Methods for Inverse Problems:
VII. Uncertainty quantification and adjoint based gradient computation

Barbara Kaltenbacher, University of Klagenfurt, Austria

General Model Setting

$$A(u, \theta) = 0$$

$u \in V$ ... state
$\theta \in \Theta \subseteq \mathbb{R}^{n_\theta}$ ... parameter vector
$A : V \times \Theta \to W$ ... model operator

General Model Setting

$$A(u,\theta) = 0$$

$u \in V \dots$ state
$\theta \in \Theta \subseteq \mathbb{R}^{n_\theta} \dots$ parameter vector
$A : V \times \Theta \to W \dots$ model operator

e.g., dynamical state space model

$$\dot{x}(t) = f(t, x(t), \theta)\,, \quad t \in [0, T]\,, \qquad x(0) = x_0(\theta)$$

with $u = x - x_0(\theta)$

$$A(u,\theta) = \dot{u} - f(\cdot, x_0(\theta) + u, \theta)\,, \quad V = C_0^1(0, T; \mathbb{R}^n),\ W = C(0, T; \mathbb{R}^n)$$

General Model Setting

$$A(u,\theta) = 0$$

$u \in V \ldots$ state
$\theta \in \Theta \subseteq \mathbb{R}^{n_\theta} \ldots$ parameter vector
$A : V \times \Theta \to W \ldots$ model operator

e.g., finite element model in continuum mechanics

$$\tilde{K}(u,\theta) - f(\theta)$$

$u \ldots$ nodal displacements $\quad f \ldots$ load
$\tilde{K} \ldots$ (possibly nonlinear) finite element model

$$A(u,\theta) = \tilde{K}(u,\theta) - f(\theta)\,, \quad V = W = \mathbb{R}^n$$

General Model + Observations

$$A(u,\theta) = 0$$
$$y = G(u,\theta) + \eta$$

$u \in V \ldots$ state
$\theta \in \Theta \subseteq \mathbb{R}^{n_\theta} \ldots$ parameter vector
$A : V \times \Theta \to W \ldots$ model operator
$G : V \times \theta \to Z \ldots$ observation operator
$y \in Z \ldots$ measurement data
$\eta \in Z \ldots$ noise

General Model + Observations

$$A(u, \theta) = 0$$
$$y = G(u, \theta) + \eta$$

$u \in V \ldots$ state
$\theta \in \Theta \subseteq \mathbb{R}^{n_\theta} \ldots$ parameter vector
$A : V \times \Theta \to W \ldots$ model operator
$G : V \times \theta \to Z \ldots$ observation operator
$y \in Z \ldots$ measurement data
$\eta \in Z \ldots$ noise

e.g., dynamical state space model

$$\dot{x}(t) = f(t, x(t), \theta)\,, \quad t \in [0, T]\,, \qquad x(0) = x_0(\theta)$$
$$y(t) = \tilde{G}(t, x(t), \theta) + \eta(t)\,, \quad t \in [0, T]\,, \text{ or } y_i = \tilde{G}_i(x(t_i), \theta) + \eta_i$$

General Model + Observations

$$A(u,\theta) = 0$$
$$y = G(u,\theta) + \eta$$

$u \in V \dots$ state
$\theta \in \Theta \subseteq \mathbb{R}^{n_\theta} \dots$ parameter vector
$A : V \times \Theta \to W \dots$ model operator
$G : V \times \theta \to Z \dots$ observation operator
$y \in Z \dots$ measurement data
$\eta \in Z \dots$ noise

e.g., finite element model in continuum mechanics

$$\tilde{K}(u,\theta) - f(\theta)$$
$$y = G(u,\theta) + \eta$$

$u \dots$ nodal displacements $f \dots$ load
$\tilde{K} \dots$ (possibly nonlinear) finite element model
$y \dots$ e.g., strain measurements

All-at-Once versus Reduced Formulation

All-at-Once

Consider

$$A(u, \theta) = 0$$
$$y = G(u, \theta) + \eta$$

as a large system of equations with unknowns (u, θ)

All-at-Once versus Reduced Formulation

All-at-Once

Consider

$$A(u,\theta) = 0$$
$$y = G(u,\theta) + \eta$$

as a large system of equations with unknowns (u,θ)

Reduced

Use parameter-to-state map $S : \mathbb{R}^{n_\theta} \to V$, $\theta \mapsto u$, satisfying $A(S(\theta),\theta) = 0$ to reduce the problem to

$$y = G(S(\theta),\theta) + \eta = g(\theta) + \eta$$

as a system of equations for θ only.

All-at-Once versus Reduced Formulation

All-at-Once

Consider

$$A(u,\theta) = 0$$
$$y = G(u,\theta) + \eta$$

as a large system of equations with unknowns (u,θ)

Reduced

Use parameter-to-state map $S : \mathbb{R}^{n_\theta} \to V$, $\theta \mapsto u$, satisfying $A(S(\theta),\theta) = 0$ to reduce the problem to

$$y = G(S(\theta),\theta) + \eta = g(\theta) + \eta$$

as a system of equations for θ only.

Reduced form needs solvabilty of $A(\cdot,\theta) = 0$ for fixed $\theta \in \Theta$, e.g., $\frac{\partial A}{\partial u}(u,\theta)$ boundedly invertible + Implicit Function Theorem.

All-at-Once versus Reduced Formulation

All-at-Once

Consider

$$A(u,\theta) = 0$$
$$y = G(u,\theta) + \eta$$

as a large system of equations with unknowns (u,θ)

Reduced

Use parameter-to-state map $S : \mathbb{R}^{n_\theta} \to V$, $\theta \mapsto u$, satisfying $A(S(\theta),\theta) = 0$ to reduce the problem to

$$y = G(S(\theta),\theta) + \eta = g(\theta) + \eta$$

as a system of equations for θ only.

example dynamical state space model: $\dot{x}(t) = f(t, x(t), \theta)$

regularity of $\frac{\partial A}{\partial u}(u,\theta) = \frac{\partial}{\partial t} - \frac{\partial f}{\partial x}(x,\theta)$ by smoothness/monotonicity of $f(\cdot,\theta)$

All-at-Once versus Reduced Formulation

All-at-Once

Consider

$$A(u, \theta) = 0$$
$$y = G(u, \theta) + \eta$$

as a large system of equations with unknowns (u, θ)

Reduced

Use parameter-to-state map $S : \mathbb{R}^{n_\theta} \to V$, $\theta \mapsto u$, satisfying $A(S(\theta), \theta) = 0$ to reduce the problem to

$$y = G(S(\theta), \theta) + \eta = g(\theta) + \eta$$

as a system of equations for θ only.

example FE model: $\tilde{K}(u, \theta) = f(\theta)$

regularity of $\frac{\partial A}{\partial u}(u, \theta) = \frac{\partial \tilde{K}}{\partial u}(u, \theta)$ by ellipticity.

