Supplement: A Variational Approach to Parameter Estimation in Ordinary Differential Equations

Daniel Kaschek

July 4, 2012

Contents

1 Parameter estimation in ODE systems with inputs 1
2 The Gâteaux differential of $\chi^2[x]$ 2
3 Discretization of the variational problem 4
4 Analytical solution with strict boundary condition 5
5 Analytical solution with relaxed boundary condition 6
6 Working with noise and observation function 7

1 Parameter estimation in ODE systems with inputs

Let

$$\dot{y} = f(y, [x], p)$$  \hspace{1cm} (1)

be an ODE for components $y \in \mathbb{R}^N$ with continuous input function $[x] \in C^0(\mathbb{R}, \mathbb{R}^M)$. The system depends on dynamic parameters $p \in \mathbb{R}^K$ and on initial conditions $y(t = 0)$ which are combined to $P = (p, y(t = 0))$. The solution to eq. (1) is denoted by $y_P[x]$ and is a differentiable function of time $t$. Let us assume that the components $y$ are measured at time points $t_i$ denoted by $y_i$. For fixed $[x]$ the $\chi^2$ function

$$\chi^2_{\text{red}}(P) = \sum_i \|y_P[x](t_i) - y_i\|^2 =: \sum_i \|\text{res}_y(t_i)\|^2$$  \hspace{1cm} (2)

is a function of parameters $P$ and the optimal parameters $\hat{P}$ minimizing eq. (2) are approached by an optimization method. For varying input function $[x]$, the solution trajectory $y_P[x]$ and the optimal parameters $\hat{P}$ vary as well and depend on the entire course of $[x]$. The question is, what is the optimal choice for the input function resulting in the lowest total $\chi^2$ value.

To this purpose, we are going to deal with variational calculus. First, the functional of optimal parameters $\hat{P}[x]$ is characterized by

$$\forall k : \left. \frac{\partial \chi^2_{\text{red}}}{\partial P_k} \right|_{P[x]} = \left. \frac{\partial}{\partial P_k} \sum_i \|y_P[x](t_i) - y_i\|^2 \right|_{P=\hat{P}[x]} = 0.$$  \hspace{1cm} (3)
Thereby, the total $\chi^2$ becomes a functional of $[x]$:

$$\chi^2[x] = \sum_i \|y_{\tilde{P}[x]}(t_i) - y_i\|^2 + \|x(t_i) - x_i\|^2$$  

The second term introduces an objective function for deviations from the $x$-measurements. The Gâteaux differential of $\chi^2[x]$ is defined as

$$d\chi^2([x], h) := \frac{d\chi^2[x + eh]}{de} \bigg|_0 = 2 \sum_i \left\{ \text{res}_y(t_i), d y_{\tilde{P}}([x], h)(t_i) \right\} + \left\{ \text{res}_x(t_i), h(\tilde{t}) \right\}.$$  

which is homogeneous, i.e. $\forall \alpha \in \mathbb{R} : d\chi^2([x], \alpha h) = \alpha d\chi^2([x], h)$ but in general it is whether continuous nor is it linear. If $d\chi^2([x], h)$ is continuous in $h$,

$$\forall h : d\chi^2([\hat{x}], h) = 0$$  

is a necessary condition for an extremal solution $[\hat{x}]$. If $d\chi^2([x], h) = \delta\chi^2[x]h$ is in addition linear in $h$ then $\delta\chi^2[x]$ is called the first variation of $\chi^2$ and the extremal condition becomes $\delta\chi^2[\hat{x}] = 0$. From the practical point of view, linearity is a major feature for other reasons. Assume, we want to find an approximate extremal solution $[\hat{x}]$ in a finite dimensional function space $X_0$ defined by a finite basis $\{h_1, \ldots, h_n\}$. Then by linearity,

$$\forall h \in X_0 : \chi^2[x + h] \approx \chi^2[x] + \sum_k \alpha_k \delta\chi^2[x]h_k$$  

$$= \chi^2[x] + \langle \alpha, \nabla\chi^2[x] \rangle$$  

where $\nabla\chi^2[x]_k := \delta\chi^2[x]h_k$. I.e. $[x]$ is approximately extremal only if $\nabla\chi^2[x] = 0$ and starting from a non-extremal $[x]$, $h = \sum_k \nabla\chi^2[x]_k h_k$ indicates the direction to the extremum. In the framework of numerical optimization theory, linearity is absolutely necessary. 

