Tech. Rpt. CRSC-TR02-27,* NCSU, Raleigh, September, 2002; *Math. Biosci.,* **192** (2004), 193–225.

S. Bonhoeffer, M. Rembiszewski, G.M. Ortiz, and D.F. Nixon, Risks and benefits of structured antiretroviral drug therapy interruptions in HIV-1 infection, *AIDS,* **14** (2000), 2313-2322.

D. S. Callaway and A. S. Perelson, HIV-1 infection and low steady state viral loads, *Bull. Math. Biol.,* **64** (2001),

29-64.

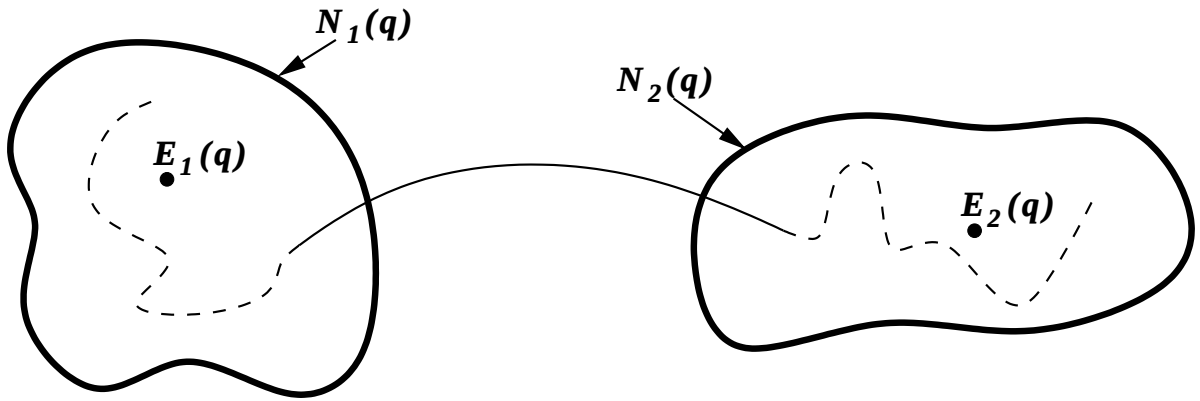


Figure 4: $E_1(q)$: "unhealthy" locally asymptotically stable equilibrium point with its domain of attraction $N_1(q)$; $E_2(q)$: "healthy" locally asymptotically stable equilibrium point with its domain of attraction $N_2(q)$; (- - -) uncontrolled trajectory; (—) controlled trajectory.

