F. Kappel
Institute for Mathematics and Scientific Computing, University of Graz, Graz, Austria

M. Munir
Government Postgraduate College No. 1, Abbottabad, Pakistan

A new approach to optimal design problems

Abstract

In this contribution we present an approach to optimal design problems based on sensitivity functions and on generalized sensitivity functions. Roughly speaking, generalized sensitivity functions as introduced in [6] describe the sensitivity of parameter estimates with respect to variations of available measurements for model outputs. The approach developed in [2] applies to all design criteria, which are continuous functions of the Fisher information matrix. The theoretical background is provided by minimization/maximization of continuous functions on the space of probability measures on a given time interval endowed with the Prohorov metric. Probability measures describe generalized measurement procedures. In order to obtain concrete information on the number and distribution of time instants, where measurements should be taken, one has to approximate the "optimal measurement distribution" by discrete probability measures.

Key words: optimal design, sensitivity function, optimal measurement distribution

1. Introduction

Assume that

$$y(t) = f(t;\theta), \quad t \geq 0, \tag{1}$$

is the output of a dynamical system modeling some real world process, where $\theta \in \mathbb{R}^p$ is the vector of parameters. As usual, during the modeling process a considerable number of parameters are picked up, which in general cannot be determined a priori by direct measurements. The following two important tasks have to be accomplished in each modeling project:

a) We have to obtain sufficiently reliable estimates for the parameters θ on the basis of measurements for the output $y(\cdot)$.
b) We have to provide guidelines concerning the experiments which have to be conducted in order to get the desired measurements. These guidelines have to address questions like how many measurements one has to take at what time instants in order to get parameter estimates of a given quality. Since measurements can be quite costly, it is also important not to take unnecessary many measurements.

In this contribution we shall only address problems related to b). Our approach will be based on the so called generalized sensitivity functions first introduced in [6]. Roughly speaking, generalized sensitivity functions describe the sensitivity of the parameter estimate with respect to variations of the nominal parameter values. The theory is developed under the following basic assumptions (additional assumptions will be introduced below):

A measurement at a time instant $t \in [0, T]$, where $T > 0$ is fixed, is given as

$$\xi(t) = f(t, \theta_0) + \varepsilon(t), \tag{2}$$

where θ_0 is the 'true' parameter vector (also called the nominal parameter vector) and $\varepsilon(\cdot)$ is a representation of a noise process $\mathcal{E}(\cdot)$, where the random variables $\mathcal{E}(t)$, $t \in [0, T]$, are independent and identically distributed with expected value zero and variance $\sigma^2(t)$, which is independent of the parameters, but may depend on time. It is clear that the measurements $\xi(t)$ are realizations of random variables $\Xi(t)$ with expected value $f(t, \theta_0)$ and variance $\sigma^2(t)$. We assume that the output function f is continuous in t and has continuous second order derivatives with respect to θ.

We shall study the parameter identification problem for nominal parameters in a neighborhood $\mathcal{U}$ of θ_0. In Section 2 we introduce generalized sensitivity functions in an abstract setting following the approach given in [2]. We refer also to this paper and to [3], [6] for interpretations of the properties of generalized sensitivity functions concerning information on parameters contained in the given measurements. In Section 3 we indicate shortly the importance of generalized sensitivity functions for experimental design, again referring for details to [2]. Finally, in Section 4 we present new results connecting linear behavior of generalized sensitivity functions with identifiability of parameters.

2. Abstract measurement procedures and generalized sensitivity functions

Let $m(t)$ denote the measurement density at time $t \in [0,T]$, i.e., $m(t)$ is the number of measurements per unit time at t. We assume that m is continuous on $[0,T]$. By definition of m, $\int_0^t m(\tau)d\tau$ is the number of measurements in the interval $[0,t]$. In order to define the least squares error $J(\xi,\theta)$ for a given measurement procedure $\xi(\cdot,\theta_0)$ and a nominal parameter vector $\theta \in \mathcal{U}$ we start with the weighted error for one measurement at time t, which is given by $\sigma(t)^{-2}\big(\xi(t,\theta_0) - f(t;\theta)\big)^2$.

