Stochastic Optimization Without Integration

João P. Hespanha

Center for Control
Dynamical Systems and Computation

UC SANTA BARBARA
engineering

This talk..

Problem addressed:

$$\min_u \mathrm{E}_d \left[f(u,d) \right]$$

- no closed form for expected value
- willing to sacrifice accuracy for speed
- error bounds

Talk outline:

1. Motivating problems
2. Additive/multiplicative bounds for expected values
3. Bounds based on Laplace method for integration

In collaboration with:
Raphael Chinchilla, Murat Erdal, Dr. Guosong Yang (UCSB)
Prof. Ramon Costa (Federal Univ. Rio Janeiro, Brazil)
Prof. Kevin Plaxco (UCSB)

Stochastic Optimal Control

Dubins vehicle (discrete-time)

angular velocity (control)

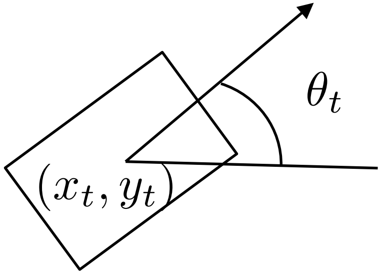
$$x_{t+1} = x_t + v\Big(\cos\theta_t - \frac{u_t}{2}\sin\theta_t\Big) + d_t$$

Cartesian disturbances

$$y_{t+1} = y_t + v\Big(\sin\theta_t + \frac{u_t}{2}\cos\theta_t\Big) + w_t$$

$$\theta_{t+1} = \theta_t + u_t + \nu_t$$

rotational disturbance



$$\min_{u_1,\dots,u_T} \mathrm{E}\Big[\sum_{t=1}^{T} x_t^2 + y_t^2 + u_t^2\Big]$$

- not possible to compute expectation in closed form, even if d_t, w_t, v_t multivariable Gaussian/von Misses

yet…

- need to solve fast for receding-horizon / Model Predictive Control (MPC)

Estimation

Dynamical system:

$$\begin{aligned} x_{t+1} &= f(x_t;\theta) + d_t \\ y_t &= g(x_t;\theta) + w_t \end{aligned}$$

state disturbances

measurement noise

unknown parameter

Maximum likelihood estimation:

prob. density function (pdf) of output

$$\hat{\theta}_{\mathrm{MLE}} = \arg\max_{\theta} p(y_1, \ldots, y_K; \theta)$$

Estimation

Dynamical system:

$$x_{t+1} = f(x_t;\theta) + d_t$$
$$y_t = g(x_t;\theta) + w_t$$

state disturbances

measurement noise

unknown parameter

Maximum likelihood estimation:

prob. density function (pdf) of output

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} p(y_1,\ldots,y_K;\theta)$$

- output pdf is generally difficult to compute, but
- conditional output pdf given state is typically very easy to compute
- can write MLE as

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} \mathrm{E}\left[\overbrace{p(y_1,\ldots,y_K \mid x_1,\ldots,x_K;\theta)}\right]$$

law of total expectation

expectation w.r.t. x_t

- generally no closed-form expression for expectation

yet,…

- need to solve fast to compute θ-dependent a-posterior estimate of state, given measurements for output-feedback control

Experiment Design

Dynamical system:

$$\begin{aligned} x_{t+1} &= A(\theta)x_t + B(\theta)u_t + d_t \\ y_t &= C(\theta)x_t + w_t \end{aligned}$$

state disturbances

measurement noise

unknown parameter

Goal: select $u_1, \ldots, u_T$ to minimize error for estimator $\hat{\theta}(y_1, y_2, \ldots, y_T)$

Experiment Design

Dynamical system:

$$\begin{aligned} x_{t+1} &= A(\theta)x_t + B(\theta)u_t + d_t \\ y_t &= C(\theta)x_t + w_t \end{aligned}$$

state disturbances

measurement noise

unknown parameter

Goal: select $u_1, \ldots, u_T$ to minimize error for estimator $\hat{\theta}(y_1, y_2, \ldots, y_T)$

Assuming unbiased estimator that achieves Cramer-Rao lower bound

$$\mathrm{E}\left[(\hat{\theta}-\theta)(\hat{\theta}-\theta)'\right] = \mathrm{FIM}(\theta)^{-1} \qquad \mathrm{FIM}(\theta) := -\,\mathrm{E}\left[\frac{\partial^2 \log p(y_1, \ldots, y_T; \theta)}{\partial \theta^2}\right]$$

Fisher information matrix

w.r.t noise/disturbances

Hessian matrix of measurements pdf

A-optimality:

$$\min_{u_1,\ldots,u_T} \mathrm{E}\left[\operatorname{trace} \mathrm{FIM}(\theta)^{-1}\right]$$

w.r.t θ

minimizes parameter estimates MSE

D-optimality:

$$\min_{u_1,\ldots,u_T} \mathrm{E}\left[\log\det FIM(\theta)^{-1}\right]$$

w.r.t θ

minimizes volume of parameters estimates confidence ellipsoids

Monte Carlo Methods

Monte Carlo Integration:

$$\mathrm{E}\left[f(d)\right] \approx \frac{1}{K}\sum_{k=1}^{K} f(d_k) \qquad \text{error} \quad \propto O\left(\frac{\sigma}{K^{\frac{1}{2}}}\right)$$

random variable

indep. identically distributed (iid) samples of d

- central limit theorem
- dimension-independent

In stochastic optimization, leads to Sample Average Approximation (SAA):

$$\min_u \mathrm{E}\left[f(u,d)\right] \approx \min_u \frac{1}{K}\sum_{k=1}^{K} f(u,d_k)$$

random variable

[Shalev-Shwartz at al. 2010]

$$\text{error} \quad \propto \begin{cases} O\left(\frac{1}{K}\right) & \text{strongly convex and Lipschitz} \\ O\left(\frac{1}{K^{\frac{1}{2}}}\right) & \text{Lipschitz} \end{cases}$$

[Kleywegt et al, 2001

e.g., stochastic gradient descent:

$$u(i+1) = u(i) - \gamma(i)\frac{1}{K}\sum_{k=1}^{K} \nabla f_u\big(u(i), d_k\big)$$

step size

gradient of empirical average

Monte Carlo Methods

MC-based methods

- can be made arbitrarily accurate by making $K \to \infty$
- generally slow (need large K, # iterations)
- often computationally expensive to sample from desired distribution (especially for conditional distribution given measurements, typically requiring Markov Chain Monte Carlos method, e.g., Metropolis Hastings or Gibbs sampling)

Our goal…

- Trade accuracy for speed
- If possible, get bounds on error

[Shalev-Shwartz at al. 2010]

random variable

$$\text{error} \quad \propto \begin{cases} O\left(\frac{1}{K}\right) & \text{strongly convex and Lipschitz} \\ O\left(\frac{1}{K^{\frac{1}{2}}}\right) & \text{Lipschitz} \end{cases}$$

[Kleywegt et al, 2001

e.g., stochastic gradient descent:

$$u(i+1) = u(i) - \gamma(i) \frac{1}{K} \sum_{k=1}^{K} \nabla f_u\big(u(i), d_k\big)$$

step size

gradient of empirical average

Outline

Problem addressed:

$$\min_u \mathrm{E}_d\left[f(u,d)\right]$$

- no closed form for expected value
- willing to sacrifice accuracy for speed
- error bounds

Talk outline:

1. Motivating problems
2. Additive/multiplicative bounds for expected values
3. Bounds based on Laplace method for integration

In collaboration with:
Raphael Chinchilla, Murat Erdal, Dr. Guosong Yang (UCSB)
Prof. Ramon Costa (Federal Univ. Rio Janeiro, Brazil)
Prof. Kevin Plaxco (UCSB)

