Issues in sampling and estimating continuous-time models with stochastic disturbances

Lennart Ljung [a]   Adrian Wills [b]

[a] *Division of Automatic Control, Linköpings Universitet, SE-581 80 Linköping, Sweden (e-mail: `ljung@isy.liu.se`)*

[b] *School of Electrical Engineering and Computer Science, University of Newcastle, Callaghan, NSW, 2308, Australia (e-mail: `adrian.wills@newcastle.edu.au`)*

Abstract

The standard continuous time state space model with stochastic disturbances contains the mathematical abstraction of continuous time white noise. To work with well defined, discrete time observations, it is necessary to sample the model with care. The basic issues are well known, and have been discussed in the literature. However, the consequences have not quite penetrated the practise of estimation and identification. One example is that the standard model of an observation being a snapshot of the current state plus noise independent of the state cannot be reconciled with this picture. Another is that estimation and identification of time continuous models require a more careful treatment of the sampling formulas. We discuss and illustrate these issues in the current contribution. An application of particular practical importance is the estimation of models based on irregularly sampled observations.

1 Continuous Time Models

A standard mathematical abstraction in control theory is the continuous time linear model with stochastic disturbances:

$$\dot{x}(t) = Ax(t) + Bu(t) + \dot{w}(t) \tag{1a}$$
$$\dot{z}(t) = Cx(t) + \dot{e}(t) \tag{1b}$$

Here $\dot{w}(t)$ and $\dot{e}(t)$ are stochastic disturbances. For $x(t)$ to be a *state* in the sense that it condenses knowledge of the past, and thus becomes a Markov process, it is necessary that both $\dot{w}(t)$ and $\dot{e}(t)$ are unpredictable from past data. This means that they have to be *white noises*. It is well known that the mathematical description of this is somewhat advanced: These processes will have to have infinite variances, and a formal mathematical description is via stochastic integrals of their integrated versions, that are Wiener processes (see e.g. Åström (1970) for an account of this).

Innovations Form   The model (1) allows any correlation between the (white noise) stochastic disturbance

signals $\dot{w}$ and $\dot{e}$. In particular, they could be fully correlated:

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + K\dot{\nu}(t) \tag{2a}$$
$$\dot{z}(t) = C\hat{x}(t) + \dot{\nu}(t) \tag{2b}$$

It is well known how a general process (1) can be transformed to a representation (2) by letting K be determined as the Kalman gain (see e.g. p247 in Åström (1970)). This gives a representation with the same second order properties (same spectrum) for $\dot{z}$ (and since it is a Gaussian process, it will then also have the same pdf.) The spectra for x and $\hat{x}$ will in general be different, though. This form, (2) is known as the *Innovations Form* of (1). If $\dot{\nu}$ in (2a) is replaced by the expression from (2b) we obtain the *Kalman Filter*:

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + K(\dot{z}(t) - C\hat{x}(t)) \tag{3}$$

All this is of course well known.

2 Discrete Time Measurements

Now, in real life the time continuous output $\dot{z}(t)$ is of course not observed, and neither are any instantaneous snapshots since they would have infinite variance. Various ways to formulate a realistic discrete time version of this have been suggested, and this paper is about several aspects and issues involved in such formulation.

One idea is simply to postulate that discrete time noisy observations of the state are available, without detailing on how they are obtained from (1b):

$$y(t_k) = H_k x(t_k) + v(t_k) \qquad (4)$$

with a suitably defined matrix H_k. This is the approach taken in Jazwinski (1970), Section 7.2, and from this a continuous-discrete Kalman filter is developed, updating estimates of the continuous state process (1a). Åström takes the same approach for identifying continuous time state space models in Åström (1980), eq (3.1)-(3.2). The same philosophy is used for Differential Algebraic Equations (DAE's) in Gerdin et al. (2007). See also Johansson et al. (1999).

Remark 1: For the filtering calculations to work, it is essential that $v(t_k)$ is independent of $x(t_k)$. It is important to note that *it is impossible to have such a finite-variance observation at time t_k based on (1b)!*. The reason is that in order to achieve finite variance, the signal $\dot{z}$ must be low-pass, pre-sample filtered before sampling. This filtering must occur after time t_k in order for the filtered noise-component be independent of $x(t_k)$. *So, a proper time index of y in (4) must be larger than t_k*. Now, if our concern is to predict future values of y (as in an identification application), this is not so essential, since it is really just an issue of labeling the time stamps. But for state estimation, it may affect what is considered to be a predicted and a filtered state estimate.