For mesh points $t_i = i\Delta t$, $i = 0,\dots,N$, $\Delta t = T/N$. Then the number of measurements in the time intervals $[t_i, t_i+\Delta t]$ is given by $m(t_i^*)\Delta t$ with $t_i^* \in [t_i, t_i+\Delta t]$. The average of the weighted error on $[t_i, t_i+\Delta t]$ is given by $\sigma(t_i^{**})^{-2}\big(\xi(t_i^{**},\theta_0) - f(t_i^{**};\theta)\big)^2$, where $t_i^{**} \in [t_i, t_i+\Delta t]$. Then the weighted errors in the intervals $[t_i, t_i+\Delta t]$ are given by

$$\frac{m(t_i^*)}{\sigma(t_i^{**})^2}\big(\xi(t_i^{**},\theta_0) - f(t_i^{**};\theta)\big)^2\Delta t, \quad i = 0,\dots,N-1.$$

Taking $\Delta t \to 0$ we get

$$J(\xi,\theta) = \int_0^T \frac{m(t)}{\sigma(t)^2}\big(\xi(t,\theta_0) - f(t;\theta)\big)^2\,dt.$$

This integral can be viewed as an integral with respect to an absolutely continuous measure P on $[0,T]$ with density $m(\cdot)$. This is the motivation to consider error functionals of the form

$$J(\xi,\theta) = \int_0^T \frac{1}{\sigma(t)^2}\big(\xi(t,\theta_0) - f(t;\theta)\big)^2\,dP,$$

where P is a general measure on $[0,T]$. Of course, we can restrict ourselves to probability measures on $[0,T]$, because the minima of J do not change, if J is multiplied by a positive constant. If, for points $t_0 < t_1 < \cdots < t_N$, we take

$$P = \sum_{i=0}^N \delta_{t_i},$$

where δ_a denotes the delta distribution with support $\{a\}$, we get

$$J(\xi,\theta) = \sum_{i=0}^N \frac{1}{\sigma(t_i)^2}\big(\xi(t_i,\theta_0) - f(t_i;\theta)\big)^2.$$

This is exactly the type of quadratic error functionals, which one would consider for real measurement procedures (and was the basis in [6]).

For a nominal parameter vector $\theta \in \mathcal{U}$ we denote the minimizer of J by $\hat{\theta}(\theta)$, i.e.,

$$\hat{\theta}(\theta) = \operatorname*{argmin}_{\tau} J(\xi, \tau), \quad \theta \in \mathcal{U}.$$

Here we assume that,

(ii) for any $\theta \in \mathcal{U}$, the functional J has a unique minimum in a neighborhood of θ, i.e., we have unique local identifiability, and that $\mathsf{E}\,\hat{\theta}(\theta) = \theta$ for $\theta \in \mathcal{U}$, i.e., the estimation procedure is unbiased.

A necessary condition, which $\hat{\theta}(\theta)$ has to satisfy, is

$$\nabla_\theta J(\xi, \hat{\theta}(\theta)) = 0, \quad \theta \in \mathcal{U}. \tag{3}$$

For each realization $\xi(\cdot, \theta)$ of $\Xi(\cdot, \theta)$, i.e., for each realization ε of $\mathcal{E}$, we get a specific value for $\hat{\theta}(\theta)$, which is a realization of a random variable we denote by $\hat{\Theta}(\theta)$.

We are really interested in the dependence of $\hat{\theta}(\theta)$ on θ. Therefore we compute the Jacobian

$$\frac{\partial \hat{\theta}}{\partial \theta}(\theta), \quad \theta \in \mathcal{U}.$$

In order to do so we differentiate (3) with respect to θ and get

$$0 = \nabla^2_{\theta\theta} J(\xi, \hat{\theta}(\theta)) \frac{\partial \hat{\theta}}{\partial \theta}(\theta) + \nabla^2_{\theta\xi} J(\xi, \hat{\theta}(\theta)) \frac{\partial \xi}{\partial \theta}(\theta).$$