Coarse Bounds on Expectations

Suppose d is random variable taking values on $\mathcal{D}$, and

$$f_{\min} \leqslant f(d) \leqslant f_{\max} \qquad \text{with probability 1}$$

by monotonicity

$$f_{\min} \leqslant \mathrm{E}\left[f(d)\right] \leqslant f_{\max}$$

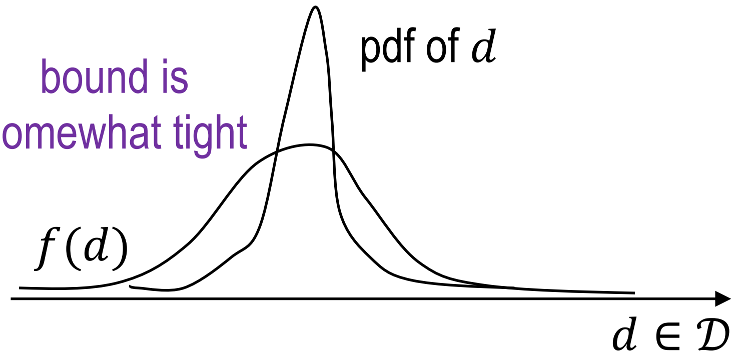
tightest such (upper) bound

can discount zero-measure sets

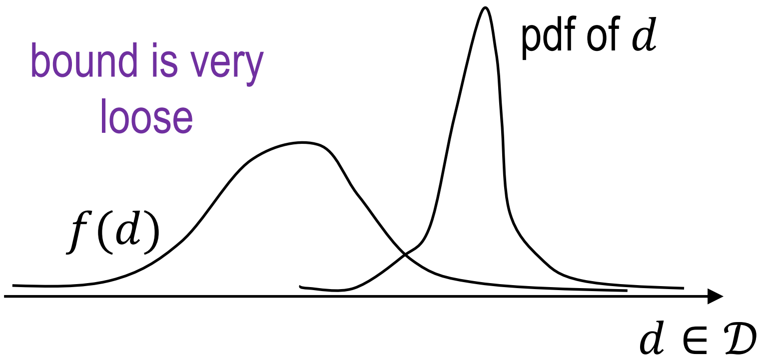
$$\mathrm{E}\left[f(d)\right] \leqslant \underbrace{\operatorname*{ess\,sup}_{d\in\mathcal{D}} f(d)}$$

values of d with positive but small probability increase supremum but have little impact on expectation

bound is somewhat tight



bound is very loose



Coarse Bounds on Expectations

Suppose d is random variable taking values on $\mathcal{D}$, and

$$f_{\min} \leqslant f(d) \leqslant f_{\max} \qquad \text{with probability 1}$$

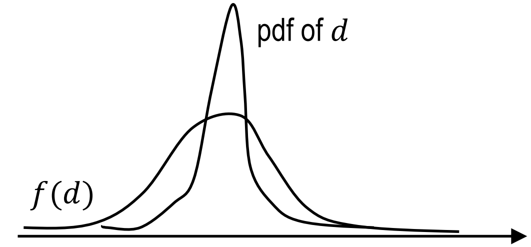
by monotonicity

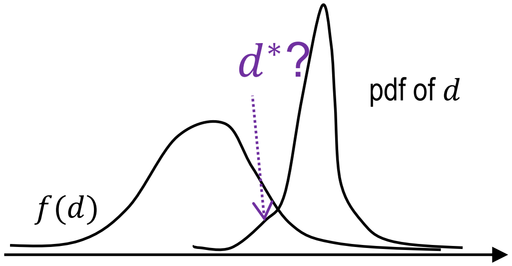
$$f_{\min} \leqslant \mathrm{E}\left[f(d)\right] \leqslant f_{\max}$$

tightest such (upper) bound

can discount zero-measure sets

$$\mathrm{E}\left[f(d)\right] \leqslant \operatorname*{ess\,sup}_{d \in \mathcal{D}} f(d)$$





“naïf” idea:

$$\mathrm{E}[f(d)] \leqslant f(d^*)$$

prob. density function (pdf) of d

$$d^* = \arg\max_{d \in \mathcal{D}} \Big(f(d) + \log p(d) \Big)$$

max will not pick value for d with low probability
(very negative log-pdf)

A Precise Bound

Additive bound: For every $\varepsilon \geq 0$ (and assuming all expectations are finite),

prob. density function (pdf) of d

$$\operatorname*{ess\,inf}_{d\in\mathcal{D}} \Big(f(d) - \epsilon \log p(d)\Big) - \epsilon H_d \leqslant \mathrm{E}[f(d)] \leqslant \operatorname*{ess\,sup}_{d\in\mathcal{D}} \Big(f(d) + \epsilon \log p(d)\Big) + \epsilon H_d$$

inf at value for which
$f(d)$ is small
but pdf not very small

sup at value for which
$f(d)$ is large
but pdf not very small

differential entropy of d

$$H_d := \mathrm{E}\big[-\log p(d)\big]$$

A Precise Bound

Additive bound: For every $\varepsilon \geq 0$ (and assuming all expectations are finite),

prob. density function (pdf) of d

$$\operatorname*{ess\,inf}_{d\in\mathcal{D}} \Big(f(d) - \epsilon \log p(d)\Big) - \epsilon H_d \leqslant \mathrm{E}[f(d)] \leqslant \operatorname*{ess\,sup}_{d\in\mathcal{D}} \Big(f(d) + \epsilon \log p(d)\Big) + \epsilon H_d$$

inf at value for which
$f(d)$ is small
but pdf not very small

sup at value for which
$f(d)$ is large
but pdf not very small

differential entropy of d
$H_d := \mathrm{E}\big[-\log p(d)\big]$

Why? (upper bound)

$$\begin{aligned}
\mathrm{E}\big[f(d)\big] &= \mathrm{E}\big[f(d) + \epsilon \log p(d) - \epsilon \log p(d)\big] \\
&= \underbrace{E\big[f(d) + \epsilon \log p(d)\big]}_{\text{use “trivial” bound on this term}} + \epsilon H_d
\end{aligned}$$

$$E\big[f(d) + \epsilon \log p(d)\big] \leqslant \operatorname*{ess\,sup}_{d\in\mathcal{D}} f(d) + \epsilon \log p(d)$$

bound typically tighter as probability mass
concentrated around large values of
$f(d) + \epsilon \log p(d)$

More Bounds

Additive bound: For every $\varepsilon \geq 0$ (and assuming all expectations are finite),

prob. density function (pdf) of d

$$\operatorname*{ess\,inf}_{d\in\mathcal{D}} \Big(f(d) - \epsilon \log p(d) \Big) - \epsilon H_d \leqslant \mathrm{E}[f(d)] \leqslant \operatorname*{ess\,sup}_{d\in\mathcal{D}} \Big(f(d) + \epsilon \log p(d) \Big) + \epsilon H_d$$

differential entropy of d

$$H_d \coloneqq \mathrm{E}\big[-\log p(d) \big]$$

Multiplicative bound: For every $\varepsilon > 0$ (and assuming all expectations are finite),

$$\operatorname*{ess\,inf}_{d\in\mathcal{D}} \Big(f(d) p(d)^{-\epsilon} \Big) I_d(-\epsilon) \leqslant \mathrm{E}[f(d)] \leqslant \operatorname*{ess\,sup}_{d\in\mathcal{D}} \Big(f(d) p(d)^{\epsilon} \Big) I_d(\epsilon)$$

$$I_d(\epsilon) \coloneqq \mathrm{E}\big[p(d)^{-\epsilon} \big]$$

More generally: For every function $\alpha\colon \mathcal{D} \to \mathbb{R}$

Any group operation that is

1. right ordered $a \leqslant b \Leftrightarrow a \oplus c \leqslant b \oplus c$
2. E-distributive $a \oplus \mathrm{E}[z] = \mathrm{E}[a \oplus z]$

$$\mathrm{E}[f(d)] \leqslant \operatorname*{ess\,sup}_{d\in\mathcal{D}} \Big(f(d) \oplus \alpha(d) \Big) \oplus A_d$$

$$A_d \coloneqq \mathrm{E}\big[\neg\alpha(d) \big]$$

for tight bound: pick $\alpha, \oplus$ so that prob. mass concentrated around max. of $f(d) \oplus \alpha(d)$

Application to Optimization

$$J^* := \min_u \left\{ \mathrm{E}\left[f(d,u)\right] : \mathrm{E}\left[g(d,u)\right] \leqslant 0 \right\}$$

replacing by upper bound will find “conservative/pessimistic” solution

replacing by upper bound will guarantee feasibility

or…
replacing both by lower bounds will find “optimistic” solution

E.g., using additive upper bounds:

$$J^{\perp} := \min_u \left\{ \operatorname*{ess\,sup}_{d \in \mathcal{D}} \Big(f(d,u) + \epsilon \log p(d) \Big) : g(\bar{d},u) + \epsilon \log p(\bar{d}) \leqslant 0, \forall \bar{d} \in \mathcal{D} \right\}$$