An approach to describe how (4) relates to (1b) is to assume *integrated sampling*:

$$y(t_k + \delta_k) = \frac{1}{\delta_k} \int_{t_k}^{t_k+\delta_k} \dot{z}(s) ds \qquad (5)$$

This is used e.g. in (10.17) in Åström (1970) (which does not divide by the time interval) and in eq. (22.10.84) in Goodwin et al. (2001). Integrated sampling is also discussed in e.g. Goodwin et al. (1995) (which considers duality with a control problem) and Feuer & Goodwin (1996), Chapter 6 (which considers "delta-versions" of integrated sampling), which is also discussed in Yuz (2006) where attention is given to sampling zeros. Another relevant reference is Shats & Shaked (1991). All these authors describe the case without input.

3 Discrete Time Models

The state process $x(t)$ is a well defined signal from (1a) and its values at discrete time instances $t_k, k = 1, \ldots$ can readily be found. Assume that the input is piecewise constant:

$$u(t) = u(t_k) = u_k \quad t_k \leq t < t_{k+1} \qquad (6)$$

Furthermore, if $y(t_k + \delta_k)$ relates to x via (1) and (5), the observations can be described by the discrete-time model

$$x_{k+1} = F_k x_k + G_k u_k + \tilde{w}_k \qquad (7a)$$
$$y_k = H_k x_k + D_k u_k + v_k \qquad (7b)$$

where $x_k = x(t_k)$ and $y_k = y(t_k + \delta_k)$. The expressions for the matrices and covariance matrices have been given in several papers and are repeated in Appendix A for reference.

Innovations Form From these equations, the Kalman gain $\tilde{K}_k$ can be computed via the time-varying Riccati equation in the well known manner (cf Appendix A) to yield the *innovations form*:

$$\hat{x}_{k+1} = F_k \hat{x}_k + G_k u_k + \tilde{K}_k \epsilon_k \qquad (8a)$$
$$y_k = H_k \hat{x}_k + D_k u_k + \epsilon_k \qquad (8b)$$

This representation has the property that the second order properties of the inputs and outputs are the same in (7) and (8). By replacing ϵ in (8a) by the expression from (8b), equation (8a) becomes the Kalman filter for estimating the state. The predicted output is

$$\hat{y}_{k+1} = H_{k+1} \hat{x}_{k+1} + D_{k+1} u_{k+1} \qquad (9)$$

Simplistic Sampling The computation of $\tilde{K}_k$ in (8) according to Equations (A.1)–(A.2) is rather complicated. It is tempting to use a simplified formula, that ignores the averaging over x in (5) and uses an assumption that the noise is piecewise constant over the sampling interval. This sounds reasonable if the sample period $\Delta_k \triangleq t_{k+1} - t_k$ is short compared to the time constants of the system. In other words it means that the continuous time innovations description (2) is sampled as if $\nu(t)$, like $u(t)$ is piecewise constant, (6), yielding

$$\hat{x}_{k+1} = F_k \hat{x}_k + G_k u_k + \bar{K}_k \epsilon_k \qquad (10a)$$
$$y_k = C \hat{x}_k + \epsilon_k \qquad (10b)$$
$$\bar{K}_k = P_k K \quad P_k = A^{-1}(e^{A\Delta_k} - I) \qquad (10c)$$

If the sampling period is constant, the matrices F and G will be constant and $\bar{K}_k$ will converge to a constant matrix K. In case $F - \bar{K}C$ is not stable, the Riccati equation corresponding to $R_1 = \bar{K}\bar{K}^T$ (the state noise covariance), $R_{12} = \bar{K}$ (the cross covariance between state and measurement noise), $R_2 = I$ (the measurement noise covariance) is solved to obtain a new stabilizing Kalman gain.

We will call this *simplistic sampling* of the continuous time innovations form. This is the approach taken in the MATLAB's System Identification Toolbox, Ljung (2007). Its manual states that this is a reasonable approximation if the sampling interval is reasonably chosen w.r.t. to the system and noise dynamics.

4 The Sampling Formulas

For equidistant sampling, the simplistic sampling (10) is considerably simpler than the correct sampling in Appendix A. It is interesting to compare how they behave.