Assuming that $\nabla^2_{\theta\theta} J(\xi, \hat{\theta}(\theta))$ is nonsingular in $\mathcal{U}$ this yields

$$\frac{\partial \hat{\theta}}{\partial \theta}(\theta) = -\Big(\nabla^2_{\theta\theta} J(\xi, \hat{\theta}(\theta))\Big)^{-1} \nabla^2_{\theta\xi} J(\xi, \hat{\theta}(\theta)) \frac{\partial \xi}{\partial \theta}(\cdot, \theta), \quad \theta \in \mathcal{U}. \tag{4}$$

From the definition of $J(\xi, \theta)$ we get

$$\begin{aligned} \nabla^2_{\theta\theta} J(\xi, \theta) = 2 &\int_0^T \frac{1}{\sigma(s)^2} \nabla_\theta f(s;\theta)^{\mathsf{T}} \nabla_\theta f(s;\theta)\, dP \\ &- 2 \int_0^T \frac{1}{\sigma(s)^2} \big(\xi(s, \theta_0) - f(s;\theta)\big) \nabla^2_{\theta\theta} f(s;\theta)\, dP. \end{aligned}$$

Analogously we get

$$\nabla^2_{\theta\xi} J(\xi,\theta)\frac{\partial \xi}{\partial \theta}(\cdot,\theta) = 2\int_0^T \frac{1}{\sigma(s)^2}\nabla_\theta f(s;\theta)^{\mathsf{T}}\frac{\partial \xi}{\partial \theta}(s,\theta)\,dP.$$

We take these quantities at the expected values of the random variables corresponding to their arguments. Observing (2), $\mathsf{E}(\mathcal{E}(t)) = 0$, i.e., $\mathsf{E}(\Xi(t,\theta)) = f(t;\theta)$, and $\mathsf{E}(\hat\theta(\theta)) = \theta$ we obtain

$$\big(\nabla^2_{\theta\theta} J\big)(f(\cdot;\theta),\theta) = 2\int_0^T \frac{1}{\sigma(s)^2}\nabla_\theta f(s;\theta)^{\mathsf{T}}\nabla_\theta f(s;\theta)\,dP$$

and

$$\big(\nabla^2_{\theta\xi} J\big)(f(\cdot;\theta),\theta)\frac{\partial \xi}{\partial \theta}(\cdot,\theta) = -2\int_0^T \frac{1}{\sigma(s)^2}\nabla_\theta f(s;\theta)^{\mathsf{T}}\nabla_\theta f(s;\theta)\,dP.$$

Using this in (4) we get

$$\frac{\partial \hat\theta}{\partial \theta}(\theta) = \mathcal{F}(T)^{-1}\int_0^T \frac{1}{\sigma(s)^2}\nabla_\theta f(s;\theta)^{\mathsf{T}}\nabla_\theta f(s;\theta)\,dP \equiv I, \qquad (5)$$

where

$$\mathcal{F}(T,\theta) = \int_0^T \frac{1}{\sigma(s)^2}\nabla_\theta f(s;\theta)^{\mathsf{T}}\nabla_\theta f(s;\theta)\,dP \qquad (6)$$

is the so called Fisher information matrix. Equation (4) reflects the assumption that the estimates are unbiased. If we want to indicate the measure P used in (6), then we also write $\mathcal{F}_P(T,\theta)$.

In order to obtain information at what time instants measurements contain more information on the parameters than measurement at other time instants we consider the following procedure. The value of the cost functional J still depends on all measurements on the interval $[0,T]$. However, the variation of the parameter estimate is only determined with respect to measurements up to time $t \in [0,T]$. Computations analogous to those leading to (5) give the following result (we set denote by $\mathcal{G}(t,\theta)$ the Jacobian $\partial\hat\theta/\partial\theta$ in this case):

$$\mathcal{G}(t,\theta) = \mathcal{F}(T,\theta)^{-1}\mathcal{F}(t,\theta), \quad \theta \in \mathcal{U},\ 0 \le t \le T. \qquad (7)$$

The diagonal elements of $\mathcal{G}$ are called the generalized sensitivity functions with respect to the parameters $\theta_1,\dots,\theta_p$,

$$g_i(t,\theta) = \big(\mathcal{G}(t,\theta)\big)_{i,i}, \quad \theta \in \mathcal{U},\ 0 \le t \le T,\ i = 1,\dots,p.$$

In case P equals the Lebesgue measure on $[0,T]$ we call $\mathcal{G}$ the continuous generalized sensitivity matrix and denote it by $\mathcal{G}_c$.