1. arg-min guaranteed to be feasible for original problem & perform no worst than $J^{\perp}$
2. cost $J^{\perp}$ is an upper bound for J^*
3. **replaced expectation/integration by optimization ➪ min-max problem**
4. lower bound for J^* computable using lower bounds for expected values

Application to Output-Feedback MPC

state disturbances

$$x_{t+1} = f(x_t, u_t; \theta) + d_t$$
$$y_t = g(x_t; \theta) + w_t$$

measurement noise

unknown parameter

$$\min_{u_0,\ldots,u_{T-1}} \mathrm{E}\left[\sum_{t=0}^{T} \ell(x_t, u_t) \,\middle|\, y_{-K}, \ldots, y_0\right]$$

future controls

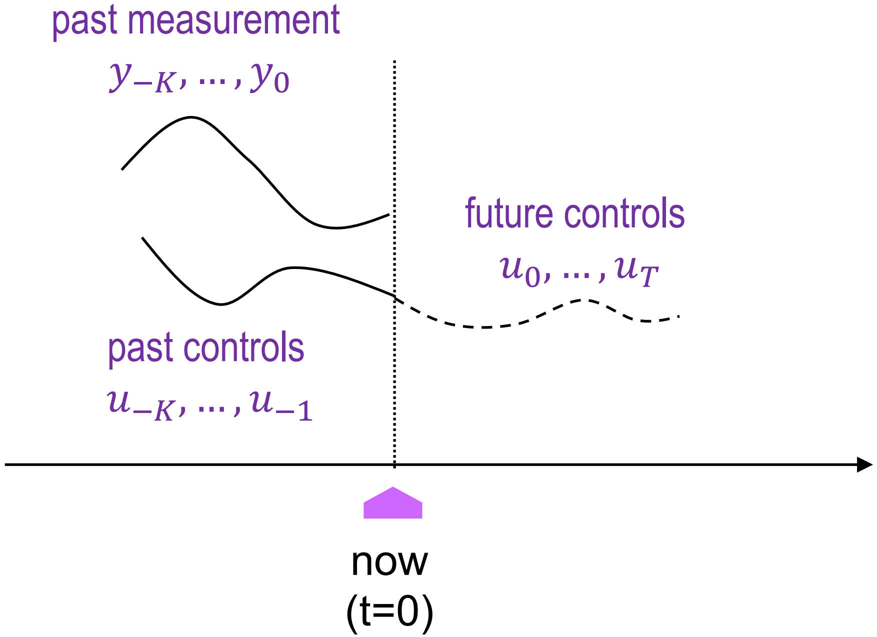
running cost, over future horizon

past measurements

a-posteriori (conditional) expectation with respect to:

1. unknown parameter θ
2. past noise $w_{-K}, \ldots, w_0$
3. past and future disturbances $d_{-K}, \ldots, d_{T-1}$

given past measurement $y_{-K}, \ldots, y_0$



In Model Predictive Control (MPC), optimization repeated at each time step under a receding horizon

Output-Feedback MPC

$$\begin{aligned} x_{t+1} &= f(x_t, u_t; \theta) + d_t \\ y_t &= g(x_t; \theta) + w_t \end{aligned}$$

$$\min_{u_0, \dots, u_{T-1}} \mathrm{E}\left[\sum_{t=0}^{T} \ell(x_t, u_t) \;\middle|\; y_{-K}, \dots, y_0 \right]$$

Observation: assuming x_{-K}, d_t, w_t, θ independent, joint pdf of x_{-K}, d_t, y_t, θ is easy to write

pdfs of w_t, d_t, x_{-K}, θ

$$\left(\prod_{t=-K}^{0} p_w\big(y_t - g(x_t; \theta)\big) \right) \left(\prod_{t=-K}^{T-1} p_d\big(x_{t+1} - f(x_t, u_t; \theta)\big) \right) p(x_{-K}) p(\theta)$$

Why? Proof by induction to compute

$$p\big(x_{t+1}, y_t \mid d_{-K}, \dots, d_t, y_{-K}, \dots, y_{t-1}, x_{-K}, \theta\big)$$

starting at $t = -K$ & using Bayes rule for induction step

Therefore,…

- closed-form solution for a-posteriori distribution (modulo normalization by pdf of $y_{-K}, \dots, y_0$)
- but not easy to draw samples from a-posteriori distribution (need Metropolis-Hastings or Gibbs sampling)

Output-Feedback MPC

$$\begin{aligned} x_{t+1} &= f(x_t, u_t; \theta) + d_t \\ y_t &= g(x_t; \theta) + w_t \end{aligned}$$

$$\min_{u_0,\dots,u_{T-1}} \mathrm{E}\Big[\sum_{t=0}^{T} \ell(x_t, u_t) \;\Big|\; y_{-K}, \dots, y_0\Big]$$

Can be restated as: $J^* = \min_U \mathrm{E}\big[V(U,D) \mid Y\big]$

$$\begin{aligned} U &\coloneqq \{u_0, \dots, u_{T-1}\} \\ D &\coloneqq \{x_{-K}, d_{-K}, \dots, d_{T-1}, \theta\} \\ Y &\coloneqq \{y_{-K}, \dots, y_0\} \end{aligned}$$

Using additive bound:

$$J^{\perp} = \min_U \Big(\operatorname*{ess\,sup}_D V(U,D) + \epsilon \log p(D|Y) \Big) + \epsilon H_D$$

$$H_D \coloneqq \mathrm{E}\big[-\log p(D)|Y\big]$$

$$\log p(D|Y) = \log(D,Y) - \log P(Y)$$

conditional pdf
of D, given Y

joint pdf of
D, given Y
⇩
closed-form
solution

marginal
pdf of Y
⇩
indep. of U

diff. entropy of D
⇩
indep. of U

Output-Feedback MPC

$$\begin{aligned} x_{t+1} &= f(x_t, u_t; \theta) + d_t \\ y_t &= g(x_t; \theta) + w_t \end{aligned}$$

$$\min_{u_0,\dots,u_{T-1}} \mathrm{E}\left[\sum_{t=0}^{T} \ell(x_t, u_t) \;\middle|\; y_{-K}, \dots, y_0 \right]$$

Can be restated as: $J^* = \min_U \mathrm{E}\big[V(U,D) \mid Y\big]$

$$\begin{aligned} U &\coloneqq \{u_0, \dots, u_{T-1}\} \\ D &\coloneqq \{d_{-K}, \dots, d_{T-1}, x_{-K}, \theta\} \\ Y &\coloneqq \{y_{-K}, \dots, y_0\} \end{aligned}$$

Using additive bound:

$$U^{\perp} = \arg\min_U \left(\operatorname*{ess\,sup}_D \sum_{t=1}^{T} \ell(x_t, u_t) + \epsilon \log p(D, Y) \right)$$

$$\left(\sum_{t=-K}^{0} \log p_w\big(y_t - g(x_t;\theta)\big) \right) + \left(\sum_{t=-K}^{T-1} \log p_d\big(x_{t+1} - f(x_t, u_t; \theta)\big) \right) + \log p(x_{-K}) + \log p(\theta)$$

- all optimization terms computed in closed form without need for integration
- stochastic optimization replaced by min-max problem
- one-to-one map from d_t's to x_t's, so we can regard x_t's as the optimization variables (computationally much better for 2nd order/Newton methods because of Hessian sparsity)
- this and all optimizations in remainder of paper solved using TensCalc toolbox