We first note that the expressions for F_k and G_k will be the same using both the simplistic and correct sampling formulas. For the correct sampling formulas, a Taylor expansion about $\delta_k = 0$ reveals

$$H_k = C + \frac{\delta_k}{2}CA + \frac{\delta_k^2}{6}CA^2 + O(\delta_k^3) \tag{11a}$$

$$D_k = \frac{\delta_k}{2}CB + \frac{\delta_k^2}{6}CAB + O(\delta_k^3) \tag{11b}$$

$$\tilde{K}_k = \delta_k K + \frac{\delta_k^2}{2}(AK - KCK) + O(\delta_k^3) \tag{11c}$$

while the simplistic formulas provide

$$H_k = C \tag{11d}$$

$$D_k = 0 \tag{11e}$$

$$\bar{K}_k = \delta_k K + \frac{\delta_k^2}{2}AK + O(\delta_k^3) \tag{11f}$$

For small δ_k note that $H_k \approx C$ and $D_k \approx 0$ while $\tilde{K} \approx \bar{K} \approx 0$ and $\tilde{K}/\delta_k \approx \bar{K}/\delta_k \approx K$. For larger δ_k the difference between the system matrices for correct and simplistic sampling is not negligible. While this is difficult to analyse in general, the effect on dynamic response for a specific example can be readily illustrated as follows.

Example 1 *Consider the system in innovations form*

$$\dot{x} = Ax + Bu + K\nu$$
$$= \begin{bmatrix} -0.357 & 1 \\ -0.243 & 0 \end{bmatrix} x + \begin{bmatrix} 0.254 \\ 1.820 \end{bmatrix} u + \begin{bmatrix} 1.287 \\ 0.786 \end{bmatrix} \nu \tag{12}$$
$$y = Cx + e = \begin{bmatrix} 1 & 0 \end{bmatrix} x + \nu$$

This system has a recommended sampling interval (by the CONTROL SYSTEM TOOLBOX*) of 0.31 s. We consider the step responses from u and e for systems that are sampled with 0.2 s and 2 s respectively. The results are shown in Figures 1 and 2. The effects on the matrices for the different sampling rules are not insignificant. For sampling time $T = 2$ we have*

$$H = \begin{bmatrix} 0.6057 & 0.7397 \end{bmatrix}; \quad D = 1.1665; \quad \tilde{K} = \begin{bmatrix} 1.0279 \\ 0.2763 \end{bmatrix}$$

using the correct sampling formulas in Appendix A and

$$H = C = \begin{bmatrix} 1 & 0 \end{bmatrix}; \quad D = 0; \quad \bar{K} = \begin{bmatrix} 1.2815 \\ 0.2395 \end{bmatrix}$$
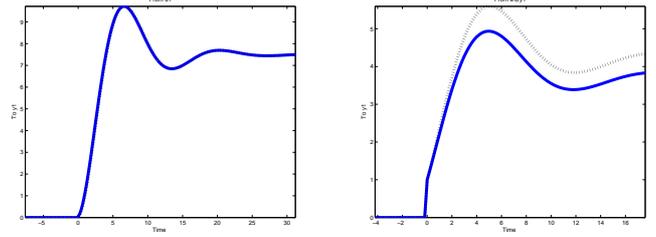

Fig. 1. Step response for DT system. Sampling interval 0.2. Solid: correct sampling. Dotted: simplistic sampling. Left: from input (the dotted and solid lines overlap in the resolution of the figure). Right: from noise source
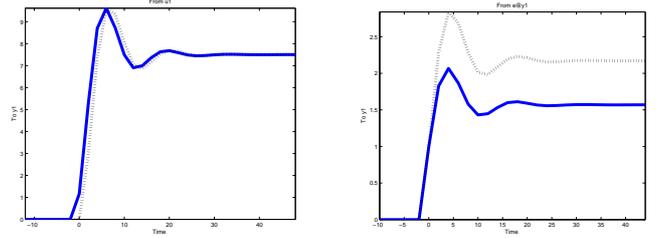

Fig. 2. Step response for DT system. Sampling interval 2. Solid: correct sampling. Dotted: simplistic sampling. Left: from input. Right: from noise source.

using the simplistic sampling formulas in (10). Still the difference in dynamic response in Figure 2(Left) is not very big. The difference in the noise response Figure 2(Right) is more evident.

5 System Identification

There is an extensive literature on System identification of continuous time systems e.g. Young (1981), Garnier et al. (2003), Garnier & Wang (2008), but few papers have dealt with the noise effects and the impact on the system dynamics of integrating sampling.

Estimation of system parameters is closely related to the state estimation/prediction problem. If A, B, C, K in (2) or $A, B, C, R_1, R_{12}, R_2$ in (1) contain unknown parameters θ, then these can be consistently identified by minimizing the prediction error criterion

$$\sum_{k=1}^{N} \|y_k - \hat{y}_k(\theta)\|^2 \tag{13}$$

or via the closely related Maximum Likelihood criterion. Here $\hat{y}$ is the expression from (9), depending on θ via the expressions (8), (9) and (A.2). See, e.g. Ljung (1999). The main work in minimizing (13) concerns finding the gradients of the criterion, which is based on the derivative $\psi_k = \frac{\partial}{\partial \theta}\hat{y}_k(\theta)$. In the identification toolbox Ljung (2007) the gradients of the matrices F, G etc. with respect to the parameters of the continuous time model are found by numerical differentiation.