3. The Fisher information matrix and experimental design criteria

Most of the optimal design criteria are based on the Fisher information matrix. Here we mention some of the most frequently used criteria (see also [4], [5]):

a) D-optimal design requires to maximize $\det \mathcal{F}(T,\hat{\theta})$.

b) E-optimal design requires to maximize the smallest eigenvalue $\lambda_{\min}\big(\mathcal{F}(T,\hat{\theta})\big)$ of the Fisher information matrix $\mathcal{F}(T,\hat{\theta})$.

c) c-optimal design requires to minimize the variance of a given linear combination $c^{\mathsf{T}}\theta$ of the parameters (for instance, if we choose $c = e_i$, then we want to minimize the variance of $\hat{\theta}_i$). This is asymptotically equivalent to minimizing $c^{\mathsf{T}}\mathcal{F}(T,\theta)^{-1}c$.

These design criteria can all be incorporated in the following formulation. Let $\mathcal{P}(0,T)$ denote the space of probability measures on $[0,T]$ and assume that $\mathcal{J} : \mathbb{R}^{p\times p} \to \mathbb{R}^+$ is continuous. The optimal measurement procedure is given by the probability measure $\hat{P} \in \mathcal{P}(0,T)$ such that

$$\mathcal{J}(\mathcal{F}_{\hat{P}}(T,\theta)) = \min_{P\in\mathcal{P}(0,T)} \mathcal{J}(\mathcal{F}_P(T,\theta).$$

In order to solve this problem we introduce the following abstract setting. Let (Ω,d) be a metric space (in our case $\Omega \in [0,T]$ and $d(x,y) = |x-y|$) and denote by $\mathcal{P}(\Omega)$ the set of all Borel probability measures on Ω. On $\mathcal{P}(\Omega)$ we introduce the Prohorov metric ρ defined by

$$\rho(\pi,\tilde{\pi}) = \inf_{A\subset\Omega \text{ closed}} \big\{\epsilon \mid \pi(A) \le \tilde{\pi}(A^\epsilon) + \epsilon\big\}, \quad \pi,\tilde{\pi} \in \mathcal{P}(\Omega).$$

Here we have set $A^\epsilon = \{y \in \Omega \mid d(y,A) \le \epsilon\}$. We make essential use of the following properties (see [1]):

a) If (Q,d) is a complete metric space, then $(\mathcal{P}(Q),\rho)$ is also a complete metric space.

b) ρ-convergence is equivalent to weak*–convergence in $\mathcal{P}(Q)$.

c) If (Q,d) is compact, then also $(\mathcal{P},\rho)$ is compact.

d) If (Q, d) is separable, then also $(\mathcal{P}, \rho)$ is separable.

Based on these result the following design procedure was developed in [2]:

Step 1. Determine T using the condition number of the Fisher information matrix (choose T such that the condition number decreases fast up to T and slowly after T).

Step 2. Determine the minimum number n of uniformly spaced measurements on $[0, T]$ such that the standard errors SE_k of the estimates for the parameters $\theta_1, \dots \theta_p$ are smaller than a given bound. Here we use the asymptotic formula

$$SE_k = \sqrt{\frac{T}{n}\mathcal{F}^{-1}_{k,k}}, \quad k = 1, \dots, p.$$

Here the Fisher information matrix is computed with the discrete probability measure which corresponds to the uniform mesh.
Step 3. Determine the optimal $\hat{P} \in \mathcal{P}(0, T)$ ($\hat{P}$ exists according to property c) from above).

Step 4. Approximate $\hat{P}$ by discrete probability measures with n atoms (this can be done by property d) from above).