With these considerations in mind the Gâteaux differential of $\chi^2$ is analyzed. The following section will show linearity as well as continuity.

## 2 The Gâteaux differential of $\chi^2[x]$

The differential of $\chi^2$ depends on the differential of the trajectory $y_{\tilde{P}[x]}[x]$. Since every trajectory satisfies the differential equation $\dot{y}_{\tilde{P}[x]}[x] = f(y_{\tilde{P}[x]}[x], [x], \tilde{P}[x])$, we have

$$\frac{d}{dt} y_{\tilde{P}}(x, h) = D_y f(y_{\tilde{P}}(x, h) + D_x f \ h + D_P f \ d\tilde{P}(x, h),$$  

where $D_y f$ denotes the Jacobian of $f$ with respect to $y$ and analogously $D_x f$ and $D_P f$. The optimal choice of parameters $\tilde{P}[x]$ is defined by eq. (3) and explicitly reads:

$$\frac{\partial\chi^2_{\text{red}}}{\partial \tilde{P}_k} \bigg|_{\tilde{P}} = 2 \sum_i \left\{ \text{res}_y(t_i), \frac{\partial y_{\tilde{P}}[x]}{\partial \tilde{P}_k} (t_i) \right\} = 0 \equiv \Phi_k(\tilde{P}[x], [x]).$$  

Consequently, the Gâteaux differential of $\Phi$ restricted to $(\tilde{P}[x], [x])$ vanishes and we get

$$d\Phi(x, h) = D_{\tilde{P}}\Phi|_{\tilde{P}[x]} \ d\tilde{P}(x, h) + \partial\Phi(x, h) = 0$$  

$$\Rightarrow d\tilde{P}(x, h) = -C \partial\Phi(x, h)$$
with \( \partial \Phi(x, h) := \frac{d\Phi(P[x], P[x+h])}{dt} \) and \( \Phi'_{\dot{P}} \) is the inverse of the covariance matrix \( C \). In the following, differentials denoted by \( \partial \) mean that \( P = \dot{P} \) is fixed. By eq. (10)

\[
\partial \Phi_k(x, h) = 2 \sum_i \left( \partial y_P(x, h)(t_i), \left. \frac{\partial y_P[x]}{\partial P_k} \right|_{\dot{P}} (t_i) \right) + \left. \left( \text{res}_y(t_i), \partial \frac{\partial y_P[\dot{P}]}{\partial P_k} \right|_{\dot{P}} (x, h)(t_i) \right).
\]

The unknown functions are expressed by their corresponding differential equations, hence

\[
\frac{d}{dt} \frac{\partial y_P(x, h)}{\partial P_k} = D_g f \frac{\partial y_P(x, h)}{\partial P_k} + D_x f \ h
\]

\[
\frac{d}{dt} \frac{\partial y_P[x]}{\partial P_k} = D_g f \left. \frac{\partial y_P[x]}{\partial P_k} \right|_{\dot{P}} + \left. \frac{\partial f}{\partial P_k} \right|_{\dot{P}} (x, h)
\]

\[
\left( \frac{\partial}{\partial P_k} \left. \frac{\partial y_P}{\partial P_k} \right|_{\dot{P}} \right)(x, h) = \partial \left( D_g f \left. \frac{\partial y_P[x]}{\partial P_k} \right|_{\dot{P}} + \left. \frac{\partial f}{\partial P_k} \right|_{\dot{P}} (x, h) \right) = \left. \frac{\partial y_P[x]}{\partial P_k} \right|_{\dot{P}} (x, h) + D_g f \left. \frac{\partial y_P[x]}{\partial P_k} \right|_{\dot{P}} (x, h) + \left. \frac{\partial f}{\partial P_k} \right|_{\dot{P}} (x, h).
\]