To make these ideas concrete we include two examples

3

of parameter estimation in the following. The first example considers the very common situation of equidistant samples, and the second considers the case of non-equidistant samples.

Example 2 - *Estimating Continuous Systems Using Equidistant Samples:* *Consider the system (12) again. Produce a "continuous time simulation" by sampling it at sampling interval 1 ms. Let the input be piecewise constant over time intervals of 2 s, and let the variance of e at this sampling rate be 1000. (Corresponding to a noise intensity of the continuous source of 1.) Generate 2560 seconds of such a data record (2.56 million samples). Create integrated sampled data using (5) for $\delta = \Delta = 0.2$ and $\delta = \Delta = 2$. The resulting data are shown in Figure 3. A continuous time model of the same structure as (12) with six unknown parameters, corresponding to the matrix elements $A_{1,1}$, $A_{2,1}$, $B_{1,1}$, $B_{2,1}$, $K_{1,1}$ and $K_{2,1}$ was estimated using (13) with the correct sampling formula and the simplistic sampling formula for the two sampling intervals. The estimation was carried out in the System Identification Toolbox, Ljung (2007), using the `idgrey` model object with the different sampling routines implemented in the model-defining m-file.*

Remark 2: *In connection with (5) we commented upon the time labeling of the outputs. While it is essentially a label in the filtering context, it plays a role for the alignment of inputs and outputs for the identification application. Therefore we tested the simplistic sampling both for the case of time labeling as in (5) and with a shifted version, naming $y(t_k) = y((k+1)\Delta)$ The identified continuous time models with simplistic sampling for the two cases are shown in Figure 4. Clearly, such a time shift (which corresponds to an "anti-causal" shift of the input (`InputDelay = -Ts`)) is beneficial for the model, and we adopt this practice for the simplistically sampled model. For the correctly sampled model, there is no confusion about the time labels, since the formulas correctly account for the dependencies.*

The resulting estimated parameters with their estimated standard deviations are shown in Tables 1 and 2. We see that the estimated parameters are fine when the correct sampling has been used. The true values are well inside reasonable confidence regions. The models obtained by the simplistic sampling have worse accuracy, especially for the larger sampling interval. The true values are in several cases not within reasonable confidence intervals.

If the estimated continuous-time models are sampled according to their "own" sampling rules, the difference in sampled behavior is less pronounced, as illustrated in Figure 5.

Example 3 - *Estimating Continuous Systems from Unequally Sampled Data:* *Consider again the system in (12), simulated as before with a sample interval of 1ms to emulate a "continuous" system. The noise properties are the same as those in Example 2.*

Par	True	Est. Correct samp	Est. Simpl.samp
$A_{1,1}$	-0.3567	-0.3511 ± 0.0129	-0.3518 ± 0.0129
$A_{2,1}$	-0.2426	-0.2434 ± 0.0058	-0.2439 ± 0.0056
$B_{1,1}$	0.2540	0.2169 ± 0.0572	0.0329 ± 0.0568
$B_{2,1}$	1.8198	1.8575 ± 0.0584	1.8652 ± 0.0569
$K_{1,1}$	1.2865	1.2556 ± 0.0415	1.1373 ± 0.0313
$K_{2,1}$	0.7864	0.7398 ± 0.0421	0.6278 ± 0.0372

Table 1
Estimated parameter for sampling interval $\Delta = 0.2$. ± indicates the estimated standard deviation

Par	True	Est. Correct samp	Est. Simpl.samp
$A_{1,1}$	-0.3567	-0.3611 ± 0.0151	-0.3357 ± 0.0143
$A_{2,1}$	-0.2426	-0.2366± 0.0067	-0.2041± 0.0048
$B_{1,1}$	0.2540	0.2623 ± 0.0897	-0.8392 ± 0.0440
$B_{2,1}$	1.8198	1.7563 ± 0.0637	1.4489 ± 0.0505
$K_{1,1}$	1.2865	1.1997 ± 0.2022	14.0573 ± 50.6497
$K_{2,1}$	0.7864	0.6977 ± 0.0848	9.8242 ± 38.8295