4. Linear behavior of generalized sensitivities and non-identifiability

In this section we take the measure P to be the Lebesgue measure on $[0, T]$. For a fixed $T > 0$ and a nominal parameter vector θ_0 let $\mathcal{G}_c(t; T)$ denote the generalized sensitivity matrix on the interval $[0, T]$,

$$\mathcal{G}_c(t, \theta) = \mathcal{F}_c(T, \theta)^{-1} \int_0^t \nabla_\theta f(\tau; \theta_0)^\mathsf{T} \nabla_\theta f(\tau; \theta_0)\, d\tau, \quad 0 \le t \le T, \tag{8}$$

where the Fisher information matrix $\mathcal{F}_c(T, \theta)$ is given by (6) with P being the Lebesgue measure on $[0, T]$,

$$\mathcal{F}_c(T, \theta) = \int_0^T \nabla_\theta f(\tau; \theta_0)^\mathsf{T} \nabla_\theta f(\tau; \theta_0)\, d\tau. \tag{9}$$

In [2] the following theorem is proved:

Theorem 1. *Assume that* $\det \mathcal{F}_c(T, \theta) \neq 0$. *Then*

$$\det \mathcal{F}_c(T + \Delta, \theta) \neq 0 \quad \textit{and} \quad \left\| \mathcal{F}_c(T + \Delta; \theta)^{-1} \right\|_2 \le \left\| \mathcal{F}_c(T, \theta)^{-1} \right\|_2$$

for all $\Delta \ge 0$, *where* $\|A\|_2 = (\lambda_{\max}(A^* A))^{1/2}$, *the spectral norm of a matrix* A.

4.1. Solutions near an exponentially stable equilibrium

We assume that the output of the system corresponding to $\theta \in \mathcal{U}$ can be written as

$$f(t;\theta_0) = h(\theta_0) + M(t,\theta_0), \quad t \geq 0, \quad t \geq 0,\ \theta \in \mathcal{U}, \tag{10}$$

where h respectively M are smooth functions defined on $\mathcal{U}$ respectively on $[0,\infty) \times \mathcal{U}$. Furthermore, we assume that

$$\|\nabla_\theta M(t,\theta_0)\| \leq c(\theta) e^{-\alpha(\theta)t}, \quad t \geq 0,\ \theta \in \mathcal{U}, \tag{11}$$

with positive functions $c, \alpha : \mathcal{U} \to \mathbb{R}$. Let T be sufficiently large and fix $T_0 \in (0,T)$. Then we get from (8) and (9)

$$\mathcal{G}_c(t,\theta) = \mathcal{F}_c(T,\theta)^{-1}\Big(\mathcal{F}_c(T_0,\theta) + \int_{T_0}^{t} \nabla_\theta f(\tau;\theta_0)^{\mathsf{T}} \nabla_\theta f(\tau;\theta_0)\, d\tau\Big), \quad 0 \leq t \leq T. \tag{12}$$

From equation (10) we get

$$\nabla_\theta f(t;\theta_0)^{\mathsf{T}} \nabla_\theta (f(t;\theta_0) = \nabla_\theta h(\theta_0)^{\mathsf{T}} \nabla_\theta h(\theta_0) + m(t,\theta_0), \quad t \geq 0,$$

where

$$m(t,\theta_0) = 2\nabla_\theta h(\theta_0)^{\mathsf{T}} \nabla_\theta M(t,\theta_0) + \nabla_\theta M(t,\theta_0)^{\mathsf{T}} \nabla_\theta M(t,\theta_0).$$

Assumption (11) implies

$$\|m(t,\theta_0)\| \leq 2c(\theta_0)\|\nabla_\theta h(\theta_0)\| e^{-\alpha(\theta_0)t} + c(\theta_0)^2 e^{-2\alpha(\theta_0)t}, \quad t \geq 0.$$