Maximum Likelihood Parameter Estimation

Likelihood Function

$L(\theta) =$
probabilty (density) of observed data, if θ are true parameters

Maximum Likelihood Parameter Estimation

Likelihood Function

$\mathrm{L}(\theta) =$
probabilty (density) of observed data, if θ are true parameters

e.g., in case $\eta_k \sim \mathcal{N}(0, \sigma_k)$ and mutually independent:

$$\mathrm{L}(\theta) = J(S(\theta), \theta) = \prod_{k=1}^{K} \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{1}{2}\left(\frac{y_k^\delta - G_k(S(\theta), \theta)}{\sigma_k}\right)^2\right),$$

where σ_k might be part of the searched for parameter vector.

Maximum Likelihood Parameter Estimation

Likelihood Function

$\mathrm{L}(\theta) =$
probabilty (density) of observed data, if θ are true parameters

e.g., in case $\eta_k \sim \mathcal{N}(0, \sigma_k)$ and mutually independent:

$$\mathrm{L}(\theta) = J(S(\theta), \theta) = \prod_{k=1}^{K} \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{1}{2}\left(\frac{y_k^\delta - G_k(S(\theta), \theta)}{\sigma_k}\right)^2\right),$$

where σ_k might be part of the searched for parameter vector.

Maximum Likelihood Estimator

$\hat{\theta} =$ maximizer of L = minimizer of $\ell = -\log L$

Optimization Based Parameter Estimation

All-at-Once

$$\min_{\theta \in \Theta, u \in V} J(u, \theta)$$
$$\text{s.t. } A(u, \theta) = 0$$

Optimization Based Parameter Estimation

All-at-Once

$$\min_{\theta\in\Theta, u\in V} J(u,\theta)$$
$$\text{s.t.}\ \ A(u,\theta) = 0$$

Reduced

$$\min_{\theta\in\Theta} j(\theta) = J(S(\theta),\theta)$$

where, e.g, $j(\theta) = \ell(\theta)$ negative log-likelihood

Uncertainty Analysis

Confidence Region

$$\mathrm{CR}_\alpha = \left\{ \theta \in \Theta \,\middle|\, \frac{\mathrm{L}(\theta)}{\mathrm{L}(\hat{\theta})} \geq \exp\left(-\frac{\Delta_\alpha}{2}\right) \right\},$$

$$= \left\{ \theta \in \Theta \,\middle|\, 2\left(j(\theta) - j(\hat{\theta})\right) \leq \Delta_\alpha \right\} \subseteq \Theta$$

$\Delta_\alpha \ldots$ upper α-quantile of χ^2 distr. with 1 dof (cf. Likelihood Ratio Test).

Uncertainty Analysis

Confidence Region

$$\mathrm{CR}_\alpha = \left\{ \theta \in \Theta \,\middle|\, \frac{\mathrm{L}(\theta)}{\mathrm{L}(\hat{\theta})} \geq \exp\left(-\frac{\Delta_\alpha}{2}\right) \right\},$$
$$= \left\{ \theta \in \Theta \,\middle|\, 2\left(j(\theta) - j(\hat{\theta})\right) \leq \Delta_\alpha \right\} \subseteq \Theta$$

$\Delta_\alpha \ldots$ upper α-quantile of χ^2 distr. with 1 dof (cf. Likelihood Ratio Test).

quadratic approximation by Taylor expansion, using $\frac{\partial j}{\partial \theta_i}(\hat{\theta}) = 0$

$$\widetilde{\mathrm{CR}}_\alpha = \left\{ \theta \in \Theta \,\middle|\, (\theta - \hat{\theta})^T \nabla^2 j(\hat{\theta})(\theta - \hat{\theta}) \leq \Delta_\alpha \right\}.$$

(contains $\frac{\partial S}{\partial \theta_i}$, $\frac{\partial^2 S}{\partial \theta_i \theta_l}$ since $j(\theta) = J(S(\theta, \theta))$ or

Uncertainty Analysis

Confidence Region

$$\mathrm{CR}_\alpha = \left\{ \theta \in \Theta \,\middle|\, \frac{\mathrm{L}(\theta)}{\mathrm{L}(\hat{\theta})} \geq \exp\left(-\frac{\Delta_\alpha}{2}\right) \right\},$$
$$= \left\{ \theta \in \Theta \,\middle|\, 2\left(j(\theta) - j(\hat{\theta})\right) \leq \Delta_\alpha \right\} \subseteq \Theta$$

$\Delta_\alpha \ldots$ upper α-quantile of χ^2 distr. with 1 dof (cf. Likelihood Ratio Test).

quadratic approximation by Taylor expansion, using $\frac{\partial j}{\partial \theta_i}(\hat{\theta}) = 0$

$$\widetilde{\mathrm{CR}}_\alpha = \left\{ \theta \in \Theta \,\middle|\, (\theta - \hat{\theta})^T \nabla^2 j(\hat{\theta})(\theta - \hat{\theta}) \leq \Delta_\alpha \right\}.$$

(contains $\frac{\partial S}{\partial \theta_i}$, $\frac{\partial^2 S}{\partial \theta_i \theta_l}$ since $j(\theta) = J(S(\theta, \theta))$ or
with Fisher Information Matrix FIM $\widetilde{H}(\hat{\theta})$ (skipping $\frac{\partial^2 S}{\partial \theta_i \theta_l}$ terms)

$$\widetilde{\widetilde{\mathrm{CR}}}_\alpha = \left\{ \theta \in \Theta \,\middle|\, (\theta - \hat{\theta})^T \widetilde{H}(\hat{\theta})(\theta - \hat{\theta}) \leq \Delta_\alpha \right\}.$$

$\rightsquigarrow$ confidence ellipsoids

Uncertainty Analysis

consider model property $q(\theta) = Q(S(\theta), \theta)$,
e.g, $Q(u, \theta) = \theta_i$
and quantify uncertainty in $q(\theta)$ due to measurement noise:

Confidence Interval

$$\mathrm{CI}_{\alpha,q} = \{c \in \mathbb{R} \,|\, \exists\theta \in \mathrm{CR}_\alpha \ :\ q(\theta) = c\}.$$

Uncertainty Analysis

consider model property $q(\theta) = Q(S(\theta), \theta)$,
e.g, $Q(u, \theta) = \theta_i$
and quantify uncertainty in $q(\theta)$ due to measurement noise:

Confidence Interval

$$\mathrm{CI}_{\alpha,q} = \{c \in \mathbb{R} \,|\, \exists \theta \in \mathrm{CR}_\alpha \;:\; q(\theta) = c\}\,.$$

analogously for the quadratic approximations
with CR_α replaced by $\widetilde{\mathrm{CR}}_\alpha$, $\widetilde{\widetilde{\mathrm{CR}}}_\alpha$
$\rightsquigarrow$ explicitely computable confidence intervals

Uncertainty Analysis

consider model property $q(\theta) = Q(S(\theta), \theta)$,
e.g, $Q(u, \theta) = \theta_i$
and quantify uncertainty in $q(\theta)$ due to measurement noise:

Confidence Interval

$$\mathrm{CI}_{\alpha,q} = \{c \in \mathbb{R} \,|\, \exists \theta \in \mathrm{CR}_\alpha \ : \ q(\theta) = c\}\,.$$

analogously for the quadratic approximations
with CR_α replaced by $\widetilde{\mathrm{CR}}_\alpha$, $\widetilde{\widetilde{\mathrm{CR}}}_\alpha$
$\rightsquigarrow$ explicitely computable confidence intervals
But these approximations may lead to underestimation of $\mathrm{CI}_{\alpha,q}$!