Table 2
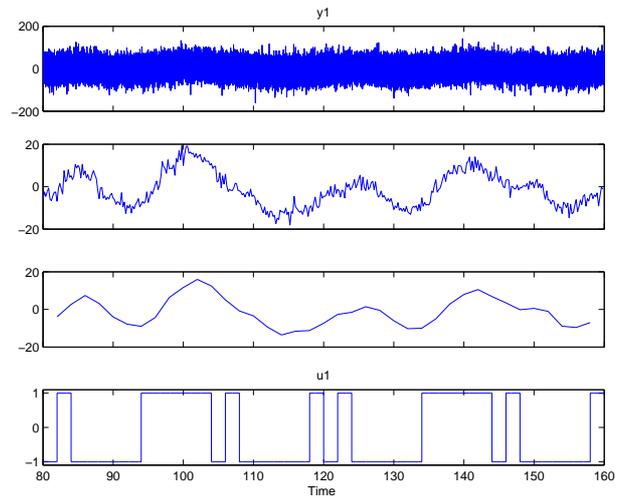Estimated parameter for sampling interval $\Delta = 2$. ± indicates the estimated standard deviation

Fig. 3. The input output data. From top (1) "Continuous time output data". (Generated by sampling interval 0.001.), Sampled output data, (2) sampling interval 0.2, (3) sampling interval 2, and (4) input data.

Again produce 2560 seconds worth of such data, but this time sample at non-equidistant points in time, so that $t_{k+1} - t_k$ is not constant over k. The average sample interval was roughly 2 seconds and the samples were distributed uniformly between 1.34 and 2.68 seconds (i.e. the ratio of largest to smallest sample interval is 2).

This "continuous" data was then sampled via (5), where we chose a constant integrating time δ_k equal to the minimum sample interval (1.34 seconds), i.e. $\delta_k = 1.34$ for all k, but $\Delta_k = t_{k+1} - t_k$ changes over k. This corresponds to an assumption that the sampling device is not adaptive over time (this assumption can be easily removed).
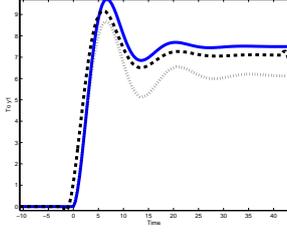
4

Fig. 4. Step response for identified continuous time system. Sampling interval 2. Solid: True system. Dotted: Estimate using simplistic sampling. Dashed: Estimate using simplistic sampling on anticausally shifted data. The step response of the identified continuous system using the correct sampling formulas coincides with the true system in this resolution.
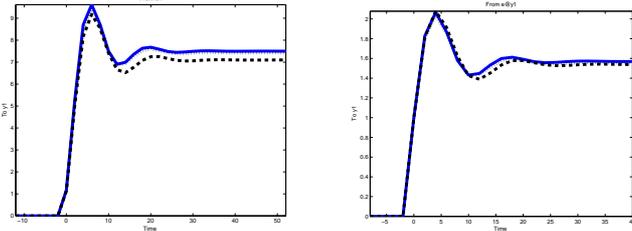

Fig. 5. Step response for identified discrete time system. Sampling interval 2. Solid: True system. Dotted: Estimate using correct sampling. Dashed: Estimate using simplistic sampling. Left: from input, Right: from noise source. The dotted lines coincide with the solid lines in this resolution.

Figure 6 shows a snapshot of the continuous and sampled data.

With the above scheme we obtained $N = 1280$ samples and based on these samples we estimated the parameters $A_{1,1}$, $A_{2,1}$, $B_{1,1}$, $B_{2,1}$, $K_{1,1}$ and $K_{2,1}$ of (12) based on the correct and simplistic sampling formulas. For the case of correct sampling, the Kalman filtering equations (A.2), (8) and (9) were used to form the predictor.

Figure 7 compares the step responses from correct and simplistic sampling models to the true model. By way of information Table 3 shows the parameter values and their standard deviations for both sampling cases. Notice that the correct sampling gives more accurate results than the simplistic case, as expected.

Par	True	Est. Correct samp	Est. Simpl.samp
$A_{1,1}$	-0.3567	-0.3662 ± 0.0133	-0.3713 ± 0.0159
$A_{2,1}$	-0.2426	-0.2436 ± 0.0055	-0.2487 ± 0.0076
$B_{1,1}$	0.2540	0.2687 ± 0.0822	0.9387 ± 0.0882
$B_{2,1}$	1.8198	1.8675 ± 0.0500	1.6867 ± 0.0621
$K_{1,1}$	1.2865	1.4170 ± 0.2072	0.7971 ± 0.1070
$K_{2,1}$	0.7864	0.7840 ± 0.1068	0.4946 ± 0.0570