LQG example

$$x_{t+1} = A(\theta)x_t + B(\theta)u_t + d_t$$
$$y_t = x_t + w_t$$

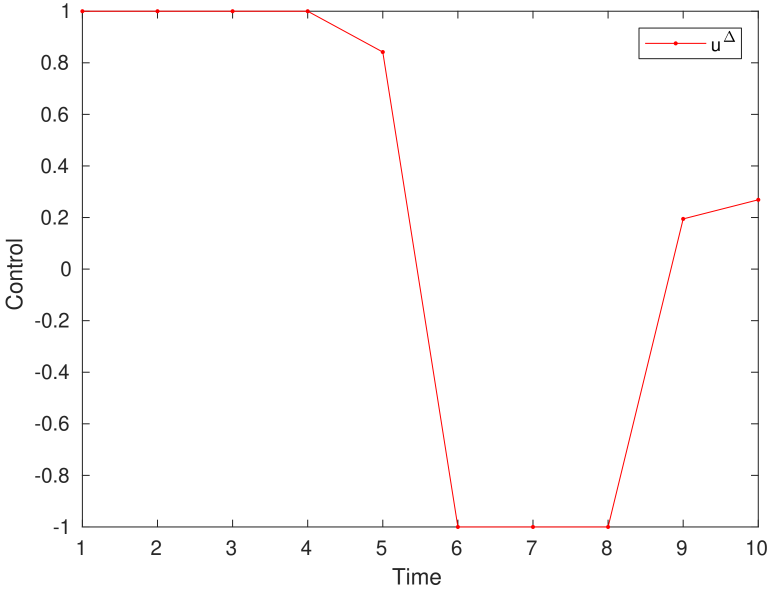
$$\min_{|u_t|\leqslant 1} \mathrm{E}\left[\sum_{t=0}^{T} \|x_t\|^2 + u_t^2 \;\middle|\; y_{-K}, \ldots, y_0\right]$$

$$A(\theta) = \begin{bmatrix} \theta_1 & \theta_2 & 0 \\ 0 & \theta_3 & \theta_4 \\ 0 & 0 & \theta_5 \end{bmatrix} \quad B(\theta) = \begin{bmatrix} 0 \\ 0 \\ \theta_6 \end{bmatrix}$$

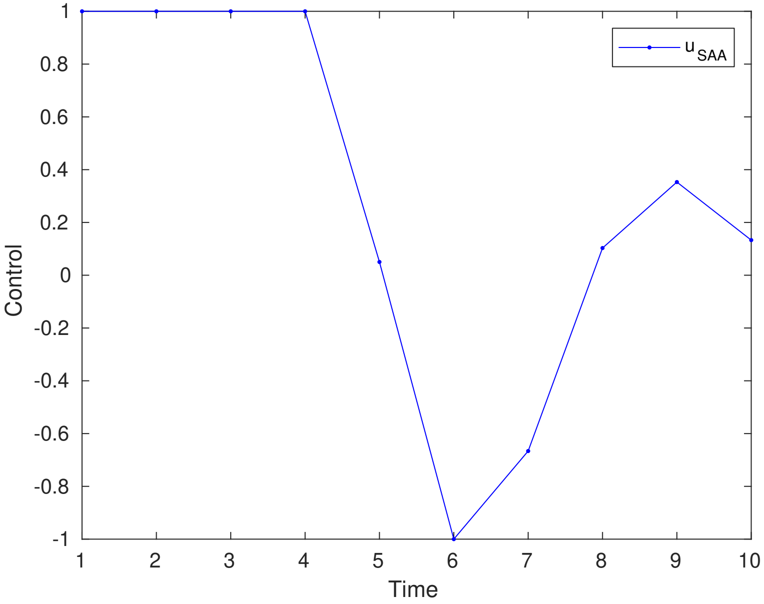
independent Gaussian parameters with

$$\mathrm{E}[\theta_i] = 1, \;\mathrm{Var}[\theta_i] = .5^2$$

$$T = 10, K = 20$$



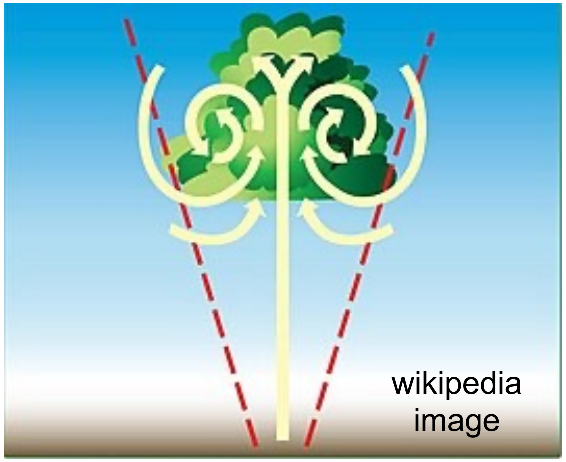
optimal control based on upper bound
(criterion value 1.79e3, computation sub-second)



optimal control based SSA, with Gibbs sampling
(criterion value 1.62e3, computation minutes)

Experiment Design

thermal column of raising air



wikipedia image

vertical wind speed at position x

$$v(x) = v_0 e^{-\gamma \|x - p_0\|^2} + w$$

thermal parameters

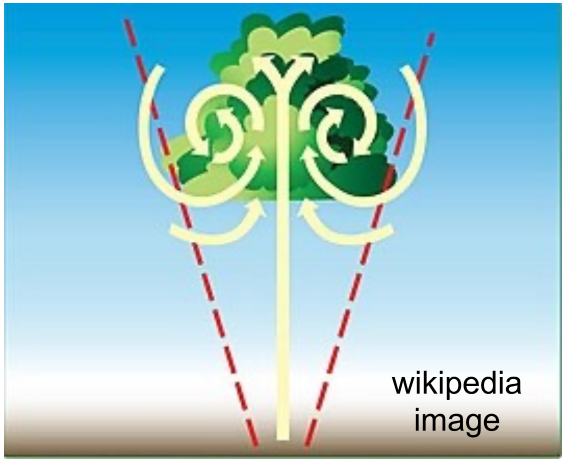
stochastic variability

Q: What is the best trajectory x_t to estimate thermal parameters $\theta = [v_0, \gamma, p_0]$, from point measurements of vertical wind speed?

[Problem proposed by Prof. Isaac Kaminer]

Experiment Design

thermal column of raising air



wikipedia image

vertical wind speed at position x

$$v(x) = v_0 e^{-\gamma\|x-p_0\|^2} + w$$

thermal parameters

stochastic variability

Q: What is the best trajectory x_t to estimate thermal parameters $\theta = [v_0, \gamma, p_0]$, from point measurements of vertical wind speed?

D-optimality:

$$\min_{\substack{x_1,\dots,x_T:\\ \|x_{i+1}-x_i\|\leq\ell}} \mathrm{E}\left[\log\det FIM(\theta)^{-1}\right]$$

Fisher Information Matrix for measurements with zero-mean Gaussian noise

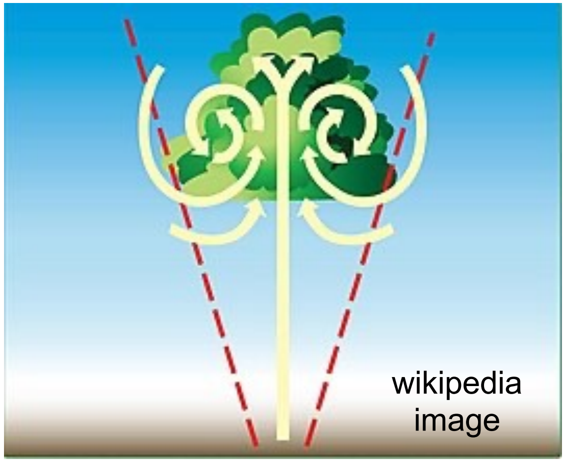
$$\mathrm{FIM}(\theta) = \frac{1}{\sigma^2}\sum_{t=1}^{T} e^{-\gamma\|x_t-p_0\|^2}\begin{bmatrix}-1\\ v_0\|x_t-p_0\|^2\\ v_0\gamma\|x_t-p_0\|^2\end{bmatrix}\begin{bmatrix}-1 & v_0\|x_t-p_0\|^2 & v_0\gamma\|x_t-p_0\|^2\end{bmatrix}$$