Uncertainty Analysis

consider model property $q(\theta) = Q(S(\theta), \theta)$,
e.g, $Q(u, \theta) = \theta_i$
and quantify uncertainty in $q(\theta)$ due to measurement noise:

Confidence Interval

$$\mathrm{CI}_{\alpha,q} = \{c \in \mathbb{R} \,|\, \exists \theta \in \mathrm{CR}_\alpha \ : \ q(\theta) = c\}\,.$$

Profile Likelihood

$$\mathrm{PL}_q(c) = \max_{\theta \in \Theta} \mathrm{L}(\theta) \text{ subject to } q(\theta) = c.$$

Uncertainty Analysis

consider model property $q(\theta) = Q(S(\theta), \theta)$,
e.g, $Q(u, \theta) = \theta_i$
and quantify uncertainty in $q(\theta)$ due to measurement noise:

Confidence Interval

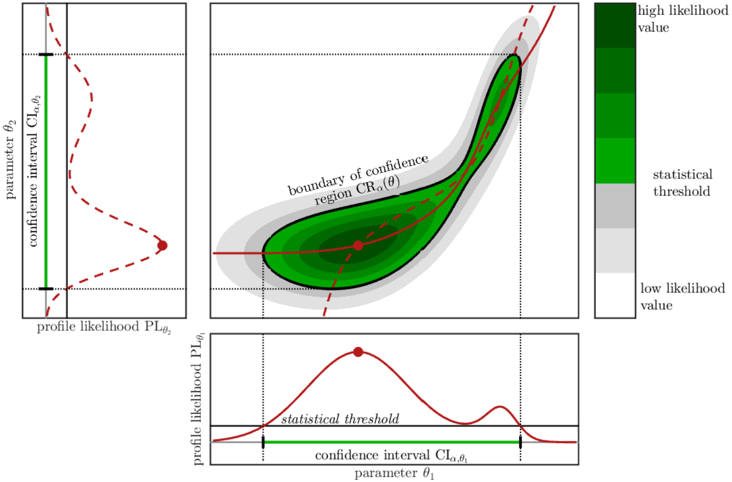
$$\mathrm{CI}_{\alpha,q} = \{c \in \mathbb{R} \,|\, \exists\theta \in \mathrm{CR}_\alpha \;:\; q(\theta) = c\}.$$

Profile Likelihood

$$\mathrm{PL}_q(c) = \max_{\theta\in\Theta} \mathrm{L}(\theta) \text{ subject to } q(\theta) = c.$$

$$\Rightarrow \quad \mathrm{CI}_{\alpha,q} = \left\{ c \in \mathbb{R} \,\middle|\, \exists\theta \in \Theta \;:\; \frac{\mathrm{L}(\theta)}{\mathrm{L}(\hat\theta)} \geq \exp\left(-\frac{\Delta_\alpha}{2}\right) \text{ and } q(\theta) = c \right\}$$

$$= \left\{ c \in \mathbb{R} \,\middle|\, \frac{\mathrm{PL}_q(c)}{\mathrm{L}(\hat\theta)} \geq \exp\left(-\frac{\Delta_\alpha}{2}\right) \right\}$$



parameter θ_2
confidence interval CI_{α,θ_2}
profile likelihood PL_{θ_2}
boundary of confidence region $CR_\alpha(\theta)$
high likelihood value
statistical threshold
low likelihood value
profile likelihood PL_{θ_1}
statistical threshold
confidence interval CI_{α,θ_1}
parameter θ_1

Optimization Based Profile Likelihood Computation

$\mathrm{PL}_q(c_l) = \exp(-\mathrm{pl}(c_l))$, $l = 1, 2, 3 \ldots$, where

$$\mathrm{pl}_q(c_l) = \min_{\theta \in \Theta} j(\theta) \quad \text{s.t. } q(\theta) = c_l,$$

$c_0 = q(\hat{\theta})$.

Starting point $\theta_{c_l}^{(0)}$ of local optimization for c_l:

- 0 order proposal: $\theta_{c_l}^{(0)} = \theta_{c_{l-1}}$ or
- 1^{st} order proposal: linear extrapolation from $\theta_{c_{l-1}}$ and $\theta_{c_{l-2}}$:

$$\theta_{c_l}^{(0)} = \theta_{c_{l-1}} + \frac{c_l - c_{l-1}}{c_{l-1} - c_{l-2}} (\theta_{c_{l-1}} - \theta_{c_{l-2}}).$$

Integration Based Profile Likelihood Computation

$$\mathrm{pl}_q(c) = \min_{\theta \in \Theta} j(\theta)$$
$$\text{s.t. } q(\theta) = c.$$

1st order optimality conditions
$$\begin{cases} \nabla j(\theta_c) + \lambda \nabla q(\theta_c) = 0 \\ q(\theta_c) = c, \end{cases} \tag{1}$$

Differentiate with respect to c to derive a DAE system ($\dot{} = \frac{d}{dc}$)

$$\begin{pmatrix} \nabla^2 j(\theta_c) + \lambda_c \nabla^2 q(\theta_c) & \nabla q(\theta_c) \\ \nabla q(\theta_c)^T & 0 \end{pmatrix} \begin{pmatrix} \dot{\theta}_c \\ \dot{\lambda}_c \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{2}$$

Integration Based Profile Likelihood Computation

$$\mathrm{pl}_q(c) = \min_{\theta \in \Theta} j(\theta)$$
$$\text{s.t. } q(\theta) = c.$$

$$\text{1st order optimality conditions } \begin{cases} \nabla j(\theta_c) + \lambda \nabla q(\theta_c) = 0 \\ q(\theta_c) = c, \end{cases} \tag{1}$$

Differentiate with respect to c to derive a DAE system ($\dot{} = \frac{d}{dc}$)

$$\begin{pmatrix} \nabla^2 j(\theta_c) + \lambda_c \nabla^2 q(\theta_c) & \nabla q(\theta_c) \\ \nabla q(\theta_c)^T & 0 \end{pmatrix} \begin{pmatrix} \dot{\theta}_c \\ \dot{\lambda}_c \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{2}$$

or, alternatively, an ODE system

$$\begin{pmatrix} \dot{\theta}_c \\ \dot{\lambda}_c \end{pmatrix} = \begin{pmatrix} \nabla^2 j(\theta_c) + \lambda_c \nabla^2 q(\theta_c) & \nabla q(\theta_c) \\ \nabla q(\theta_c)^T & 0 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