Eq. (15) has initial condition zero for dynamical parameters \( P_k \) and one for the corresponding component of \( y \) if \( P_k \) is an initial concentration. Eqs. (14, 17) have zero initial condition which is due to the fact that both \( y_P(t=0) \) and \( \frac{\partial y_P}{\partial P_k}(t=0) \) are fixed by the corresponding parameters \( P_k \) of initial concentrations no matter what \( x \) is. Finally, the unknown differentials \( \partial D_g f \) and \( \partial \frac{\partial f}{\partial P_k} \) can be evaluated and stated explicitly:

\[
\frac{d}{dt} \frac{\partial y_P}{\partial P_k} = D_g f \frac{\partial y_P(x, h)}{\partial P_k} + D_x f \ h
\]

\[
\left( \frac{\partial}{\partial P_k} \left. \frac{\partial y_P}{\partial P_k} \right|_{\dot{P}} \right)(x, h) = \frac{\partial}{\partial P_k} \left. \frac{\partial y_P[x]}{\partial P_k} \right|_{\dot{P}} + \left. \frac{\partial f}{\partial P_k} \right|_{\dot{P}} (x, h)
\]

At this point, \( d\chi^2(x, h) \) is completely defined by derivatives of \( f \), the covariance matrix \( C \) of the reduced parameter estimation problem and the solutions to the differential equations (14-17). These differential equations have the special feature that they are of the form

\[
\dot{y}(t) = M(t)y(t) + h(t)
\]

with \( M : \mathbb{R} \rightarrow \text{Hom}(\mathbb{R}^n) \). The solution to this equation is explicitly given by

\[
y(t) = \Phi(t) \left( \int_0^t \Phi^{-1}(\tau)h(\tau) \ d\tau + y_0 \right)
\]

\[
=: \mathcal{J}[M, h](t) + \Phi(t)y_0
\]

with \( \Phi(t) := \exp \left( \int_0^t M(\tau) \ d\tau \right) \). This explicit formula is the key to the linearity of \( d\chi^2(x, h) \). By the linearity of the integral,

\[
\mathcal{J}[M, \alpha h + \beta g] = \alpha \mathcal{J}[M, h] + \beta \mathcal{J}[M, g]
\]

is linear in the second argument. Furthermore,

\[
\partial y_P(x, h) = \mathcal{J}[D_g f, D_x f h]
\]

is linear in \( h \) and thereby also eqs. (18) and (19) are linear in \( h \). Consequently,

\[
\left( \frac{\partial}{\partial P_k} \right)(\partial y_P \left. \frac{\partial y_P}{\partial P_k} \right|_{\dot{P}}) = \mathcal{J} \left[ D_g f, (\partial D_g f) \frac{\partial y_P[x]}{\partial P_k} \left|_{\dot{P}} (x, h) \right) + \left( \frac{\partial f}{\partial P_k} \right)(x, h) \right]
\]

3
is linear in $h$ as well as $\partial \Phi_k(x, h)$, $d\hat{P}(x, h)$, $dy_P(x, h)$ and finally $d\chi^2(x, h)$ so that we come to the important result:

$$d\chi^2(x, h) = \delta \chi^2(x) h$$

(26)

is a linear functional. Due to the explicit solutions of eqs. (14-17), $d\chi^2(x, h)$ is also continuous for appropriate $f$.

3 Discretization of the variational problem

One possibility to approach $[\hat{x}]$ is discretization of the function space from which $[\hat{x}]$ originates. An adequate finite dimensional subspace $X_0 \subset X$ has to be chosen. Inspired by Fourier or Laplace transformations, bases of $X_0$ could be

$$h_k(t) = \{ \sin(k\omega t), k = 1, 3, 5, \ldots \}
\{ \cos(k\omega t), k = 2, 4, 6, \ldots \}$$

(27)

or

$$h_k(t) = e^{-k\gamma t}$$

(28)

and $k$ should be bounded.