noise variance

sensor position at time t

Experiment Design

thermal column of raising air


wikipedia image

vertical wind speed at position x

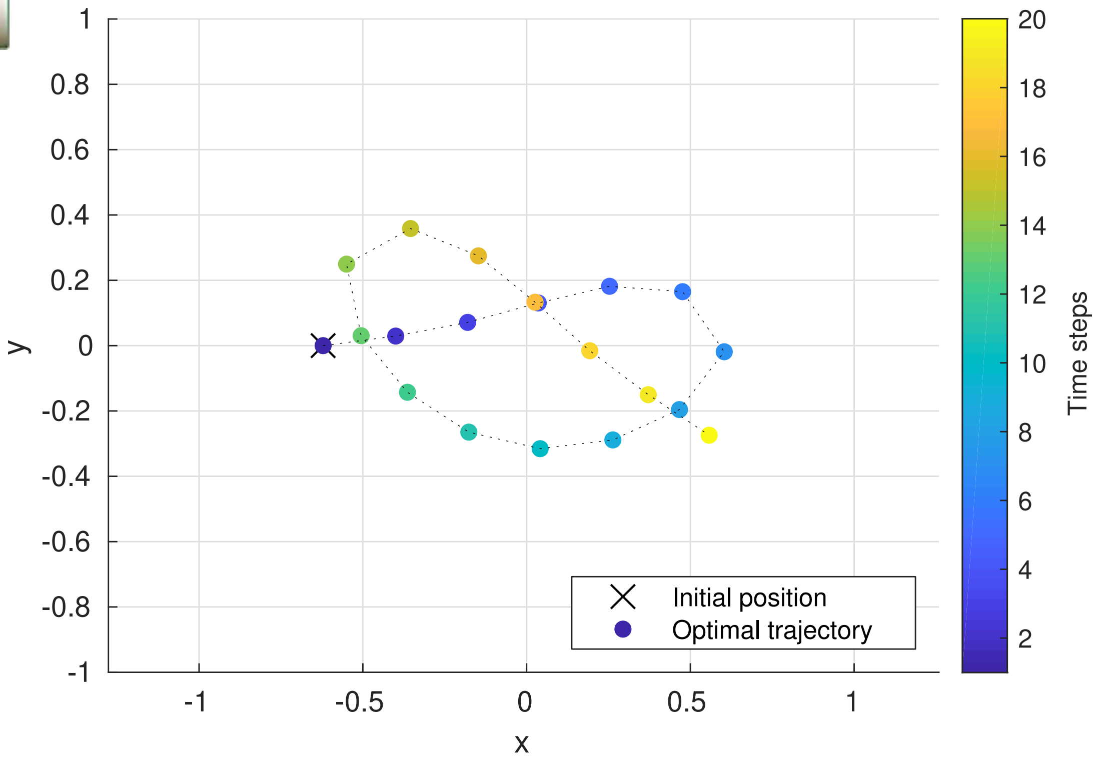
$$v(x) = v_0 e^{-\gamma \|x - p_0\|^2} + w$$

thermal parameters

stochastic variability

Q: What is the best trajectory x_t to estimate thermal parameters v_0, γ, p_0 vertical wind speed measurements?

A: a knot ;-)



Outline

Problem addressed:

$$\min_u \mathrm{E}_d\left[f(u,d)\right]$$

- no closed form for expected value
- willing to sacrifice accuracy for speed
- error bounds

Talk outline:

1. Motivating problems
2. Additive/multiplicative bounds for expected values
3. Bounds based on Laplace method for integration

In collaboration with:
Raphael Chinchilla, Murat Erdal, Dr. Guosong Yang (UCSB)
Prof. Ramon Costa (Federal Univ. Rio Janeiro, Brazil)
Prof. Kevin Plaxco (UCSB)

Laplace Integration Method

$$\int_a^b e^{f(x)} dx$$ smooth $f(x)$ with a maximum at x_0 and $f''(x_0) < 0$

Taylor series of $f(x)$ around x_0:

$$f(x) = f(x_0) + \underbrace{f'(x_0)(x - x_0)}_{\text{0 at maximum}} + \underbrace{\frac{f''(x_0)}{2}}_{< 0}(x - x_0)^2 + O\big((x - x_0)^3\big)$$

Assuming 3rd and higher order terms negligible:

$$\int_a^b e^{f(x)} \approx \int_a^b e^{f(x_0) + \frac{f''(x_0)}{2}(x - x_0)^2} dx = e^{f(x_0)} \underbrace{\int_a^b e^{\frac{f''(x_0)}{2}(x - x_0)^2} dx}$$

Gaussian integral:

- closed-form for $b = -a = \infty$

$$\int_{-\infty}^{\infty} e^{\frac{f''(x_0)}{2}(x - x_0)^2} dx = \sqrt{\frac{2\pi}{f''(x_0)}}$$

- in practice, integral is computed by finding maximum x_0

Laplace Integration Method

marginal distribution

joint distribution

$$p(Y) = \int_{\mathbb{R}^n} p(Y,Z)dZ = \int_{\mathbb{R}^n} e^{\log p(Y,Z)} dZ$$

Taylor series of Z $\mapsto \log p(Z,Y)$ around maximum Z_0:

$$\log p(Y,Z) = \log p(Y,Z_0) - \frac{1}{2}(Z-Z_0)' H(Y,Z_0)(Z-Z_0)' + O\big(\|Z-Z_0\|^3\big)$$

$$H(Y,Z_0) = -\frac{\partial^2 \log p(Y,Z)}{\partial Z^2}\Big|_{Z=Z_0}$$

Laplace Integration Method

marginal distribution

joint distribution

$$p(Y) = \int_{\mathbb{R}^n} p(Y,Z)dZ = \int_{\mathbb{R}^n} e^{\log p(Y,Z)} dZ$$

Taylor series of $Z \mapsto \log p(Z,Y)$ around maximum Z_0:

$$\log p(Y,Z) = \log p(Y,Z_0) - \frac{1}{2}(Z-Z_0)' H(Y,Z_0)(Z-Z_0)' + O\big(\|Z-Z_0\|^3\big)$$

$$H(Y,Z_0) = -\frac{\partial^2 \log p(Y,Z)}{\partial Z^2}\bigg|_{Z=Z_0}$$

Assuming 3rd and higher order terms negligible:

$$p(Y) \approx p(Y,Z_0) \int_{\mathbb{R}^n} \underbrace{e^{-\frac{1}{2}(Z-Z_0)' H(Y,Z_0)(Z-Z_0)} dZ}_{\text{Gaussian integral}} \propto \frac{p(Y,Z_0)}{\det H(Y,Z_0)}$$

marginal distribution

joint distribution computed at maximum Z_0

conditional means

$$E[Z|Y] = \int_{\mathbb{R}^n} Z e^{p(Z|Y)} dZ = \int_{\mathbb{R}^n} Z e^{p(Y,Z)-p(Y)} dZ \approx \cdots = Z_0$$

$$\mathrm{CoV}[Z|Y] = \mathrm{E}\big[(Z-\mathrm{E}[Z])(Z-\mathrm{E}[Z])' \mid Y\big] \approx \cdots = H(Y,Z_0)^{-1}$$

Laplace Integration Method

typically easy to compute

Given joint distribution $p(Y,Z)$ we can compute marginal, conditional mean, and covariance matrix through optimization:

$$Z_0 := \arg\max_{Z} \log p(Y,Z)$$

$$p(Y) \approx \frac{c\,p(Y,Z_0)}{\det H(Y,Z_0)}$$

useful for MLE

useful for Bayesian estimation

$$E[Z|Y] \approx Z_0$$

$$\mathrm{CoV}[Z|Y] \approx \frac{\partial^2 \log p(Y,Z)}{\partial Z^2}\Big|_{Z=Z_0}$$

Has been recognized as far back as 1986 [Tierney, Kadane], but now we have the optimization tools to use this in nontrivial problems !

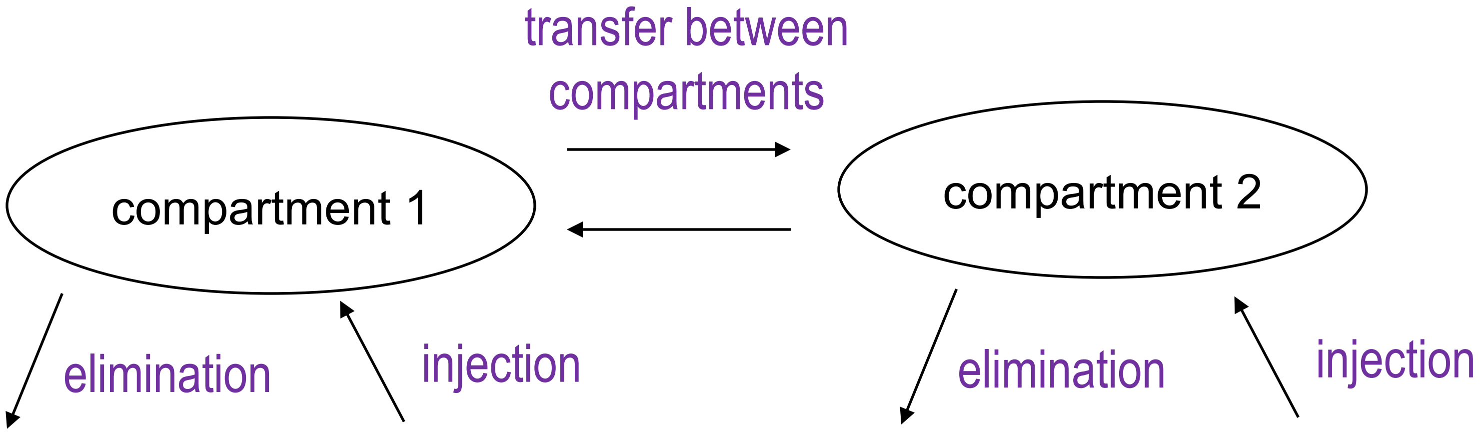
computed at maximum Z_0

conditional means

$$E[Z|Y] = \int_{\mathbb{R}^n} Z e^{p(Z|Y)} dZ = \int_{\mathbb{R}^n} Z e^{p(Y,Z)-p(Y)} dZ \approx \cdots = Z_0$$

$$\mathrm{CoV}[Z|Y] = \mathrm{E}\big[(Z-\mathrm{E}[Z])(Z-\mathrm{E}[Z])' \mid Y\big] \approx \cdots = H(Y,Z_0)^{-1}$$