Integration Based Profile Likelihood Computation

$$\mathrm{pl}_q(c) = \min_{\theta \in \Theta} j(\theta)$$
$$\text{s.t. } q(\theta) = c.$$

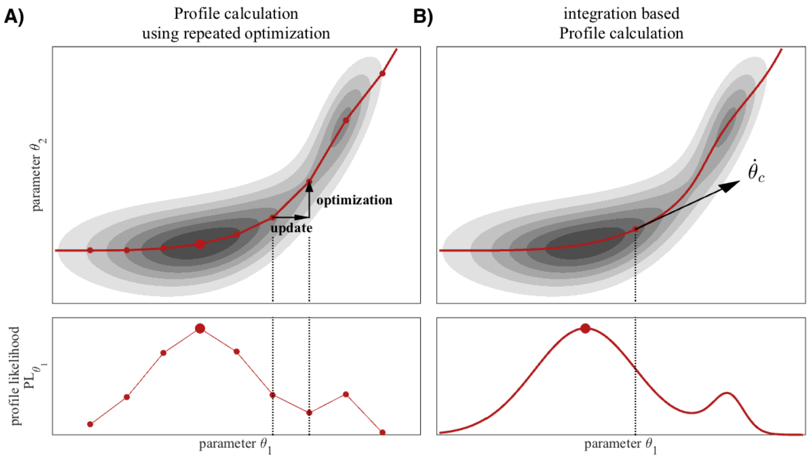
1st order optimality conditions
$$\begin{cases} \nabla j(\theta_c) + \lambda \nabla q(\theta_c) = 0 \\ q(\theta_c) = c, \end{cases} \quad (1)$$

Differentiate with respect to c to derive a DAE system ($\dot{} = \frac{d}{dc}$)

$$\begin{pmatrix} \nabla^2 j(\theta_c) + \lambda_c \nabla^2 q(\theta_c) & \nabla q(\theta_c) \\ \nabla q(\theta_c)^T & 0 \end{pmatrix} \begin{pmatrix} \dot{\theta}_c \\ \dot{\lambda}_c \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad (2)$$

Theorem

Let $j \in C^2(\mathbb{R}^n, \mathbb{R})$, $q : C(\mathbb{R}^{n_\theta}, \mathbb{R})$ and $(\theta_c, \lambda_c)_{c \in [c_0, c_{\mathrm{end}}]}$ solve (2) with $(\theta_{c_0}, \lambda_{c_0})$ solving (1) for $c = c_0$ (e.g., $c_0 = q(\hat{\theta})$, $\theta_{c_0} = \hat{\theta}$, $\lambda_{c_0} = 0$).
Then for all $c \in [c_0, c_{\mathrm{end}}]$, (θ_c, λ_c) solves (1).


A)
Profile calculation
using repeated optimization
B)
integration based
Profile calculation
parameter θ_2
optimization
update
$\dot{\theta}_c$
profile likelihood
PL_{θ_1}
parameter θ_1
parameter θ_1

Gradient (and Hessian) Computation

Chain Rule for

$$j(\theta) = J(S(\theta), \theta)$$

and Implicit Function Theorem for

$$A(S(\theta), \theta) = 0$$

Gradient (and Hessian) Computation

Chain Rule for

$$j(\theta) = J(S(\theta), \theta) \quad \Rightarrow \quad \frac{\partial j}{\partial \theta_i}(\theta) = \frac{\partial J}{\partial u}(S(\theta), \theta)\frac{\partial S}{\partial \theta_i}(\theta) + \frac{\partial J}{\partial \theta_i}(S(\theta), \theta)$$

and Implicit Function Theorem for

$$A(S(\theta), \theta) = 0$$

Gradient (and Hessian) Computation

Chain Rule for

$$j(\theta) = J(S(\theta), \theta) \quad \Rightarrow \quad \frac{\partial j}{\partial \theta_i}(\theta) = \frac{\partial J}{\partial u}(S(\theta), \theta) \frac{\partial S}{\partial \theta_i}(\theta) + \frac{\partial J}{\partial \theta_i}(S(\theta), \theta)$$

and Implicit Function Theorem for

$$A(S(\theta), \theta) = 0 \quad \Rightarrow \quad v_i = \frac{\partial S}{\partial \theta_i}(\theta) \text{ solves } \frac{\partial A}{\partial u}(S(\theta), \theta)\, v_i = -\frac{\partial A}{\partial \theta_i}(S(\theta), \theta)$$

Gradient (and Hessian) Computation

Chain Rule for

$$j(\theta) = J(S(\theta),\theta) \quad \Rightarrow \quad \frac{\partial j}{\partial \theta_i}(\theta) = \frac{\partial J}{\partial u}(S(\theta),\theta)\frac{\partial S}{\partial \theta_i}(\theta) + \frac{\partial J}{\partial \theta_i}(S(\theta),\theta)$$

and Implicit Function Theorem for

$$A(S(\theta),\theta) = 0 \quad \Rightarrow \quad v_i = \frac{\partial S}{\partial \theta_i}(\theta) \text{ solves } \frac{\partial A}{\partial u}(S(\theta),\theta)\, v_i = -\frac{\partial A}{\partial \theta_i}(S(\theta),\theta)$$

$$\begin{aligned}\frac{\partial^2 j}{\partial \theta_i \partial \theta_l}(\theta) =& \frac{\partial J}{\partial u}(S(\theta),\theta)\frac{\partial^2 S}{\partial \theta_i \partial \theta_l}(\theta) + \frac{\partial^2 J}{\partial u^2}(S(\theta),\theta)\left(\frac{\partial S}{\partial \theta_i}(\theta), \frac{\partial S}{\partial \theta_l}(\theta)\right) \\ &+ \frac{\partial^2 J}{\partial \theta_i \partial u}(S(\theta),\theta)\frac{\partial S}{\partial \theta_l}(\theta) + \frac{\partial^2 J}{\partial \theta_l \partial u}(S(\theta),\theta)\frac{\partial S}{\partial \theta_i}(\theta) + \frac{\partial^2 J}{\partial \theta_i \partial \theta_l}(S(\theta),\theta)\,,\end{aligned}$$

with second order sensitivities $w_{il} = \frac{\partial S}{\partial \theta_i \theta_l}$ solving

$$\begin{aligned}\frac{\partial A}{\partial u}(S(\theta),\theta) w_{i,l} =& -\frac{\partial^2 A}{\partial u^2}(S(\theta),\theta)(v_i, v_l) - \frac{\partial^2 A}{\partial \theta_i \partial u}(S(\theta),\theta) v_l \\ &- \frac{\partial^2 A}{\partial \theta_l \partial u}(S(\theta),\theta) v_i - \frac{\partial^2 A}{\partial \theta_i \partial \theta_l}(S(\theta),\theta)\,,\end{aligned}$$