Consequently we have

$$\int_{T_0}^{t} \nabla_\theta f(\tau;\theta_0)^{\mathsf{T}} \nabla_\theta f(\tau;\theta_0)\, d\tau = (t - T_0)\nabla_\theta h(\theta_0)^{\mathsf{T}} \nabla_\theta h(\theta_0) + \mathcal{B}(t,\theta_0), \quad t \geq T_0,$$

where $\mathcal{B}(t,\theta_0) = \int_{T_0}^{t} m(\tau,\theta_0)\, d\tau$ satisfies the estimate

$$\begin{aligned} \|\mathcal{B}(t,\theta_0)\| &\leq \int_{T_0}^{t} \|m(\tau,\theta_0)\|\, d\tau \\ &\leq 2\frac{c(\theta_0)}{\alpha(\theta_0)}\|\nabla_\theta h(\theta_0)\| e^{-\alpha(\theta_0)T_0} + \frac{c(\theta_0)^2}{2\alpha(\theta_0)} e^{-2\alpha(\theta_0)T_0} \\ &\leq b(\theta_0) e^{-\alpha(\theta_0)T_0}, \quad t \geq T_0, \end{aligned} \tag{13}$$

with

$$b(\theta_0) = 2\frac{c(\theta_0)}{\alpha(\theta_0)}\|\nabla_\theta h(\theta_0)\| + \frac{c(\theta_0)^2}{2\alpha(\theta_0)}e^{-\alpha(\theta_0)T_0}.$$

Equation (12) can be written as

$$\begin{aligned}\mathcal{G}_c(t,\theta) = \mathcal{G}_c(T_0,\theta) + (t-T_0)\mathcal{F}_c(T,\theta)^{-1}\nabla_\theta h(\theta_0)^\mathsf{T}\nabla_\theta h(\theta_0) \\ + \mathcal{F}_c(T,\theta)^{-1}\mathcal{B}(t,\theta_0), \quad T_0 \le t \le T.\end{aligned} \tag{14}$$

Assume that $\det \mathcal{F}(T_1,\theta) \neq 0$ for some fixed $T_1 < T$. Then Theorem 1 and (13) imply

$$\begin{aligned}\big\|\mathcal{F}_c(T,\theta)^{-1}\mathcal{B}(t,\theta_0)\big\|_2 &\le \|\mathcal{F}_c(T,\theta)^{-1}\|_2\|\mathcal{B}(t,\theta_0)\|_2 \\ &\le b(\theta_0)\|\mathcal{F}_c(T_1,\theta)^{-1}\|_2 e^{-\alpha(\theta_0)T_0}, \quad T_0 \le t \le T,\ T_1 \le T.\end{aligned}$$

Given $\epsilon > 0$ we choose $T_0 > 0$ such that

$$b(\theta_0)\|\mathcal{F}_c(T_1,\theta)^{-1}\|_2 e^{-\alpha(\theta_0)T_0} \le \epsilon.$$

Then we have

$$\big\|\mathcal{G}_c(t,\theta) - \mathcal{G}_c(T_0,\theta) - (t-T_0)\mathcal{F}_c(T,\theta)^{-1}\nabla_\theta h(\theta_0)^\mathsf{T}\nabla_\theta h(\theta_0)\big\|_2 \le \epsilon \tag{15}$$

for $T_0 \le t \le T$ and $T_1 \le T$. This shows that the functions $\big(\mathcal{G}_c\big)_{i,j}(t)$, $i,j = 1,\dots,p$, are close to straight lines on $[T_0,T]$ connecting $\big(\mathcal{G}_c\big)_{i,j}(T_0,\theta)$ and $\big(\mathcal{G}_c\big)_{i,j}(T,\theta) = \delta_{i,j}$ (Kronecker delta).

This result remains true if instead of (10) we assume that

$$f(t;\theta_0) = h(\theta_0) + \gamma(t) + M(t;\theta_0), \quad t \ge 0,$$

where $\gamma(\cdot)$ is a given function on $t \ge 0$ not dependent on θ and $M(t,\theta_0)$ satisfies assumption (11).