Table 3
Estimated parameter values for non-uniformly sampled example, ± indicates standard deviations.
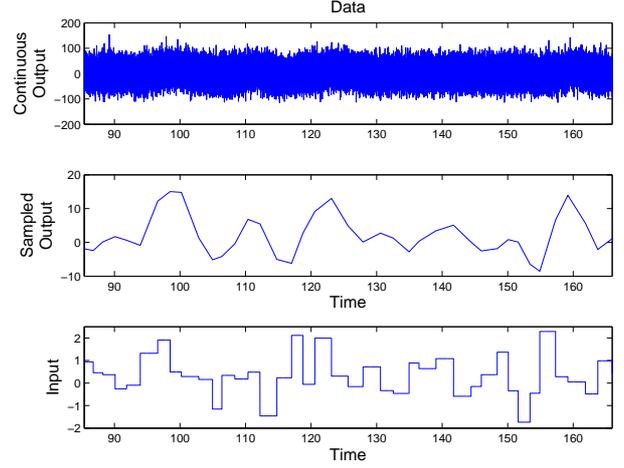
A few comments are in order:

Fig. 6. Data for non-uniformly sampled example. Top: "continuous" output (0.001 second sample interval). Middle: sampled output (non-uniform sampling). Bottom: input.
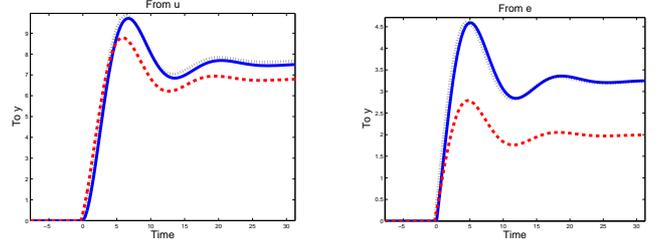

Fig. 7. Step response for identified continuous time system with non-uniformly sampled data. Solid: True system. Dotted: Estimate using correct sampling. Dashed: Estimate using simplistic sampling. Left: from input. Right: from noise source.

- In both examples, the main reason why the simplistically sampled system gives bad estimates is that it does not provide a direct D-matrix term. The sampled data has such a term due to the integrated sampling (its value is 1.17 for the 2 s equidistant sampling interval). The model suffers from not being able to explain this effect.
- The model fit is done in the metric of the sampled data, even though the parameters relate to continuous time. Comparing the discrete time responses of the models (according to their own sampling rules) with that of the correctly sampled system shows less discrepancies compared to the parametric errors, as seen in Figure 5.

6 A Collection of Why's

Why Are Continuous Time Models of Interest?
The main reason why we deal with continuous time models is that physical insight and intuition are typically best represented in continuous time. Most physical modeling work ends up with a continuous time model, and in order to make use of that knowledge it is natural to parametrize a model (1) accordingly. Even though the fit to data will have to be done in discrete time as in

(13), the dimension of the vector θ reflecting continuous time dynamics will be lower in that way.

For equidistantly sampled data one could build a black-box discrete time model, that is converted to continuous time afterwards, using the inverse sampling formulas. However to comply with any physical structure of (1) this would involve another round of optimization. If there is no physical knowledge, so (1) is black-box, and the sampling time is constant, a simple route is to first build a discrete time model and than convert it to continuous time. This route is essentially equivalent to estimating the continuous model directly for this case.

In the case of unequally sampled data, the discrete time system (8) is time varying, even if the underlying continuous system is time invariant. Then this continuous time model is the natural basis to deal with the sampled observations. This is the pragmatic reason for working with the abstract time-continuous noise models: It allows us to work with time invariant parametrization of the noise properties.

Why Is It Essential to Have Correct Sampling Formulas? Well, actually it is not that essential. You could have any weird sampling formula from A, B, C, K to $F, G, H, D, \bar{K}$ that is surjective. The model fit is done in the metric of the sampled systems, so the resulting sampled system would be the best discrete time model available. The underlying parametrization of the continuous time model may however be irrelevant. Then, on the other hand, you could as well fit a discrete time model directly. So, only if you have a genuine interest in the continuous time model (see previous question), are the sampling formulas essential.

Why Haven't the Correct Sampling Formulas Been Extensively Used? The sampling formulas in Appendix A are not new, and not difficult to derive. Still the results for integrating sampling are not widely used. They are for example typically not given in textbooks and the MATLAB CONTROL SYSTEM TOOLBOX offers no implementation. (There is more attention to the conceptually related, but different, issue of piecewise linear inputs – 'First order hold' –, i.e. an integrator between a piecewise constant input source and the actual input.) Also, we are not aware of any explicit inverse formulas ("d2c") for the integrating sampling case.

In a way this is a case of "double standards:" On the one hand the observations of the ideal continuous time description (1) will need lowpass filtering to achieve realistic measurements. On the other hand one is not willing to take the consequences of this when it comes to the effects of the input. The instantaneous sampling, giving $H = C$ and $D = 0$ are simpler and more attractive. The engineering rules of thumb say that a signal should be filtered by a presampling filter with cutoff frequency below the Nyquist frequency. This motivates the use of $\delta = \Delta$ for $t_k = k\Delta$ in (5). For the double standards to be acceptable, one must consequently sample fast enough

(Δ small enough) so that the effects on the dynamics of $H \neq C$ and $D \neq 0$ in (9) are negligible.