Gradient (and Hessian) Computation

Chain Rule for

$$j(\theta) = J(S(\theta),\theta) \quad \Rightarrow \quad \frac{\partial j}{\partial \theta_i}(\theta) = \frac{\partial J}{\partial u}(S(\theta),\theta)\frac{\partial S}{\partial \theta_i}(\theta) + \frac{\partial J}{\partial \theta_i}(S(\theta),\theta)$$

and Implicit Function Theorem for

$$A(S(\theta),\theta) = 0 \quad \Rightarrow \quad v_i = \frac{\partial S}{\partial \theta_i}(\theta) \text{ solves } \frac{\partial A}{\partial u}(S(\theta),\theta)\, v_i = -\frac{\partial A}{\partial \theta_i}(S(\theta),\theta)$$

$$\begin{aligned}\frac{\partial^2 j}{\partial \theta_i \partial \theta_l}(\theta) =& \frac{\partial J}{\partial u}(S(\theta),\theta)\frac{\partial^2 S}{\partial \theta_i \partial \theta_l}(\theta) + \frac{\partial^2 J}{\partial u^2}(S(\theta),\theta)\left(\frac{\partial S}{\partial \theta_i}(\theta), \frac{\partial S}{\partial \theta_l}(\theta)\right)\\ &+ \frac{\partial^2 J}{\partial \theta_i \partial u}(S(\theta),\theta)\frac{\partial S}{\partial \theta_l}(\theta) + \frac{\partial^2 J}{\partial \theta_l \partial u}(S(\theta),\theta)\frac{\partial S}{\partial \theta_i}(\theta) + \frac{\partial^2 J}{\partial \theta_i \partial \theta_l}(S(\theta),\theta)\,,\end{aligned}$$

with second order sensitivities $w_{il} = \frac{\partial S}{\partial \theta_i \theta_l}$ solving

$$\begin{aligned}\frac{\partial A}{\partial u}(S(\theta),\theta) w_{i,l} =& -\frac{\partial^2 A}{\partial u^2}(S(\theta),\theta)(v_i, v_l) - \frac{\partial^2 A}{\partial \theta_i \partial u}(S(\theta),\theta) v_l\\ &- \frac{\partial^2 A}{\partial \theta_l \partial u}(S(\theta),\theta) v_i - \frac{\partial^2 A}{\partial \theta_i \partial \theta_l}(S(\theta),\theta)\,,\end{aligned}$$

$\rightsquigarrow$ requires solutions of n_θ linear models for gradient and $n_\theta(n_\theta+1)/2$ linear models for Hessian!

Three Ways of Avoiding (1st and) 2nd Order Sensitivities

- adjoint approach
- retraction
- all-at-once formulation

Adjoint Approach

Defining $p = P(\theta)$ as the solution of the adjoint equation

$$\frac{\partial A}{\partial u}(S(\theta), \theta)^* p = -\frac{\partial J}{\partial u}(S(\theta), \theta)\,.$$

we obtain

$$\frac{\partial j}{\partial \theta_i}(\theta) = \frac{\partial J}{\partial \theta_i}(S(\theta), \theta) + \langle \frac{\partial A}{\partial \theta_i}(S(\theta), \theta), P(\theta) \rangle_W \,,$$

Adjoint Approach

Defining $p = P(\theta)$ as the solution of the adjoint equation

$$\frac{\partial A}{\partial u}(S(\theta),\theta)^* p = -\frac{\partial J}{\partial u}(S(\theta),\theta)\,.$$

we obtain

$$\frac{\partial j}{\partial \theta_i}(\theta) = \frac{\partial J}{\partial \theta_i}(S(\theta),\theta) + \langle \frac{\partial A}{\partial \theta_i}(S(\theta),\theta), P(\theta)\rangle_W\ ,$$

Defining z_l, w_l as the solutions of

$$\begin{aligned}\frac{\partial A}{\partial u}(S(\theta),\theta) z_l &= -\ \frac{\partial A}{\partial \theta_l}(S(\theta),\theta)\\ \frac{\partial A}{\partial u}(S(\theta),\theta)^* w_l &= -\ \frac{\partial J}{\partial \theta_l}(S(\theta),\theta) - \frac{\partial^2 J}{\partial u^2}(S(\theta),\theta)^* z_l - \dots\end{aligned}$$

we arrive at

$$\begin{aligned}\frac{\partial^2 j}{\partial \theta_i \theta_l}(\theta) = \frac{\partial^2 J}{\partial \theta_i \theta_l}(S(\theta),\theta) + \frac{\partial^2 J}{\partial \theta_i \partial u}(S(\theta),\theta) z_l + \langle \frac{\partial A}{\partial \theta_i}(S(\theta),\theta), w_l\rangle_W\\ + \langle \frac{\partial^2 A}{\partial \theta_i \theta_l}(S(\theta),\theta) + \frac{\partial^2 A}{\partial \theta_i u}(S(\theta),\theta) z_l, P(\theta)\rangle_W\end{aligned}$$

Adjoint Approach

Defining $p = P(\theta)$ as the solution of the adjoint equation

$$\frac{\partial A}{\partial u}(S(\theta),\theta)^* p = -\frac{\partial J}{\partial u}(S(\theta),\theta)\,.$$

we obtain

$$\frac{\partial j}{\partial \theta_i}(\theta) = \frac{\partial J}{\partial \theta_i}(S(\theta),\theta) + \langle \frac{\partial A}{\partial \theta_i}(S(\theta),\theta), P(\theta)\rangle_W\,,$$

Defining z_l, w_l as the solutions of

$$\begin{aligned}\frac{\partial A}{\partial u}(S(\theta),\theta)z_l &= -\,\frac{\partial A}{\partial \theta_l}(S(\theta),\theta)\\ \frac{\partial A}{\partial u}(S(\theta),\theta)^* w_l &= -\,\frac{\partial J}{\partial \theta_l}(S(\theta),\theta) - \frac{\partial^2 J}{\partial u^2}(S(\theta),\theta)^* z_l - \ldots\end{aligned}$$

we arrive at

$$\begin{aligned}\frac{\partial^2 j}{\partial \theta_i \theta_l}(\theta) = {} & \frac{\partial^2 J}{\partial \theta_i \theta_l}(S(\theta),\theta) + \frac{\partial^2 J}{\partial \theta_i \partial u}(S(\theta),\theta)z_l + \langle \frac{\partial A}{\partial \theta_i}(S(\theta),\theta), w_l\rangle_W\\ & + \langle \frac{\partial^2 A}{\partial \theta_i \theta_l}(S(\theta),\theta) + \frac{\partial^2 A}{\partial \theta_i u}(S(\theta),\theta)z_l, P(\theta)\rangle_W\end{aligned}$$

$\rightsquigarrow$ requires solutions of 1 linear model for gradient and $2n_\theta$ linear models for Hessian!