Control results with “suboptimal” STI

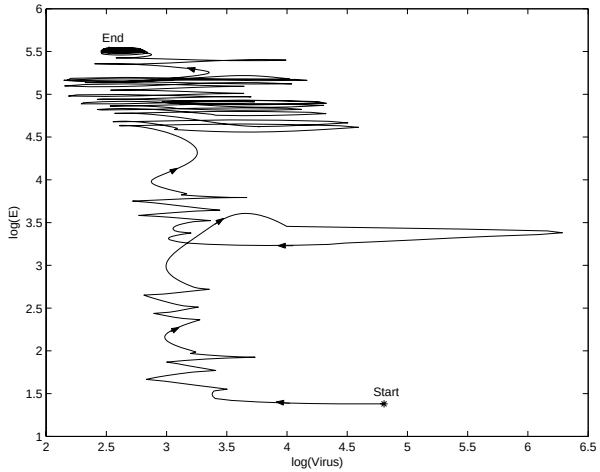


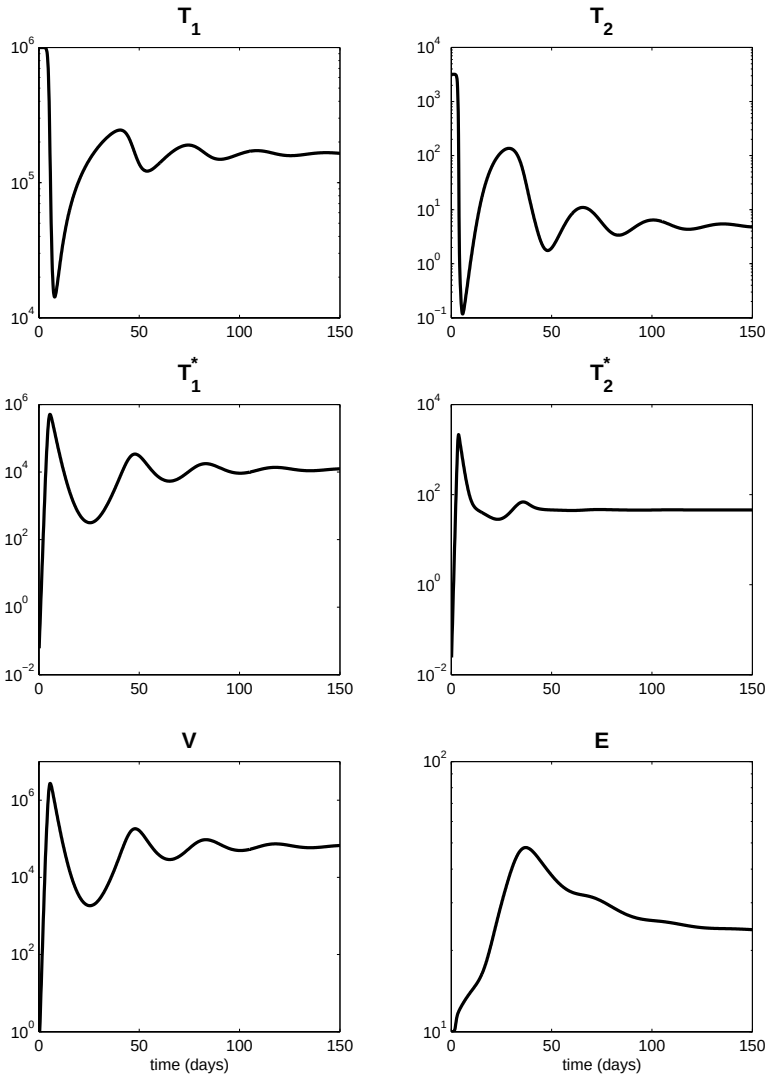
Figure 5: Phase plane plot of virus versus immune effectors (log scale) using “unhealthy” steady state as initial condition ($T_1(0) = 163573$, $T_2(0) = 5$, $T_1^*(0) = 11945$, $T_2^*(0) = 46$, $V(0) = 63919$ and $E(0) = 24$), with suboptimal STI treatment. This plot demonstrates the interplay of immune response and viral load during a dynamic transfer of the model between the “unhealthy” equilibrium (labeled Start) and “healthy” equilibrium (labeled End), around which the solution oscillates.

Typical Model Parameters

$$
\begin{aligned}
\dot{\mathbf{T_1}} &= \lambda_1 - d_1\mathbf{T_1} - (1-\epsilon_1)k_1\mathbf{V}\mathbf{T_1} \\
\dot{\mathbf{T_2}} &= \lambda_2 - d_2\mathbf{T_2} - (1-f\epsilon_1)k_2\mathbf{V}\mathbf{T_2} \\
\dot{\mathbf{T_1^*}} &= (1-\epsilon_1)k_1\mathbf{V}\mathbf{T_1} - \delta\mathbf{T_1^*} - m_1\mathbf{E}\mathbf{T_1^*} \\
\dot{\mathbf{T_2^*}} &= (1-f\epsilon_1)k_2\mathbf{V}\mathbf{T_2} - \delta\mathbf{T_2^*} - m_2\mathbf{E}\mathbf{T_2^*} \\
\dot{\mathbf{V}} &= (1-\epsilon_2)N_T\delta(\mathbf{T_1^*}+\mathbf{T_2^*}) - c\mathbf{V} \\
&\quad - \left[(1-\epsilon_1)\rho_1 k_1\mathbf{T_1} + (1-f\epsilon_1)\rho_2 k_2\mathbf{T_2}\right]\mathbf{V} \\
\dot{\mathbf{E}} &= \lambda_E + \frac{b_E(\mathbf{T_1^*}+\mathbf{T_2^*})}{(\mathbf{T_1^*}+\mathbf{T_2^*})+K_b}\mathbf{E} - \frac{d_E(\mathbf{T_1^*}+\mathbf{T_2^*})}{(\mathbf{T_1^*}+\mathbf{T_2^*})+K_d}\mathbf{E} - \delta_E\mathbf{E}
\end{aligned}
$$

parameter	value	parameter	value	parameter	value
λ_1	$10,000$	k_2	1×10^{-4}	ρ_1	1
d_1	0.01^{**}	δ	0.7^{*}	ρ_2	1
ϵ_1	$\in[0,1)$	m_1	1.0×10^{-5}	λ_E	1
ϵ_2	$\in[0,1)$	m_2	1.0×10^{-5}	b_E	0.3
k_1	8.0×10^{-7}	N_T	100^{*}	K_b	100
λ_2	31.98	c	13^{*}	d_E	0.25
d_2	0.01^{**}			K_d	500
f	$0.34\,(\in[0,1])$			δ_E	0.1^{*}

Sample Model Solution: Early Infection Scenario



ODE (Initial Value Problem) Solver

System can be stiff, especially when computing sensitivity matrices. In all cases we use stiff solvers based on numerical differentiation formulas (typically BDFs of order 1 to 5).)