Another possibility would be to choose smooth spline functions. Let $x_S(t; x_1, \ldots, x_n, \lambda)$ be the smooth spline function defined by nodes $x_i$ at times $t_i$, i.e. $x_S$ is the smooth spline through the measurement points of $x$. Then

$$h_k(t) = \frac{\partial x_S(t; x_1, \ldots, x_n, \lambda)}{\partial x_k}$$

(29)

defines a local basis of a function space of smooth spline functions. This approach is closely connected to the estimation problem by the objective function

$$\chi^2(P, x_1, \ldots, x_n) = \sum_i \| y_p[x_S](t_i) - y_i \|^2 + \| x_S(t_i; x_1, \ldots, x_n) - x_i \|^2.$$  

(30)

When computing $\frac{\partial \chi^2}{\partial x_k}$, derivatives of $x_S$ and $y_p[x_S]$ have to be computed:

$$\frac{\partial x_S}{\partial x_k}(t_i) = \frac{d}{d\epsilon} \bigg|_{\epsilon=0} x_S(t_i, x_k + \epsilon)$$

(31)

$$\frac{\partial y_p}{\partial x_k}(t_i) = \frac{d}{d\epsilon} \bigg|_{\epsilon=0} y_p[x_S(t_i; x_k + \epsilon)]$$

(32)

$$= \frac{d}{d\epsilon} \bigg|_{\epsilon=0} y_p[x_S(t_i; x_k) + \epsilon \frac{\partial x_S}{\partial x_k}(t_i)]$$

(33)

$$= \delta y_p(x_S) \frac{\partial x_S}{\partial x_k}(t_i)$$

(34)

$$= \delta y_p(x_S) h_k(t_i).$$

(35)

In other words

$$\left( \nabla \chi^2[x_S] \right)_k := \delta \chi^2(x_S) h_k = \frac{\partial \chi^2(P, x_1, \ldots, x_n)}{\partial x_k}.$$  

(36)
Conclusion  The functional of the first variation applied to the basis functions coincides with the gradient which is usually employed by the optimizer to optimize the parameters \( x_1, \ldots, x_n \).

4 Analytical solution with strict boundary condition

If the differential equation \( \dot{y} = f(y, x) \) is linear with respect to \( x \), a differential equation for \( x(t) \) can be derived from the condition of vanishing first variation. Since in \( \delta \chi^2(x)h \) the variation \( h(t) \) occurs one time under the integral and one time not, this has to be resolved by partial integration. Therefore, \( \text{res}_y \) and \( \text{res}_x \) as defined in eq. (2) for discrete time points need to become continuous functions of \( t \). This can be realized by interpolation splines

\[
S_x(t) = \text{spline}(x_1, \ldots, x_n)(t) \quad \text{and} \quad S_y(t) = \text{spline}(y_1, \ldots, y_n)(t). \tag{37, 38}
\]

Thereby, \( y(t) - S_y(t) \) becomes a function of \( t \) with \( y(t_i) - S_y(t_i) = y(t_i) - y_i = \text{res}_y(t_i) \) and analogously for \( x \). Another issue however is not resolvable. The variation \( \delta \hat{P}(x)h \) does not depend on \( t \) at all. Therefore, \( h \) cannot be extracted from \( \delta \hat{P}(x)h \) by means of differentiation or partial integration. The problem has to be reformulated as finding the optimal \( x(t) \) for given data and given parameters. Thereby, we have to minimize

\[
\chi^2[x|P] = \int_0^T \left( (y(t) - S_y(t))^2 + (x(t) - S_x(t))^2 \right) dt \quad \text{(39)}
\]

For the first variation we have

\[
\frac{1}{2} \delta \chi^2(x|P)h = \int_0^T \left< \text{res}_y(t), \Phi(t) \int_0^t \Phi^{-1}(\tau) \Phi_x f(\tau)h(\tau) \right> \text{d}\tau + \left< \text{res}_x(t), h(t) \right> \text{d}t \quad \text{(41)}
\]

and by partial integration

\[
= \left[ \int_0^t \Phi(t)^* \text{res}_y(t) \text{d}t, \int_0^t \Phi^{-1}(t) \Phi_x f(t)h(t) \text{d}t \right]_0^T - \int_0^T \left\langle \int_0^t \Phi(\tau)^* \text{res}_y(\tau) \text{d}\tau, \Phi^{-1}(t) \Phi_x f(t)h(t) \right\rangle \text{d}t. \tag{42, 43}
\]