Another view on this, that is applicable to the equidistantly sampled case, is to build a discrete time model that describes the observed data as well as possible, and then regard any sampling devices, including presampling filters as part of the model. But, again, if a continuous time model is essential, one must be careful in converting back to such a model.

7 Conclusions

The main conclusion is that estimating continuous time systems requires some care and understanding of how the sampling was done. There is always some kind of averaging or low-pass filtering involved in the A/D conversion of measured data. If the sampling is fast compared to the system dynamics of interest this filtering/averaging effect may be negligible compared to the dynamics. We have shown in this contribution how to properly handle the case of integrating sampling, which is one variant of real sampling. The effect of the estimated models may be significant for slow sampling. The use of continuous time models may be tied to an interest in physical models, i.e. "grey-box" identification. It could also be due to non-equidistantly sampled data, for which the continuous time model is the natural basis.

When dealing with a realistic sampling model, several issues arise in the choice of sampling interval Δ. A short interval gives higher variance of the output measurement, but this is basically compensated for by obtaining more measurements over a given time period. For any sampling formula to be correct it is also essential that the true, continuous time input can be correctly reconstructed from the sampled measurements. This may also influence the choice of Δ. Finally it is clear that the issues raised in this paper are also relevant for possible down-sampling after the experiment.

A   Computing $F_k, G_k, H_k, D_k, \tilde{R}_1, \tilde{R}_{12}, \tilde{R}_2$

For reference, the duration between successive time samples t_k and t_{k+1} is denoted by Δ_k, i.e. $\Delta_k = t_{k+1} - t_k$ With this in place, the following Result provides formulas for sampling (1) based on (5), together with a numerically stable implementation based on Bernstein et al. (1986) and Van Loan (1978).

Result 1 *Assume that the input is piecewise constant so that $u(t) = u(t_k) = u_k$ for $t_k \leq t < t_{k+1}$ and assume that the outputs are observed via integrated sampling given by (5). Denote $\Delta_k = t_{k+1} - t_k$, $x_k = x(t_k)$ and $y_k = y(t_k + \delta_k)$, then*

$$x_{k+1} = F_k x_k + G_k u_k + \tilde{w}_k \qquad \text{(A.1a)}$$
$$y_k = H_k x_k + D_k u_k + v_k \qquad \text{(A.1b)}$$

where the noise terms are Gaussian distributed via

$$\begin{bmatrix} \tilde{w}_k \\ v_k \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tilde{R}_1(k) & \tilde{R}_{12}(k) \\ \tilde{R}_{12}^T(k) & \tilde{R}_2(k) \end{bmatrix} \right) \quad \text{(A.1c)}$$

The matrix terms are given by

$$F_k = M_{33}^T(\Delta_k), \qquad G_k = M_{34}^T(\Delta_k)B \quad \text{(A.1d)}$$

$$H_k = \frac{1}{\delta_k} C M_{34}^T(\delta_k), \qquad D_k = \frac{1}{\delta_k} C M_{35}^T(\delta_k)B \quad \text{(A.1e)}$$

$$\tilde{R}_1(k) = M_{33}^T(\Delta_k) M_{34}(\Delta_k) \quad \text{(A.1f)}$$

$$\tilde{R}_{12}(k) = \frac{1}{\delta_k} \left(M_{33}^T(\delta_k) M_{24}(\delta_k) C^T + M_{34}^T(\delta_k) R_{12} \right) \quad \text{(A.1g)}$$

$$\tilde{R}_2(k) = \frac{1}{\delta_k^2} C \left(M_{33}^T(\delta_k) M_{14}(\delta_k) + M_{14}^T(\delta_k) M_{33}(\delta_k) \right) C^T$$
$$+ \frac{1}{\delta_k^2} \left(C M_{35}^T(\delta_k) R_{12} + R_{12}^T M_{35}(\delta_k) C^T \right) + \frac{1}{\delta_k} R_2 \quad \text{(A.1h)}$$

where each $M_{ij}(\cdot)$ denotes an $n \times n$ block matrix defined by the matrix exponential $M(\tau) = \exp(\Pi\tau)$, where