- Matlab's `ODE15s` (interpreted): used for development and most small experiments
- CASC at LLNL's `LSODE` (widely used Fortran code): used when MCMC group needed fast model solutions in Matlab, wrote a Matlab MEX gateway to interface directly with Matlab
- CASC at LLNL's `CVODE` (based on VODE which is similar to LSODE, is integrated into SUNDIALS package): used to quickly generate model solutions (in parallel) for virtual patients and at nodes in estimating probability distributions. (I wanted an excuse to learn MPI in C.)

Optimization Algorithms

All in Matlab:

- Gauss-Newton `gaussn.m` and Levenberg-Marquardt `levmar_old.m` algorithms supplied by C.T. Kelley (NCSU) – both use line searches.
- Matlab's `quadprog`: handles our bound constraints and linear equality constraints, solves with SQP active set method.
- `DIRECT` as implemented by Dan Finkel (NCSU): Hyper-rectangle sampling algorithm that iteratively samples the objective function at rectangle centers and selectively divides potentially optimal rectangles. For use on bound constrained problems.
- Matlab's `lsqnonlin`: specifically for solving nonlinear least squares problems – handles bound constraints and includes option of difference or analytic derivatives. subspace trust region method and is based on an interior-reflective Newton method.

Simulated Data Generation Detail

- Have clinical data from patients corresponding to model states

$$\mathbf{x} = \begin{bmatrix} T_1 + T_1^* \\ V \\ E \end{bmatrix} \quad \begin{matrix} \text{(CD4 T-cells)} \\ \text{(free virions)} \\ \text{(CTL response)} \end{matrix}$$

- Verify methods on simulated data generated for times $t^i, i = 1, \ldots, N$:

$$y_s^i = x_s(t^i; \mathbf{q}) + \epsilon_s^i$$

and state $s = 1, 2, 3$, by taking $y_m^i = \exp\{z_s^i\}$ where

$$z_s^i \sim \mathcal{N}\left(\log x_s^i - \log(\sigma_s^2 + 1)/2, \log(\sigma_s^2 + 1)\right), \text{ so}$$

⋆ **Mean response** given by ODE model: $E(y_s^i) = x_s^i(\mathbf{q})$

⋆ **Variance model:** $Var(y_s^i) = \sigma_s^2 \left(x_s^i(\mathbf{q})\right)^2$

(constant coeff. of variation model – typical for blood draw assays)

Theory for General Inverse Problem

$$P^* = \arg\min_{P \in \mathcal{Q}} J(P) = \frac{1}{N_P} \sum_{j=1}^{N_P} \frac{1}{N_j} \sum_{i=1}^{N_j} \left| \mathcal{E}\left[\mathbf{x}(t^{ij}; \mathbf{q}) | P(\mathbf{q}) \right] - \mathbf{y}^{ij} \right|^2$$

- Problem assumes **model parameters** of interest $\mathbf{q}$ for each patient **are realizations of a random variable with probability distribution** P.
- P belongs to a probability space $\mathcal{Q} \subset \mathcal{P}(Q)$, where $\mathcal{P}(Q)$ denotes all probability distributions on the admissible parameter space Q. For example $\mathcal{Q}$ could be the set
  - ⋆ of normal distributions over Q, parameterized by $\mathcal{N}(\mu, \sigma^2)$
  - ⋆ $\mathcal{P}^M = \{P = \sum_{k=0}^{M} p_k \delta_{q_k}\}$, where $q_k \in Q, p_k \geq 0, \sum_{k=0}^{M} p_k = 1$ (point masses – for inverse problem well-posedness results and examples in this context: Banks and Bihari (2001), Banks and Potter (2003))

Inverse Problem: Supporting Theory

Inverse problem: Minimize $J(P)$ over $P \in \mathcal{Q} \subset \mathcal{P}(Q)$.

Banks and Bihari (2001) review: **continuity of $P \to J(P)$** in the Prohorov ρ metric, with **compactness of $\mathcal{P}(Q)$** in the ρ metric (guaranteed by Q compact) is sufficient to establish a minimizer's existence.

Inverse Problem: Supporting Theory

Inverse problem: Minimize $J(P)$ over $P \in \mathcal{Q} \subset \mathcal{P}(Q)$.