Pharmacokinetic Model



transfer between compartments

compartment 1

compartment 2

elimination

injection

elimination

injection

$$C_i(t+1) = (1 - K_{E_i})C_i(t) + \sum_j K_{ji}\big(C_j(t) - C_i(t)\big) + \frac{1}{V_i}u_i(t)$$

drug concentration in compartment i at time $t+1$ [moles/liter]

elimination rate

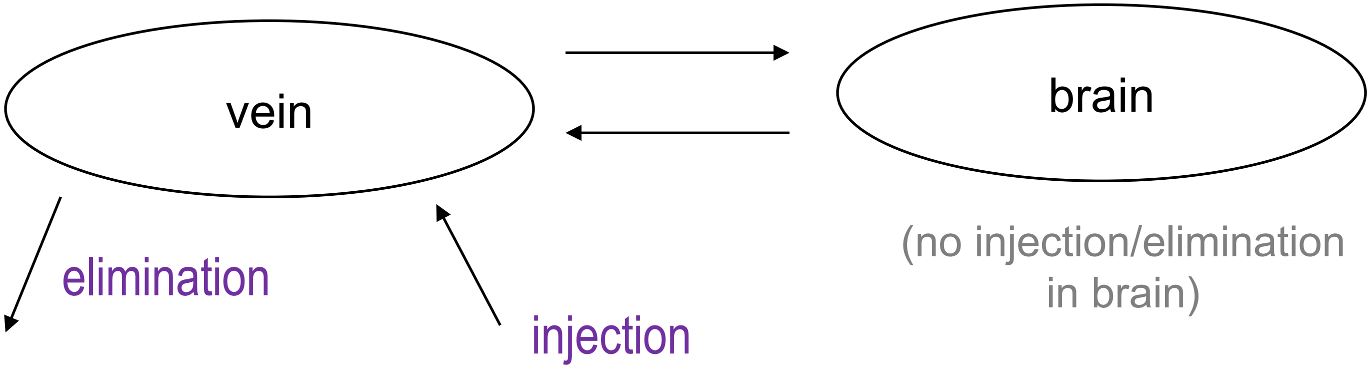
transfer rate between compartments i & j (Fick’s law)

volume of compartment i [liter]

injection at time t [moles]

Vancomycin Rat Model

Vancomycin is an antibiotic medication used to treat a number of bacterial infections



(no injection/elimination in brain)

Vancomycin concentrations

typically small since $\propto \frac{\text{brain volume}}{\text{vein volume}}$

(vein) injection at time t [moles]

$$C_{\text{vein}}(t+1) = (1-K_E)C_{\text{vein}}(t) + K_{BV}C_{\text{brain}}(t) + \frac{u_{\text{vein}}(t)}{V_{\text{vein}}}$$

$$C_{\text{brain}}(t+1) = K_{IO}\big(C_{\text{vein}}(t) - C_{\text{brain}}(t)\big)$$

volume of vein compart. [liter]

meas. noise

$$\left.\begin{aligned} y_{\text{vein}}(t) &= C_{\text{vein}}(t) + w_{\text{vein}}(t) \\ y_{\text{brain}}(t) &= C_{\text{brain}}(t) + w_{\text{brain}}(t) \end{aligned}\right\}$$

measurements from E-AB (Electrochemical-aptamer based) sensors [Plaxco Lab]

- in-vivo
- time-resolution of second

Recall…

$$\begin{aligned} x_{t+1} &= f(x_t, u_t; \theta) + d_t \\ y_t &= g(x_t; \theta) + w_t \end{aligned}$$

Observation: assuming x_{-K}, d_t, w_t independent, joint pdf of x_{-K}, d_t, y_t is easy to write

pdfs of w_t, d_t, x_{-K}

$$\left(\prod_{t=-K}^{0} p_w\big(y_t - g(x_t;\theta)\big) \right) \left(\prod_{t=-K}^{T-1} p_d\big(x_{t+1} - f(x_t, u_t; \theta)\big) \right) p(x_{-K})$$

Why? Proof by induction to compute

$$p\big(x_{t+1}, y_t \mid d_{-K}, \dots, d_t, y_{-K}, \dots, y_{t-1}, x_{-K}; \theta\big)$$

starting at $t = -K$ & using Bayes rule for induction step

Joint state-output pdf of a nonlinear system is easy to compute…

Vancomycin Rat Model

Vancomycin concentrations

$$C_{\text{vein}}(t+1) = (1-K_E)C_{\text{vein}}(t) + \frac{u_{\text{vein}}(t)}{V_{\text{vein}}}$$

$$C_{\text{brain}}(t+1) = K_{IO}\big(C_{\text{vein}}(t) - C_{\text{brain}}(t)\big)$$

(vein) injection at time t [moles]

$$y_{\text{vein}}(t) = C_{\text{vein}}(t) + w_{\text{vein}}(t)$$

$$y_{\text{brain}}(t) = C_{\text{brain}}(t) + w_{\text{brain}}(t)$$

Defining

$$Y := \Big\{y_{\text{vein}}(1), y_{\text{brain}}(1), \dots, y_{\text{vein}}(T), y_{\text{brain}}(T)\Big\}$$

measurements

$$Z := \Big\{K_E, 1/V_{\text{vein}}, K_{IO},$$

unknown parameters – large subject-2-subject variability

$$C_{\text{brain}}(1), C_{\text{vein}}(1)\}$$

initial conditions (known to be zero for our Vancomycin experiments)

$$\theta := \Big\{\sigma^2_{\text{vein}}, \sigma^2_{\text{brain}}\Big\}$$

unknown variances for measurement errors

– large experiment-2-experiment variability

Joint pdf $p(Y, Z; \theta)$ *is easy to compute…*

Vancomycin Rat Model

Defining

$$Y := \left\{ y_{\text{vein}}(1), y_{\text{brain}}(1), \ldots, y_{\text{vein}}(T), y_{\text{brain}}(T) \right\}$$

measurements

$$Z := \Big\{ K_E, 1/V_{\text{vein}}, K_{IO},$$

unknown parameters – large subject-2-subject variability

$$C_{\text{brain}}(1), C_{\text{vein}}(1) \}$$

initial conditions (known to be zero for our Vancomycin experiments)

$$\theta := \left\{ \sigma^2_{\text{vein}}, \sigma^2_{\text{brain}} \right\}$$

unknown variances for measurement errors
– large experiment-2-experiment variability

From Laplace integration formulas …

noise variance estimation:

$$\max_{\theta} p(Y;\theta) \propto \max_{\theta} \max_{Z} \frac{p(Y,Z;\theta)}{\det H(Y,Z;\theta)}$$

MLE to estimate
for noise variances

optimization
for Laplace integration

parameter estimation:

arg-max from
MLE estimate

$$E[Z|Y] \approx Z_0$$

$$\text{CoV}[Z|Y] \approx \frac{\partial^2 \log p(Y,Z)}{\partial Z^2}\bigg|_{Z=Z_0}$$