$$M(\tau) = \begin{bmatrix} M_{11}(\tau) & M_{12}(\tau) & M_{13}(\tau) & M_{14}(\tau) & M_{15}(\tau) \\ 0 & M_{22}(\tau) & M_{23}(\tau) & M_{24}(\tau) & M_{25}(\tau) \\ 0 & 0 & M_{33}(\tau) & M_{34}(\tau) & M_{35}(\tau) \\ 0 & 0 & 0 & M_{44}(\tau) & M_{45}(\tau) \\ 0 & 0 & 0 & 0 & M_{55}(\tau) \end{bmatrix}$$

$$\Pi = \begin{bmatrix} -A & I & 0 & 0 & 0 \\ 0 & -A & R_1 & 0 & 0 \\ 0 & 0 & A^T & I & 0 \\ 0 & 0 & 0 & 0 & I \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The expressions for (A.1)(a–c) are presented in Bernstein et al. (1986) and are not new. Formulas for computing the required matrices based on the matrix exponential expressions for $M(\tau)$ are partly provided in Bernstein et al. (1986), but for a slightly different case where $\Delta_k = \delta_k$. However, by referring to the original work of Van Loan (1978) it is possible to straightforwardly extend the results in Bernstein et al. (1986) to the case where $\Delta_k \neq \delta_k$. In addition, the above implementation employs fewer matrix exponential calculations.

Using the above sampling formulas for $F_k, H_k, \tilde{R}_1, \tilde{R}_{12}, \tilde{R}_2$, the time-varying Kalman gain $\tilde{K}_k$ can be computed using the following standard expressions:

$$P_{k+1} = F_k P_k F_k^T + \tilde{R}_1(t_k, \Delta_k) - \tilde{K}_k \Lambda_k \tilde{K}_k^T \quad \text{(A.2a)}$$

$$\tilde{K}_k = (F_k P_k H_k^T + \tilde{R}_{12}(t_k, \delta_k)) \Lambda_k^{-1} \quad \text{(A.2b)}$$

$$\Lambda_k = H_k P_k H_k^T + \tilde{R}_2(t_k, \delta_k) \quad \text{(A.2c)}$$

References

Bernstein, D., Davis, L. & Greeley, S. (1986), 'The optimal projection equations for fixed-order sampled-data dynamic compensation with computation delay', *IEEE Transaction on Automatic Control* **31**(9), 859–862.

Feuer, A. & Goodwin, G. C. (1996), *Sampling in Digital Signal Processing and Control*, Birkhauser.

Garnier, H., Mensler, M. & Richard, A. (2003), 'Continuous-time model identification from sampled data: Implementation issues and performance evaluation', *Int. J. Control* **76**(13), 1337–1357.

Garnier, H. & Wang, L., eds (2008), *Identification of Continuous-time Models from Sampled Data*, Springer-Verlag, London.

Gerdin, M., Schön, T., Glad, T., Gustafsson, F. & Ljung, L. (2007), 'On parameter and state estimation for linear differential-algebraic equations', *Automatica* **43**, 416–425.

Goodwin, G. C., Graebe, S. F. & Salgado, M. E. (2001), *Control System Design*, Prentice Hall, Upper Saddle River, New Jersey.

Goodwin, G., Mayne, D. Q. & Feuer, A. (1995), 'Duality of hybrid optimal regulator and hybrid optimal filter', *Int. J. Control* **61**(6), 1465–1471.

Jazwinski, A. (1970), *Stochastic Process and Filtering Theory*, Vol. 64 of *Mathematics in Science and Engineering*, Academic Press, New York.

Johansson, R., Verhaegen, M. & Chou, C. T. (1999), 'Stochastic theory of continuous-time state-space identification', *IEEE Trans. Signal Processing* **47**(1), 41–51.

Ljung, L. (1999), *System Identification - Theory for the User*, 2nd edn, Prentice-Hall, Upper Saddle River, N.J.

Ljung, L. (2007), *The System Identification Toolbox: The Manual*, The MathWorks Inc. 1st edition 1986, 7th edition 2007, Natick, MA, USA.

Shats, S. & Shaked, U. (1991), 'Discrete-time filtering of noise correlated continuous-time process: Modeling and derivation of the sampling period sensitivities', *IEEE Trans. Autom. Control* **36**(1), 115–119.

Van Loan, C. (1978), 'Computing Integrals Involving the Matrix Exponential', *IEEE Transaction on Automatic Control* **23**(3), 395–404.

Young, P. C. (1981), 'Parameter estimation for continuous-time models - a survey', *Automatica* **17**(1), 23–39.

Yuz, J. I. (2006), Sampled-data models for linear and nonlinear systems, PhD thesis, School of Electrical Engineering and Computer Science, University of Newcastle, Australia.

Åström, K. J. (1970), *Introduction to Stochastic Control Theory*, Academic Press, New York.

Åström, K. J. (1980), 'Maximum likelihood and prediction error methods', *Automatica* **16**, 551–574.