We illustrate this result for a solution of the so called logistic equation

$$\begin{aligned}\dot{x}(t) &= ax(t) - bx(t)^2, \quad t \ge 0,\\ x(0) &= x_0,\end{aligned} \tag{16}$$

where we have $\theta = (a,b,x_0)$ and $f(t;\theta) = x(t;a,b,x_0)$, $t \ge 0$.

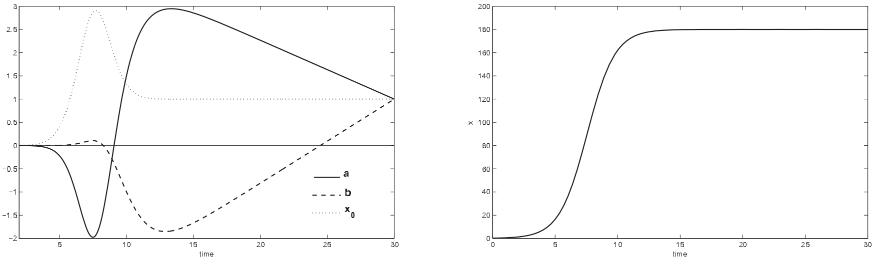

Figure 1: Generalized sensitivities for the three parameters of the logistic equation (16) (left panel) and corresponding solution for $\theta_0 = (0.9, 0.005, 0.2)$ (right panel) on the time interval $[2, 30]$.

4.2. A sufficient condition for ill-conditioning of the parameter identification problem

For $T_0 \in (0, T)$ and $0 \le t \le T$ we get

$$\begin{aligned} \mathcal{G}_c(t,\theta) &= \mathcal{F}_c(T,\theta)^{-1}\Big(\mathcal{F}_c(T_0,\theta) + \int_{T_0}^{t} \nabla_\theta f(\tau;\theta_0)^{\mathsf{T}} \nabla_\theta f(\tau;\theta_0)\, d\tau\Big) \\ &= \mathcal{G}_c(T_0,\theta) + \mathcal{F}_c(T,\theta)^{-1} \int_{T_0}^{t} \nabla_\theta f(\tau;\theta_0)^{\mathsf{T}} \nabla_\theta f(\tau;\theta_0)\, d\tau. \end{aligned} \tag{17}$$

From (8) we get

$$\mathcal{G}_c'(t,\theta) = \mathcal{F}_c(T,\theta)^{-1} \nabla_\theta f(t;\theta_0)^{\mathsf{T}} \nabla_\theta f(t;\theta_0), \quad 0 \le t \le T,$$

and

$$\mathcal{G}_c(T,\theta) = I.$$

Let $\mathcal{L}(t;\theta)$ be the linear matrix-valued function defined by

$$\begin{aligned} \mathcal{L}(t;\theta) &= I + (t-T)\mathcal{G}_c'(T,\theta) \\ &= I + (t-T)\mathcal{F}_c(T,\theta)^{-1} \nabla_\theta f(T;\theta_0)^{\mathsf{T}} \nabla_\theta f(T;\theta_0), \quad 0 \le t \le T. \end{aligned}$$

We have the following result:

Proposition 2. *Assume that for an $\epsilon > 0$ we have*

$$\max_{T_0 \le t \le T} \|\mathcal{G}_c(t,\theta) - \mathcal{L}(t;\theta)\|_2 \le \epsilon. \tag{18}$$

Then the minimal eigenvalue $\lambda_{\min}$ *of the matrix*

$$\mathcal{F}_c(T_0, \theta) = \int_{T_0}^{T} \nabla_\theta f(\tau; \theta_0)^\mathsf{T} \nabla_\theta f(\tau; \theta_0), d\tau$$

satisfies

$$\lambda_{\min} \le \|\mathcal{F}_c(T, \theta)\|_2 \epsilon.$$

Proof. From equation (17) we immediately get (for $T_0 \to T$)

$$\begin{aligned}\mathcal{G}_c(t, \theta) &= \mathcal{F}_c(T, \theta)^{-1} \Big(\mathcal{F}_c(T, \theta) - \int_t^T \nabla_\theta f(\tau; \theta_0)^\mathsf{T} \nabla_\theta f(\tau; \theta_0)\, d\tau \Big) \\ &= I - \mathcal{F}_c(T, \theta)^{-1} \int_t^T \nabla_\theta f(\tau; \theta_0)^\mathsf{T} \nabla_\theta f(\tau; \theta_0)\, d\tau, \quad 0 \le t \le T.\end{aligned}$$