Retraction

Skip second order sensitivity term in $\frac{\partial^2 j}{\partial \theta_i \partial \theta_l}$

$\rightsquigarrow$ approximation of Hessian by positive definite FIM $\widetilde{H}(\hat{\theta}_c)$;

Introduce retraction term to damp resulting error

$$\begin{pmatrix} \widetilde{H}(\theta_c) & \nabla_\theta q(\theta_c) \\ \nabla_\theta q(\theta_c)^T & 0 \end{pmatrix} \begin{pmatrix} \dot{\theta}_c \\ \dot{\lambda}_c \end{pmatrix} = \begin{pmatrix} -\gamma(\nabla_\theta j(\theta_c) + \lambda_c \nabla_\theta q(\theta_c)) \\ 1 \end{pmatrix} \tag{3}$$

retraction works due to strict monotonicity of $(\nabla_\theta j(\theta_c) + \lambda_c \nabla_\theta q(\theta_c))$ (second order sufficient condition) and implies exponential (w.r.t. c) decay of error.

Retraction

Skip second order sensitivity term in $\frac{\partial^2 j}{\partial\theta_i \partial\theta_l}$
$\rightsquigarrow$ approximation of Hessian by positive definite FIM $\widetilde{H}(\hat{\theta}_c)$;
Introduce retraction term to damp resulting error

$$\begin{pmatrix} \widetilde{H}(\theta_c) & \nabla_\theta q(\theta_c) \\ \nabla_\theta q(\theta_c)^T & 0 \end{pmatrix} \begin{pmatrix} \dot{\theta}_c \\ \dot{\lambda}_c \end{pmatrix} = \begin{pmatrix} -\gamma(\nabla_\theta j(\theta_c) + \lambda_c \nabla_\theta q(\theta_c)) \\ 1 \end{pmatrix} \tag{3}$$

retraction works due to strict monotonicity of $(\nabla_\theta j(\theta_c) + \lambda_c \nabla_\theta q(\theta_c))$ (second order sufficient condition) and implies exponential (w.r.t. c) decay of error.

optimization versus integration:

1 step explicit Euler for (3) with $\gamma := \frac{1}{c - q(\theta_c)}$

$= 1$ step Gauss-Newton for $\begin{cases} \min_{\theta \in \Theta} j(\theta) \\ \text{s.t. } q(\theta) = c. \end{cases}$

All-At-Once Formulation

$$\mathrm{pl}_q(c) = \min_{u \in V, \theta \in \Theta} J(u, \theta)$$

$$\text{s.t. } A(u, \theta) = 0 \text{ and } Q(u, \theta) = c$$

Lagrange function $\mathcal{L}(u, \theta, p) = J(u, \theta) + \langle A(u, \theta), p \rangle + \lambda(Q(u, \theta) - c)$

1st order optimality conditions $\begin{cases} \nabla_\theta \mathcal{L}(u_c, \theta_c, p_c) + \lambda_c \nabla_\theta Q(u_c, \theta_c) = 0 \\ \nabla_u \mathcal{L}(u_c, \theta_c, p_c) + \lambda_c \nabla_u Q(u_c, \theta_c) = 0 \\ \nabla_p \mathcal{L}(u_c, \theta_c, p_c) = 0 \\ Q(u_c, \theta_c) = c \end{cases}$

Differentiate with respect to c to derive a DAE system ($\dot{} = \frac{d}{dc}$)

$$\begin{pmatrix} \nabla^2_\theta \mathcal{L} + \lambda_c \nabla^2_\theta Q & \nabla_u \nabla_\theta \mathcal{L} + \lambda_c \nabla_u \nabla_\theta Q & \nabla_p \nabla_\theta \mathcal{L} & \nabla_\theta Q \\ \nabla_\theta \nabla_u \mathcal{L} + \lambda_c \nabla_\theta \nabla_u Q & \nabla^2_u \mathcal{L} + \lambda_c \nabla^2_u Q & \nabla_p \nabla_u \mathcal{L} & \nabla_u Q \\ \nabla_\theta \nabla_p \mathcal{L} & \nabla_u \nabla_p \mathcal{L} & 0 & 0 \\ \nabla_\theta Q^T & \nabla_u Q^T & 0 & 0 \end{pmatrix} \begin{pmatrix} \dot{\theta}_c \\ \dot{u}_c \\ \dot{p}_c \\ \dot{\lambda}_c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

All-At-Once Formulation

$$\mathrm{pl}_q(c) = \min_{u \in V, \theta \in \Theta} J(u,\theta)$$
$$\text{s.t. } A(u,\theta) = 0 \text{ and } Q(u,\theta) = c$$

Lagrange function $\mathcal{L}(u,\theta,p) = J(u,\theta) + \langle A(u,\theta), p\rangle + \lambda(Q(u,\theta) - c)$

1st order optimality conditions
$$\begin{cases} \nabla_\theta \mathcal{L}(u_c,\theta_c,p_c) + \lambda_c \nabla_\theta Q(u_c,\theta_c) = 0 \\ \nabla_u \mathcal{L}(u_c,\theta_c,p_c) + \lambda_c \nabla_u Q(u_c,\theta_c) = 0 \\ \nabla_p \mathcal{L}(u_c,\theta_c,p_c) = 0 \\ Q(u_c,\theta_c) = c \end{cases}$$

Differentiate with respect to c to derive a DAE system ($\dot{} = \frac{d}{dc}$)

$$\begin{pmatrix} \nabla_\theta^2 \mathcal{L} + \lambda_c \nabla_\theta^2 Q & \nabla_u \nabla_\theta \mathcal{L} + \lambda_c \nabla_u \nabla_\theta Q & \nabla_p \nabla_\theta \mathcal{L} & \nabla_\theta Q \\ \nabla_\theta \nabla_u \mathcal{L} + \lambda_c \nabla_\theta \nabla_u Q & \nabla_u^2 \mathcal{L} + \lambda_c \nabla_u^2 Q & \nabla_p \nabla_u \mathcal{L} & \nabla_u Q \\ \nabla_\theta \nabla_p \mathcal{L} & \nabla_u \nabla_p \mathcal{L} & 0 & 0 \\ \nabla_\theta Q^T & \nabla_u Q^T & 0 & 0 \end{pmatrix} \begin{pmatrix} \dot\theta_c \\ \dot u_c \\ \dot p_c \\ \dot\lambda_c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

huge system, no sensitivities!