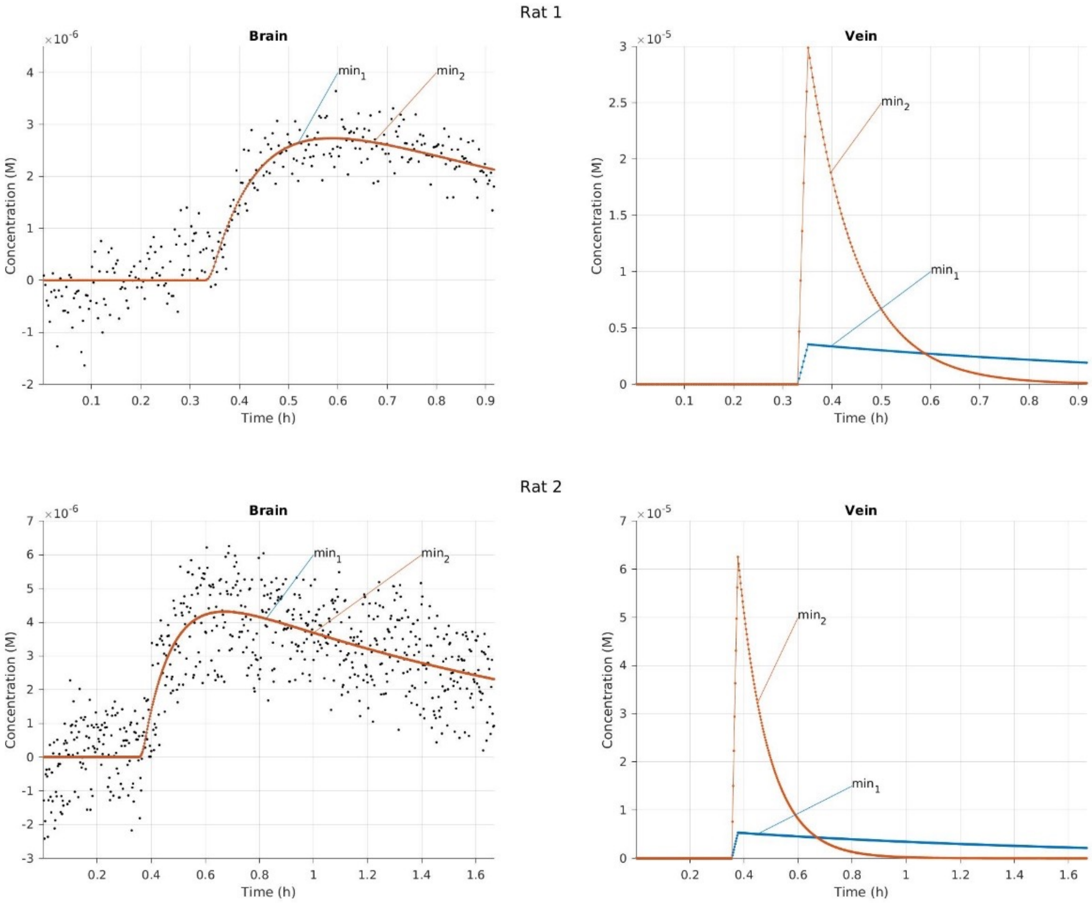
a-posteriori parameter estimates
and error covariances

Experimental results

Results based solely on measurements of brain concentration

individual 1

individual 2



concentration in brain
(measurements
& state estimates)

concentration in vein
(state estimates
for 2 global minima)

Just with brain measurements, optimization has 2 isolated global maxima

⇓

fundamental ambiguity in determining parameters & vein concentration

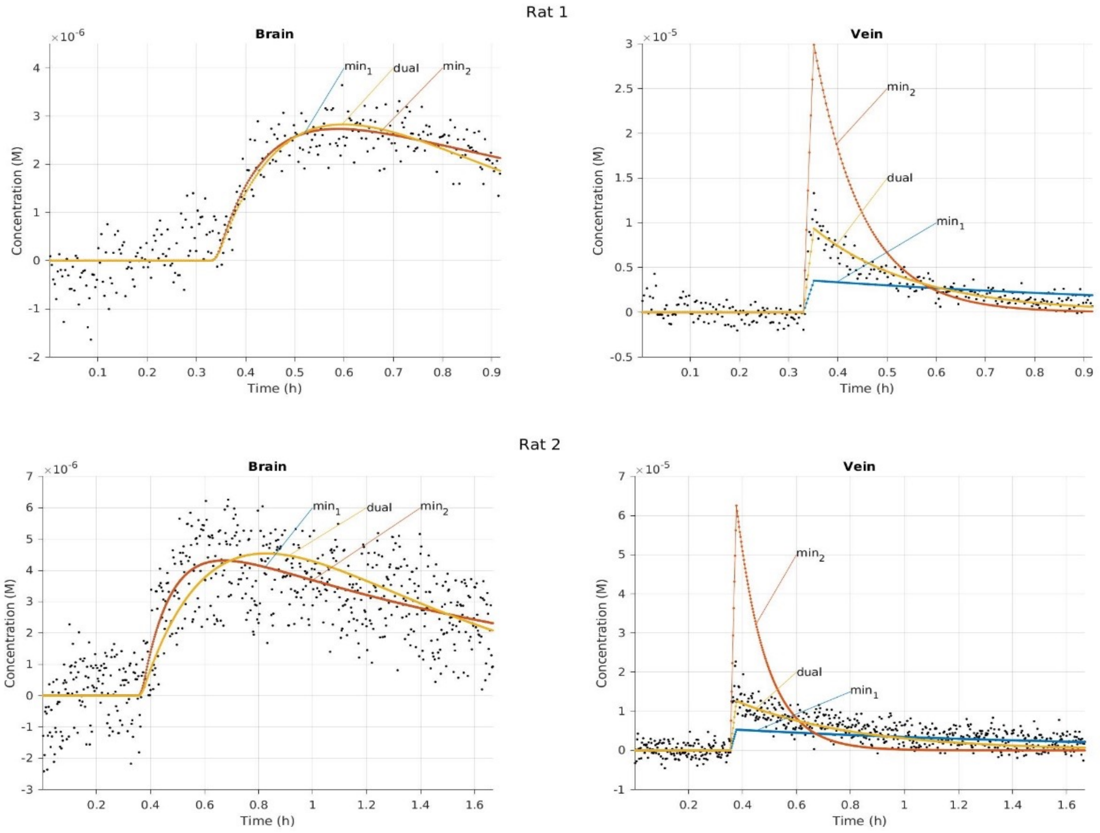
Experimental results

UC SANTA BARBARA engineering

Results based solely on measurements of brain concentration

individual 1

individual 2



concentration in brain (measurements & state estimates)

concentration in vein (measurement & state estimates)

Single maximum with both brain & vein measurements

(previous maxima also shown in right-figures)

COVID-19 forecasting

Stochastic SIR model:

susceptible

$$S(t+1) = S(t) - \nu(t)$$

$$I(t+1) = I(t) + \nu(t) - \rho(t)$$

infected

$$R(t+1) = R(t) + \rho(t)$$

removed

daily new cases

$$\nu(t) = \beta(t)\frac{I(t)}{N_0}S(t) + d_\nu(t)$$

$$\rho(t) = \gamma(t)I(t) + d_\rho(t)$$

daily new removals

stochastic disturbances

time-varying infection/removal rates

Measurement model:

measurement noise

reported daily new cases

$$y_\nu(t) = \phi(t)\nu(t) + w_\nu(t)$$

$$y_D(t) = \omega(t)I(t) + w_D(t)$$

reported daily deaths

time-varying reporting rates

A-priori parameter drift model:

$\beta(t), \gamma(t), \phi(t), \omega(t)$

random walks with unknown variances

Recall…

$$\begin{aligned} x_{t+1} &= f(x_t, u_t; \theta) + d_t \\ y_t &= g(x_t; \theta) + w_t \end{aligned}$$

Observation: assuming x_{-K}, d_t, w_t independent, joint pdf of x_{-K}, d_t, y_t is easy to write

pdfs of w_t, d_t, x_{-K}

$$\left(\prod_{t=-K}^{0} p_w\big(y_t - g(x_t;\theta)\big) \right) \left(\prod_{t=-K}^{T-1} p_d\big(x_{t+1} - f(x_t, u_t; \theta)\big) \right) p(x_{-K})$$