This and the definition of $\mathcal{L}$ imply

$$\begin{aligned}&\mathcal{G}_c(t, \theta) - \mathcal{L}(t; \theta) \\ &= \mathcal{F}_c(T, \theta)^{-1} \Big((T-t) \nabla_\theta f(T; \theta_0)^\mathsf{T} \nabla_\theta f(T; \theta_0) - \int_t^T \nabla_\theta f(\tau; \theta_0)^\mathsf{T} \nabla_\theta f(\tau; \theta_0)\, d\tau \Big),\end{aligned}$$

for $0 \le t \le T$. Consequently we have

$$\mathcal{F}_c(T_0, \theta) = (T - T_0) \nabla_\theta f(T; \theta_0)^\mathsf{T} \nabla_\theta f(T; \theta_0) - \mathcal{F}_c(T, \theta) \big(\mathcal{G}_c(T_0, T) - \mathcal{L}(T_0; \theta) \big). \tag{19}$$

Let a with $\|a\|_2 = 1$ be an eigenvector of $\nabla_\theta f(T; \theta_0)^\mathsf{T} \nabla_\theta f(T; \theta_0)$ corresponding to the eigenvector 0. Then equation (19) implies

$$a^\mathsf{T} \mathcal{F}_c(T_0, \theta) a = -a^\mathsf{T} \mathcal{F}_c(T, \theta) \big(\mathcal{G}_c(T_0, \theta) - \mathcal{L}(T_0; \theta) \big) a.$$

From assumption (18) we get the estimate

$$\lambda_{\min} \le a^\mathsf{T} \mathcal{F}_c(T_0, \theta) a \le \|\mathcal{F}_c(T, \theta)\|_2 \epsilon$$

for all $a \in \ker \nabla_\theta f(T; \theta_0)^\mathsf{T} \nabla_\theta f(T; \theta_0)$.

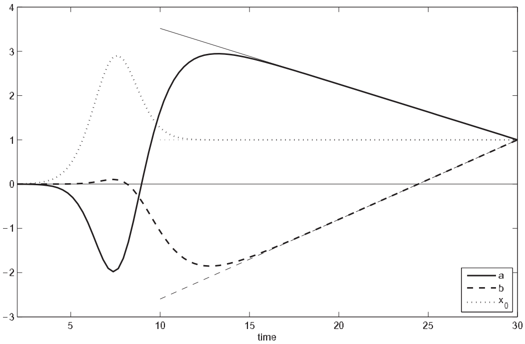

Figure 2: Generalized sensitivities for the three parameters of the logistic equation (16) and corresponding straight lines corresponding to the diagonal elements of the matrix $\mathcal{L}(t;\theta)$ for for the nominal parameters $\theta_0 = (0.9, 0.005, 0.2)$.

References

[1] *H. T. Banks, and K.L. Bihari,* Modeling and estimating uncertainty in parameter estimation. Inverse Problems **17**(2001), 95 – 111.

[2] *H. T. Banks, S. Dediu, S. Ernstberger, and F. Kappel,* A new approach to optimal design problems, submitted.

[3] Ĵ. J. Batzel, F. Kappel, D. Schneditz, and H. T. Tran, Cardiovascular and Respiratory Systems: Modeling, Analysis, and Control, Frontiers in Applied Mathematics Vol. 34, SIAM, Philadelphia 2007.

[4] *V. V. Fedorov, and P. Hackel,* Model-Oriented Design of Experiments, Springer-Verlag, Berlin 1997.

[5] *F. Puckelsheim,* Optimal Design of Experiments, John Wiley & Sons, New York, NY., 1993.

[6] *K. Thomaset, and C. Cobelli,* Generalized sensitivity functions in physiological system identification, Ann. Biomed. Eng. **27**(5) (1999), 607 – 616.