All-At-Once Formulation

$$\mathrm{pl}_q(c) = \min_{u \in V, \theta \in \Theta} J(u, \theta)$$

$$\text{s.t. } A(u,\theta) = 0 \text{ and } Q(u,\theta) = c$$

Lagrange function $\mathcal{L}(u,\theta,p) = J(u,\theta) + \langle A(u,\theta), p \rangle + \lambda(Q(u,\theta) - c)$

1st order optimality conditions
$$\begin{cases} \nabla_\theta \mathcal{L}(u_c,\theta_c,p_c) + \lambda_c \nabla_\theta Q(u_c,\theta_c) = 0 \\ \nabla_u \mathcal{L}(u_c,\theta_c,p_c) + \lambda_c \nabla_u Q(u_c,\theta_c) = 0 \\ \nabla_p \mathcal{L}(u_c,\theta_c,p_c) = 0 \\ Q(u_c,\theta_c) = c \end{cases}$$

Differentiate with respect to c to derive a DAE system ($\dot{} = \frac{d}{dc}$)

$$\begin{pmatrix} \nabla_\theta^2 \mathcal{L} + \lambda_c \nabla_\theta^2 Q & \nabla_u \nabla_\theta \mathcal{L} + \lambda_c \nabla_u \nabla_\theta Q & \nabla_p \nabla_\theta \mathcal{L} & \nabla_\theta Q \\ \nabla_\theta \nabla_u \mathcal{L} + \lambda_c \nabla_\theta \nabla_u Q & \nabla_u^2 \mathcal{L} + \lambda_c \nabla_u^2 Q & \nabla_p \nabla_u \mathcal{L} & \nabla_u Q \\ \nabla_\theta \nabla_p \mathcal{L} & \nabla_u \nabla_p \mathcal{L} & 0 & 0 \\ \nabla_\theta Q^T & \nabla_u Q^T & 0 & 0 \end{pmatrix} \begin{pmatrix} \dot\theta_c \\ \dot u_c \\ \dot p_c \\ \dot\lambda_c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

equivalence of integration and optimization like in reduced setting.

All-At-Once Formulation

$$\mathrm{pl}_q(c) = \min_{u \in V, \theta \in \Theta} J(u, \theta)$$
$$\text{s.t. } A(u,\theta) = 0 \text{ and } Q(u,\theta) = c$$

Lagrange function $\mathcal{L}(u, \theta, p) = J(u,\theta) + \langle A(u,\theta), p\rangle + \lambda(Q(u,\theta) - c)$

1st order optimality conditions
$$\begin{cases} \nabla_\theta \mathcal{L}(u_c, \theta_c, p_c) + \lambda_c \nabla_\theta Q(u_c, \theta_c) = 0 \\ \nabla_u \mathcal{L}(u_c, \theta_c, p_c) + \lambda_c \nabla_u Q(u_c, \theta_c) = 0 \\ \nabla_p \mathcal{L}(u_c, \theta_c, p_c) = 0 \\ Q(u_c, \theta_c) = c \end{cases}$$

Differentiate with respect to c to derive a DAE system ($\dot{} = \frac{d}{dc}$)

$$\begin{pmatrix} \nabla_\theta^2 \mathcal{L} + \lambda_c \nabla_\theta^2 Q & \nabla_u \nabla_\theta \mathcal{L} + \lambda_c \nabla_u \nabla_\theta Q & \nabla_p \nabla_\theta \mathcal{L} & \nabla_\theta Q \\ \nabla_\theta \nabla_u \mathcal{L} + \lambda_c \nabla_\theta \nabla_u Q & \nabla_u^2 \mathcal{L} + \lambda_c \nabla_u^2 Q & \nabla_p \nabla_u \mathcal{L} & \nabla_u Q \\ \nabla_\theta \nabla_p \mathcal{L} & \nabla_u \nabla_p \mathcal{L} & 0 & 0 \\ \nabla_\theta Q^T & \nabla_u Q^T & 0 & 0 \end{pmatrix} \begin{pmatrix} \dot{\theta}_c \\ \dot{u}_c \\ \dot{p}_c \\ \dot{\lambda}_c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

retraction approach analogous to reduced setting.

Numerical Tests

gradient formation on surface of fission yeast cell:

$$
\begin{aligned}
&u_t = Du_{xx} - \alpha u^2 + \frac{J}{\sqrt{2\pi}\rho} e^{-x^2/2\rho^2} && \text{on } (0,T)\times(-L,L) \\
&\frac{\partial u}{\partial \nu} = 0 && \text{on } (0,T)\times\{-L,L\} \\
&u = 0 && \text{on} \{t=0\}\times(-L,L)
\end{aligned}
$$

$u \ldots$ concentration of the protein Pom1p, $D \ldots$ diffusion coefficient, $\alpha \ldots$ dimerisation rate, $J \ldots$ influx rate, $\rho \ldots$ source width

measurements:

Concentration profile: $(-L,L) = \Omega_1 \dot\cup \cdots \dot\cup \Omega_{60}$

$G_k(u,\theta) = s_1 \int_{\Omega_k} u(t=100,x)dx \quad \text{for} \quad k = 1,\ldots,60,$

Time course:

$G_{60+k}(u,\theta) = s_2 \int_{-L}^{L} u(t_k,x)dx \quad \text{for} \quad k = 1,\ldots,10,$

Quantification:

$G_{71}(u,\theta) = \int_{-L}^{L} u(t=100,x)dx.$

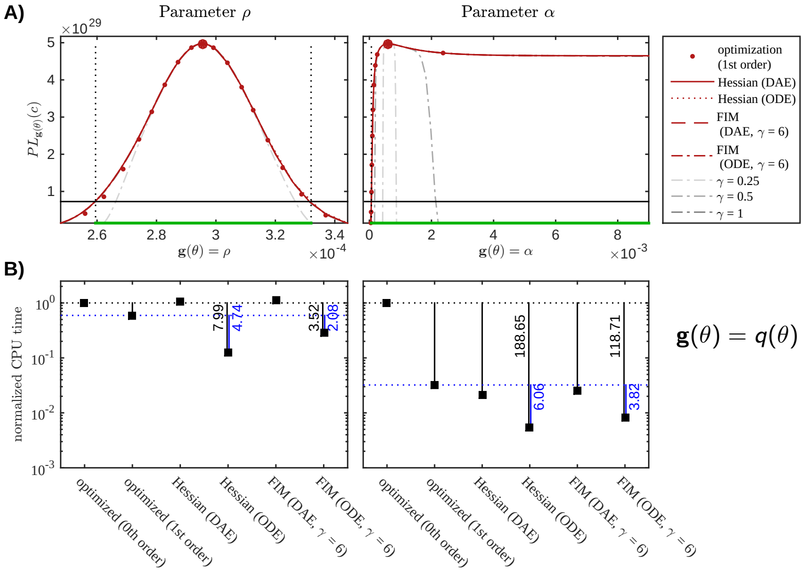
Implementation

- DAE/ODE solution, parameter identification and profile computation implemented by Sabrina Hroß and Jan Hasenauer (Helmholtz Zentrum München) in AMICI and PESTO
  https://github.com/ICB-DCM/AMICI
  https://github.com/ICB-DCM/PESTO
- simulated data with additive, normally distributed noise
- $\theta = (D, \alpha, J, \rho, s_1, s_2)$