Banks and Bihari (2001) review: **continuity of $P \to J(P)$** in the Prohorov ρ metric, with **compactness of $\mathcal{P}(Q)$** in the ρ metric (guaranteed by Q compact) is sufficient to establish a minimizer's existence.

If **distribution P characterized by its density**: For $\mathcal{F}$ a weakly compact subset of $L^2(Q)$, Q compact, define

$$\mathcal{P}_{\mathcal{F}}(Q) := \{P \in \mathcal{P}(Q) | P' = f, f \in \mathcal{F}\}$$

Banks and Pinter (2004) showed $\mathcal{P}_{\mathcal{F}}(Q)$ is a compact subset of $\mathcal{P}(Q_)$ in the Prohorov metric. Hypotheses of Banks and Bihari (2001) are satisfied and well-posedness of the inverse problem minimizing over $P \in \mathcal{P}_{\mathcal{F}}$ (distributions with densities) follows.

Quadratic Programming Problem Coefficients

$$\int_Q \mathbf{x}(t^i;q)f(q)dq \approx \sum_{k=0}^{N_S-1} \frac{\Delta q_k}{2}\left(\mathbf{x}^i(q_k)d_k + \mathbf{x}^i(q_{k+1})d_{k+1}\right)$$

So the minimization problem reduces to a **quadratic programming problem** in coefficients $d_k, k = 0, \ldots, N_S$:

$$(\dagger) \quad \boxed{J(\mathbf{d}) = \mathbf{d}^T A\mathbf{d} + 2\mathbf{b}^T\mathbf{d} + c}$$

where, for $x^i_{s,k} = x_s(t_i;q_k)$,

$$\begin{aligned}
A_{kl} &= \frac{1}{4N\,N_P}\sum_{i=1}^{N}\sum_{j=1}^{N_P}\sum_{s=1}^{3}\left(\Delta q_k x^i_{s,k}(q_k) + \Delta q_{k-1}x^i_{s,k-1}\right)\left(\Delta q_l x^i_{s,l} + \Delta q_{l-1}x^i_{s,l-1}\right)\\
b_k &= -\frac{1}{2N\,N_P}\sum_{i=1}^{N}\sum_{j=1}^{N_P}\sum_{s=1}^{3} y^{ij}_s\left(\Delta q_k x^i_{s,k} + \Delta q_{k-1}x^i_{s,k-1}\right)\\
c &= \frac{1}{N\,N_P}\sum_{i=1}^{N}\sum_{j=1}^{N_P}\sum_{s=1}^{3}\left(y^{ij}_s\right)^2.
\end{aligned}$$

Estimated Parameters – On Treatment Data

parameter	Patient 26	Patient 2
d_1	8.9252e-03	1.3288e-02
k_1	6.9104e-12	4.0266e-09
d_2	4.3851e-02	4.5115e-03
δ	4.6416e-04	2.3462e-04
k_2	5.9948e-02	5.0548e-02
N_T	5.9948e+01	8.4319e+01
c	1.1860e+01	9.4473e+00

Initial conditions: T_1 = 0.8*(Initial T-cell measurement), T_1^* = 0.2*(initial T-cell measurement), V = initial viral load measurement. Other states fixed at early infection scenario initial conditions.

Estimated Parameters – STI Data

parameter	Patient 25	Patient 5
T_1	3.9354e+02	1.1304e+03
T_2	2.7393e+02	2.6769e+00
T_1^*	6.0904e+01	2.2896e+01
T_2^*	4.6714e+02	1.4598e+01
V	1.4065e+06	1.0000e+06
E	1.1365e+00	1.1365e+02

parameter	Patient 25	Patient 5
λ_1	9.8624e+03	7.5516e+03
d_1	2.0214e-02	1.2110e-02
ϵ	5.9414e-01	8.7025e-01
k_1	6.5614e-08	1.2874e-07
λ_2	3.1623e+02	3.1623e+01
d_2	1.0000e-02	5.5048e-02
f	6.6834e+00	6.1423e+00
k_2	1.2833e-05	1.7440e-04
δ	9.4523e-02	2.8256e-01
m_1	4.6416e-05	1.0890e-06
m_2	4.2622e-06	4.6416e-05
β	1.0000e+00	1.0000e+00
N_T	9.7957e+01	1.0773e+02
c	1.0462e+01	8.6989e+00
b_E	1.0000e-01	1.0000e-01
K_b	3.9137e+02	1.1860e+00
d_E	2.1544e-02	2.1544e-02
K_d	2.3462e+00	5.9948e+03
δ_E	5.9948e-02	1.0000e-01
λ_E	5.5048e+00	3.9137e+00