Why? Proof by induction to compute

$$p\big(x_{t+1}, y_t \mid d_{-K}, \dots, d_t, y_{-K}, \dots, y_{t-1}, x_{-K}; \theta\big)$$

starting at $t = -K$ & using Bayes rule for induction step

Joint state-output pdf of a nonlinear system is easy to compute…

COVID-19 forecasting

Stochastic SIR model:

susceptible

$$S(t+1) = S(t) - \nu(t)$$

$$I(t+1) = I(t) + \nu(t) - \rho(t)$$

infected

$$R(t+1) = R(t) + \rho(t)$$

removed

$$\nu(t) = \beta(t)\frac{I(t)}{N_0}S(t) + d_\nu(t)$$

$$\rho(t) = \gamma(t)I(t) + d_\rho(t)$$

$$y_\nu(t) = \phi(t)\nu(t) + w_\nu(t)$$

$$y_D(t) = \omega(t)I(t) + w_D(t)$$

Defining

$$Y \coloneqq \Big\{y_\nu(1), y_D(1), \ldots, y_\nu(T), y_D(T)\Big\}$$

past measurements

$$Z \coloneqq \Big\{S(1), I(1), R(1), \beta(1), \gamma(1), \phi(1), \omega(1), \ldots, \omega(T+P),$$

past and future states

$$y_\nu(T+1), y_D(T+1), \ldots, y_\nu(T+P), y_D(T+P)\Big\}$$

future measurements

Joint pdf $p(Y, Z)$ *is easy to compute…*

COVID-19 forecasting

UC SANTA BARBARA
engineering

Defining

$$Y \coloneqq \Big\{ y_\nu(1), y_D(1), \dots, y_\nu(T), y_D(T) \Big\}$$

past measurements

$$Z \coloneqq \Big\{ S(1), I(1), R(1), \beta(1), \gamma(1), \phi(1), \omega(1), \dots, \omega(T+P),$$

past and future states

$$y_\nu(T+1), y_D(T+1), \dots, y_\nu(T+P), y_D(T+P) \Big\}$$

future measurements

$$Z_0 \coloneqq \arg \max_Z \log p(Y, Z)$$

Identification:

$$\max_\theta p(Y;\theta) \propto \max_\theta \max_Z \frac{p(Y,Z;\theta)}{\det H(Y,Z;\theta)}$$

MLE to estimate unknown for noise/disturbance variances

optimization for Laplace integration

Forecasting:

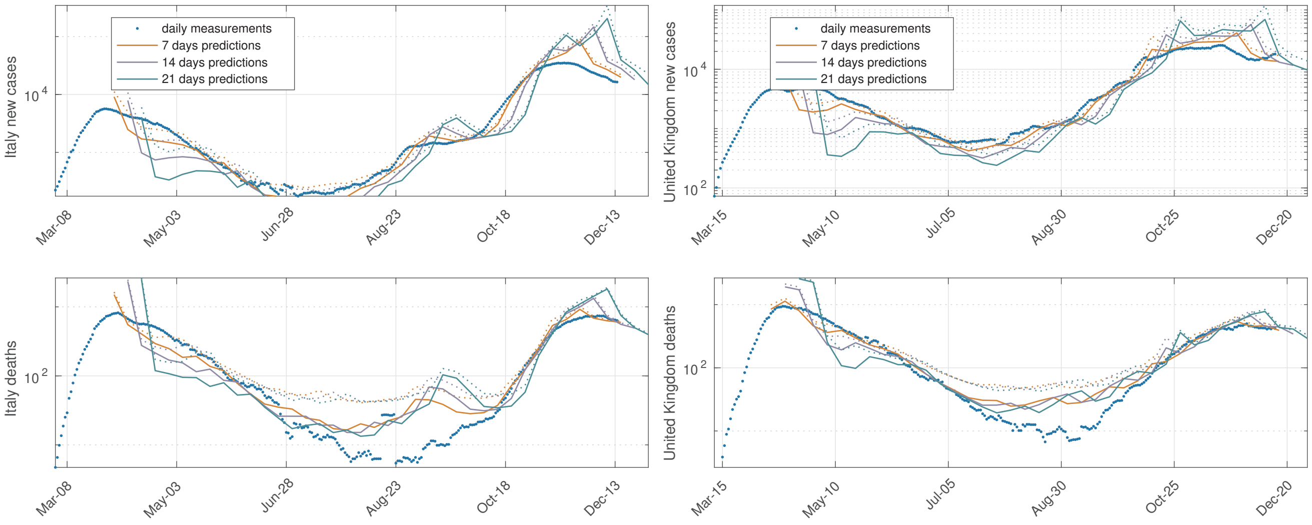
arg-max from MLE estimate

$$E[Z|Y] \approx Z_0$$

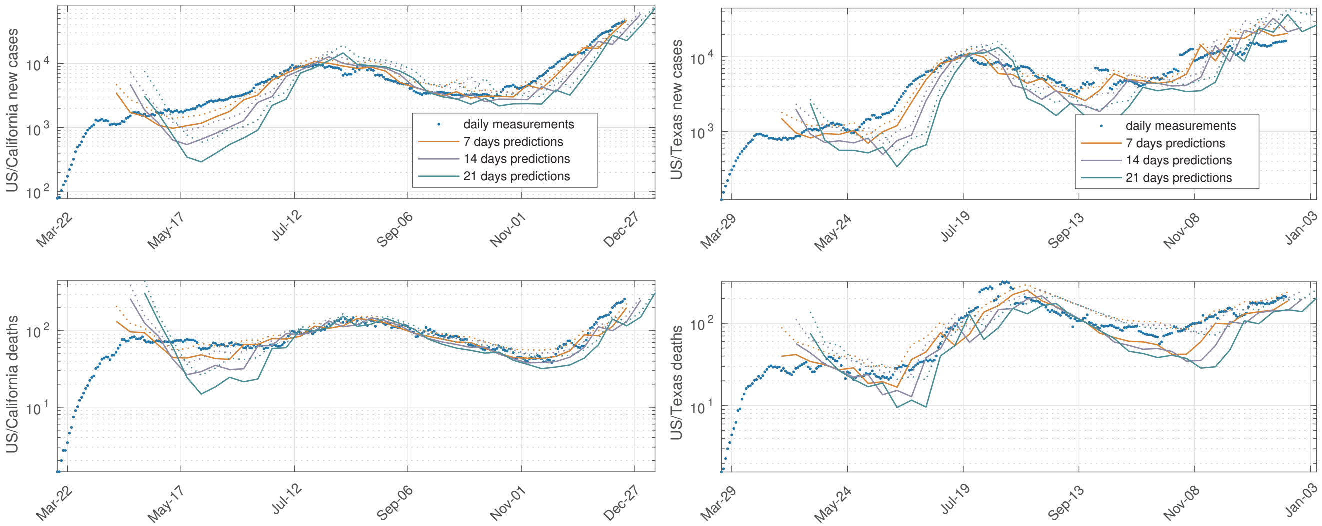
$$\mathrm{CoV}[Z|Y] \approx \frac{\partial^2 \log p(Y,Z)}{\partial Z^2}\bigg|_{Z=Z_0}$$

a-posteriori forecasts and error covariances

Results



(a) Italy (b) United Kingdom



(c) California, USA (d) Texas, USA

Model not identifiable, but still possible to reliably compute

1. Maximum likelihood estimates for unknown random walk variances
2. 7-14-21 day Bayesian forecasts for measured variables

Conclusions

Problem addressed: $\min_{u} \mathrm{E}_d\left[f(u,d)\right]$

- no closed form for expected value
- willing to sacrifice accuracy for speed
- error bounds

Optimization-based bounds
1. Additive/multiplicative bounds for expected values
2. Laplace method for integration

What I did not talk about:
- Numerical methods to solve min-max optimizations
  - primal-dual interior point methods for min-max equilibria
  - MATLAB toolbox (TensCalc)

Future/current work:
- Which methods/bounds to use?
- Combination with MC methods

In collaboration with:
Raphael Chinchilla, Murat Erdal, Dr. Guosong Yang (UCSB)
Prof. Ramon Costa (Federal Univ. Rio Janeiro, Brazil)
Prof. Kevin Plaxco (UCSB)

Conclusions

Problem addressed: $\min_u \mathrm{E}_d\left[f(u,d)\right]$

- no closed form for expected value
- willing to sacrifice accuracy for speed
- error bounds

Optimization-based bounds

1. Additive/multiplicative bounds for expected values
2. Laplace method for integration

What I did not talk about:

- Numerical methods to solve min-max optimizations
  - primal-dual interior point methods for min-max equilibria
  - MATLAB toolbox (TensCalc)

Future/current work:

- Which methods/bounds to use?
- Combination with MC methods

THANKS!

In collaboration with:
Raphael Chinchilla, Murat Erdal, Dr. Guosong Yang (UCSB)
Prof. Ramon Costa (Federal Univ. Rio Janeiro, Brazil)
Prof. Kevin Plaxco (UCSB)