Comparison

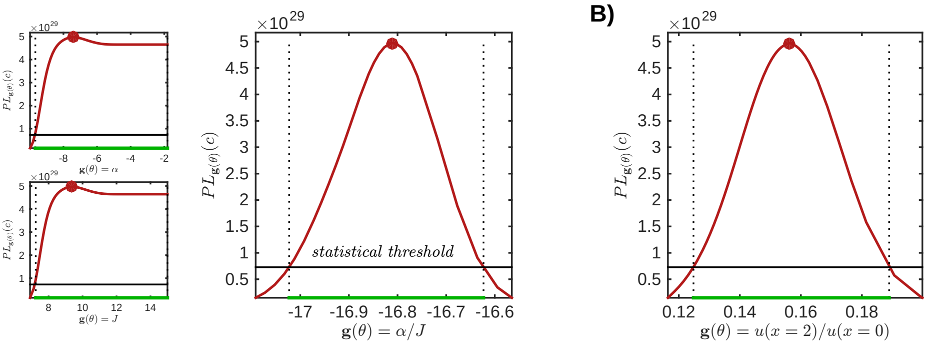
- optimization based with 0 order starting values
- optimization based with 1^{st} order starting values
- integration based with full Hessian via DAE
- integration based with full Hessian via ODE
- integration based with FIM and retraction via DAE
- integration based with FIM and retraction via ODE

Numerical Results



$$\mathbf{g}(\theta) = q(\theta)$$

Numerical Results



$$\mathbf{g}(\theta) = q(\theta)$$

Conclusions and Outlook

- likelihood profiles allow to quantify uncertainty in parameter estimation due to measurement noise

Conclusions and Outlook

- likelihood profiles allow to quantify uncertainty in parameter estimation due to measurement noise
- computation via solution of an evolutionary system with artificial time c

Conclusions and Outlook

- likelihood profiles allow to quantify uncertainty in parameter estimation due to measurement noise
- computation via solution of an evolutionary system with artificial time c
- examples of time dependent models; as well applicable to stationary models

Conclusions and Outlook

- likelihood profiles allow to quantify uncertainty in parameter estimation due to measurement noise
- computation via solution of an evolutionary system with artificial time c
- examples of time dependent models; as well applicable to stationary models
- (much) less expensive than Bayesian credible intervals (5 min vs 12 hrs)

Conclusions and Outlook

- likelihood profiles allow to quantify uncertainty in parameter estimation due to measurement noise
- computation via solution of an evolutionary system with artificial time c
- examples of time dependent models; as well applicable to stationary models
- (much) less expensive than Bayesian credible intervals (5 min vs 12 hrs)

→ infinite dimensional parameters (e.g., spatially varying coefficients, initial conditions) in PDEs

Likelihood profiles for infinite dimensional parameters θ

$$\mathrm{pl}_q(c) = \min_{u \in V, \theta \in \Theta} J(u, \theta)$$
$$\text{s.t. } A(u, \theta) = 0 \text{ and } Q(u, \theta) = c$$

infinite dimensional inverse problem $\rightsquigarrow$ ill-posed $\rightsquigarrow$ regularize

Likelihood profiles for infinite dimensional parameters θ

$$\mathrm{pl}_q(c) = \min_{u \in V, \theta \in \Theta} J(u, \theta)$$

$$\text{s.t. } A(u, \theta) = 0 \text{ and } Q(u, \theta) = c$$

infinite dimensional inverse problem $\rightsquigarrow$ ill-posed $\rightsquigarrow$ regularize

$$\mathrm{pl}_q(c) = \min_{u \in V, \theta \in \Theta} J(u, \theta)$$

$$\text{s.t. } A(u, \theta) = 0 \text{ and } Q(u, \theta) = c \text{ and } \mathcal{R}(u, \theta) \le \rho$$

Likelihood profiles for infinite dimensional parameters θ

$$\mathrm{pl}_q(c) = \min_{u \in V, \theta \in \Theta} J(u,\theta)$$

$$\text{s.t. } A(u,\theta) = 0 \text{ and } Q(u,\theta) = c$$

infinite dimensional inverse problem $\rightsquigarrow$ ill-posed $\rightsquigarrow$ regularize

$$\mathrm{pl}_q(c) = \min_{u \in V, \theta \in \Theta} J(u,\theta)$$

$$\text{s.t. } A(u,\theta) = 0 \text{ and } Q(u,\theta) = c \text{ and } \mathcal{R}(u,\theta) \leq \rho$$

- $\mathcal{R}$... regularization functional, e.g., norm;
  ρ... regularization parameter
- Ivanov regularization (method of quasi solutions)

Likelihood profiles for infinite dimensional parameters θ

$$\mathrm{pl}_q(c) = \min_{u\in V,\theta\in\Theta} J(u,\theta)$$
$$\text{s.t. } A(u,\theta) = 0 \text{ and } Q(u,\theta) = c$$

infinite dimensional inverse problem $\rightsquigarrow$ ill-posed $\rightsquigarrow$ regularize

$$\mathrm{pl}_q(c) = \min_{u\in V,\theta\in\Theta} J(u,\theta)$$
$$\text{s.t. } A(u,\theta) = 0 \text{ and } Q(u,\theta) = c \text{ and } \mathcal{R}(u,\theta) \leq \rho$$

- $\mathcal{R}$... regularization functional, e.g., norm;
  ρ... regularization parameter
- Ivanov regularization (method of quasi solutions)
- as compared to Tikhonov, avoids introducing artificial identifiability

Likelihood profiles for infinite dimensional parameters θ

$$\mathrm{pl}_q(c) = \min_{u\in V,\theta\in\Theta} J(u,\theta)$$
$$\text{s.t. } A(u,\theta) = 0 \text{ and } Q(u,\theta) = c$$

infinite dimensional inverse problem $\rightsquigarrow$ ill-posed $\rightsquigarrow$ regularize

$$\mathrm{pl}_q(c) = \min_{u\in V,\theta\in\Theta} J(u,\theta)$$
$$\text{s.t. } A(u,\theta) = 0 \text{ and } Q(u,\theta) = c \text{ and } \mathcal{R}(u,\theta) \leq \rho$$

- $\mathcal{R}$... regularization functional, e.g., norm;
  ρ... regularization parameter
- Ivanov regularization (method of quasi solutions)
- as compared to Tikhonov, avoids introducing artificial identifiability
- convergence as $\eta \to 0$ with a priori or a posteriori choice of ρ

Likelihood profiles for infinite dimensional parameters θ

$$\mathrm{pl}_q(c) = \min_{u \in V, \theta \in \Theta} J(u,\theta)$$
$$\text{s.t. } A(u,\theta) = 0 \text{ and } Q(u,\theta) = c$$

infinite dimensional inverse problem $\rightsquigarrow$ ill-posed $\rightsquigarrow$ regularize

$$\mathrm{pl}_q(c) = \min_{u \in V, \theta \in \Theta} J(u,\theta)$$
$$\text{s.t. } A(u,\theta) = 0 \text{ and } Q(u,\theta) = c \text{ and } \mathcal{R}(u,\theta) \le \rho$$

- $\mathcal{R}$... regularization functional, e.g., norm; ρ... regularization parameter
- Ivanov regularization (method of quasi solutions)
- as compared to Tikhonov, avoids introducing artificial identifiability
- convergence as $\eta \to 0$ with a priori or a posteriori choice of ρ
- computational approaches and equivalence results for likelihood profiles carry over

Thank you for your attention!