In order to find a unique solution, one requires \( \int_0^T \Phi^{-1}(t) \Phi_x f(t)h(t) \text{d}t \) to vanish. This means that only variations \( h \) are taken into consideration that leave initial and end point of \( y(t) \) untouched. Since eq. (43) has to be zero for all such \( h \), the equation

\[
\Phi^{-1}(t)^* \int_0^t \Phi(\tau)^* \text{res}_y(\tau) \text{d}\tau - \text{res}_x(t) = 0 \quad \text{(44)}
\]

has to be solved. This integral equation can be transformed into an equivalent system of differential equations. Therefor, we use that \( \Phi = \exp \left( \int_0^\tau \Phi_y f(\tau) \text{d}\tau \right) \) is a fundamental system for \( \dot{\phi} = \Phi_y f \phi \) and consequently, \( \Phi^{-1} \) is a fundamental system for \( \dot{\phi} = -\Phi_y f \phi \). From that, eq. (46)
follows. For the time derivative of $x$ we distinguish different cases. If $D_x f = \text{const.}$, it is not affected by the time derivative leading to eq. (47). The entire ODE system then reads:

\[
\begin{align*}
\dot{y} &= f(y, x) \\
\dot{u} &= -D_y f(y, x)^* u + y - S_y(t) \\
\dot{x} &= S_x(t) + D_x f^* \dot{u}.
\end{align*}
\]

(45) (46) (47)

If $D_x f$ is not constant, its time derivative is

\[
\frac{d}{dt} D_x f = D_x \frac{df}{dt} = D_x (D_y f \dot{y} + D_x f \dot{x})
\]

(48)

and we distinguish the case $D_x^2 f = 0$ where

\[
\dot{x} = S_x(t) + \dot{y}^* D_y D_x f(y, x)^* u + D_x f(y, x)^* \dot{u}.
\]

(49)

and $D_x^2 f \neq 0$ where

\[
\dot{x} = \dot{x}^* D_x^2 f(y, x)^* u + H(y, x, u, t)
\]

(50)

with some function $H$ that depends on $y$, $x$, $u$ and $t$ but not on $\dot{x}$. In this case, it is unclear whether $1 - D_x^2 f(y, x)^* u$ is invertible or not. Therefore, a unique solution for $(y, u, x)$ only exists if $1 - D_x^2 f(y, x)^* u$ is invertible alongside the whole trajectory.

5 Analytical solution with relaxed boundary condition

When optimizing $\chi^2[\hat{x}(P), P]$ with respect to $P$, the resulting $\hat{x}(\hat{P})$ looks very promising and follows the data points closely. Parameter estimation with repeated noise realizations reveals, that $\hat{P}$ is an unbiased estimator with a variance that is about the variance of the original estimation problem. However, the main problem is that starting with parameters $P$ that are slightly different from $\hat{P}$, the trajectory $x(t)$ obtained by the above differential equation for parameters $P$ differing slightly from $\hat{P}$ may be far away from the data points. In most cases it does not seem like an optimal solution at all. Though, tests with such optimal solutions $\hat{x}$ reveal that indeed, for any $h$ obeying the boundary condition (42), $\chi^2[\hat{x} + \epsilon h] > \chi^2[\hat{x}]$ for $\epsilon \neq 0$ and $\epsilon$ in a neighborhood of zero. This problem is resolved by relaxing the boundary condition.

For real valued functions $f$ and $g$ we have the following general equation which can be shown by partial integration:

\[
\int_0^t f(t) \left( \int_0^t g(\tau) \ d\tau \right) dt = \int_0^t \left( \int_0^\tau f(\tau) \ d\tau \right) g(t) \ dt
\]

(51)

Applying this formula to (41), we get

\[
\frac{1}{2} \delta \chi^2(x) h = \int_0^T \left\langle \int_0^T \Phi(\tau)^* \text{res}_y(\tau) \ d\tau, \Phi^{-1}(t) D_x f(t) h(t) \right\rangle + \left\langle \text{res}_x(t), h(t) \right\rangle dt
\]

(52)

and therefore

\[
D_x f(t)^* \Phi^{-1}(t)^* \int_0^T \Phi(\tau)^* \text{res}_y(\tau) \ d\tau + \text{res}_x(t) = 0
\]

(53)

\[
\therefore u(t)
\]
This leads to the very similar system of differential equations

\[
\dot{y} = f(y, x) \quad (54)
\]

\[
\dot{u} = -D_y f(y, x)^* u - (y - S_y(t)) \quad (55)
\]

\[
\dot{x} = \dot{S}_x(t) - D_x f^* \dot{u} - \frac{d}{dt} D_x f(t)^* u. \quad (56)
\]

with the only difference that the initial condition of \(x\) and \(u\) already depend on the solution of the ODE. In contrast, the final conditions \(u(T) = 0\) and \(x(T) = S_x(T)\) follow directly from eq. (53). Hence, the price to pay for releasing the boundary condition of \(h\) is that the initial value problem becomes a two way boundary value problem, which is numerically demanding.

6 Working with noise and observation function

The basic objective function used so far is the unweighted sum of squares. In the context of parameter estimation in chemical and biochemical reaction networks, some components may be measured more precisely than others. Several of the components might be unobserved or only linear combinations might be observable. In order to deal with these issues, the formalism needs to be adjusted.

Let \(G\) be an observation function so that

\[
\tilde{z}(t) = G(y(t), x(t)). \quad (57)
\]

In order to deal with the approach derived above, it is necessary that \(G\) separates into two parts \(G_x\) and \(G_y\), so that

\[
\tilde{y}(t) = G_y(y(t)) \quad (58)
\]

\[
\tilde{x}(t) = G_x(x(t)). \quad (59)
\]

In addition, we assume noise functions \(\sigma_{\tilde{x}}\) and \(\sigma_{\tilde{y}}\) for the data. Then, eq. (53) becomes

\[
D_x f^*(\Phi^{-1})^* \int_t^T \Phi^* D_y G_y^* \frac{\tilde{y} - S_{\tilde{y}}}{\sigma_{\tilde{y}}^2} d\tau + D_x G_x \frac{\tilde{x} - S_{\tilde{x}}}{\sigma_{\tilde{x}}^2} = 0. \quad (60)
\]

Like above, linearity in the input is sufficient to guarantee a solution. In this case this means that \(G_x\) is expected to be linear. The ODE system to solve is

\[
\dot{y} = f(y, x) \quad (61)
\]

\[
\dot{u} = -D_y f(y, x)^* u - D_y G_y^* \frac{G_y(y) - S_{\tilde{y}}}{\sigma_{\tilde{y}}^2} \quad (62)
\]

with \(x = G_x^{-1}(S_{\tilde{x}} - \sigma_{\tilde{x}}^2 (D_x G_x^*)^{-1} D_x f^* u)\).

**Remark 6.1 (Observational function for the input)** Since there are strong requirements to \(G_x\), it is most convenient to make a transformation of the data beforehand to circumvent dealing with \(\tilde{x}\). The same works for \(y\). If \(G_y\) has not full rank meaning that some components are unobserved, these components do not contribute to \(\dot{u}\). Say \(G_y = \text{diag}(1, \ldots, 1, 0, \ldots, 0)\). In other cases it might be necessary do deal at least with \(G_y\), e.g. if \(G_y\) is a nonlinear transformation which produces normally distributed data \(\tilde{y}\) whereas \(y\) itself is not normally distributed.
Remark 6.2 (Noise) If noise in the input measurements is considered, the condition for constructing \( x \) uniquely from \( u \) is that \( D^2_x f^*u + \frac{1}{\sigma_x^2} \mathbf{1} \) be invertible, analogously to eq. (50). Or equivalently

\[
\sigma_x^2 D^2_x f^*u + \mathbf{1} \text{ be invertible.} \tag{63}
\]

This is the case if the spectral radius (largest absolute eigenvalue)

\[
\rho(\sigma_x^2 D^2_x f^*) < 1 \tag{64}
\]

is smaller than 1. That can be achieved if the curvature \( D^2_x f \) is bounded over all feasible solution points. I.e. let \( U \subset \mathbb{R}^N \times \mathbb{R}^M \) be a subspace where all feasible solutions of eqs. (61-62) are contained in and let

\[
\sup_{(y,x) \in U} \rho(D^2_x f) = \bar{\rho} < \infty. \tag{65}
\]

Then, if the measurement noise \( \sigma_x \) is small enough, eq. (50) can be solved for \( \dot{x} \) in the entire solution space. At this point, existence and uniqueness of a solution of eqs. (54-56) is subject to the same considerations as in the case of functions \( f \) that are linear with respect to \( x \).