Project-Based Teaching

A CASE STUDY OF LEARNING SYSTEMS THEORY AND SIGNAL PROCESSING BY A DYNAMIC MEASUREMENTS PROJECT

Ivan Markovsky

Summary

Research shows that, in learning science and engineering, guided project work leads to deeper understanding of theoretical concepts (as well as acquisition of hands-on skills) than the classical approach of textbook reading and attending lectures. In an approach to education based on project work, the role of the teacher is to create a stimulating learning environment and to supervise the students in accomplishing their objectives. The main challenge is to come up with projects that are engaging, diverse, and feasible in view of limited time and resources. In this paper, we describe such a signal processing project. The task is to improve the speed and accuracy characteristics of a sensor by real-time signal processing. It turns out that this is an application of Kalman filtering however the students need to identify a model of the sensor and implement the Kalman filter on a DSP. The project consists of three main tasks: 1) mathematical formalization of the problem, 2) development of solution methods, and 3) implementation and testing of the methods. The testing is done on an inexpensive laboratory setup, using the Lego Mindstorms educational kit in combination with a temperature sensor. The learning outcomes are understanding of model representations, system identification, and state estimation, as well as implementation in Matlab and C of real-time signal processing algorithms. Possible extensions are adaptive signal processing, multiple sensors data fusion, and non-constant measured value estimation.

> "Success in the rapidly changing world of the future depends on being able to do well what you were not taught to do." Seymour Papert [1]

The Moore method in mathematics is to let students rediscover the theory that they are learning [2], [3]. This is done individually or in small groups, where every student has a personal contribution. The teacher's role is to supervise the work of the students, suggesting where needed possible ways of overcoming difficulties. The suggestions are hints that are just enough to direct the students into making the discovery on their own. The Moore method is developed and used for teaching of advanced mathematics courses [4]. The idea of learning through self-discovery, however, is generally applicable in education starting from early childhood [1].

Applied to science and engineering, the Moore method implies working on open-ended projects, *i.e.*, problems that are not fully specified and allow for different interpretations. They should be challenging but feasible for the students, stimulating them to apply knowledge that they have already learned in a classical lecture-based course or need to discover by doing the projects [5]–[7]. Apart from the application of already known theory and methods and learning new ones, the objective is to encourage free exploration, critical thinking, and creativity. This necessarily implies having less structure than in the traditional lecture-based education [8], see Sidebars "Structure vs Free-Exploration" and "Results-Driven vs Process-Driven Mindset".

The topics of the projects are irrelevant as long as they are motivated by and have connection to real-life applications. The connection to applications makes the projects meaningful for the students. By recognizing the practical usefulness of their work, students are more likely to engage with the projects [9].

The motto of *self-discovery by free exploration* is at the core of the program for learning systems theory presented in [10]. It consists of a set of open ended curiosity-driven questions related to the free-fall dynamics. The key step the students need to accomplish is translation of the open ended questions into well-defined mathematical problems. Indeed, problem formulation is an important but often neglected skill. Like any other skill, however, it can be developed by deliberate practice [11]. As a positive side-effect, the freedom of exploring different problem

1066-033X/20©2020IEEE

STRUCTURE VS FREE-EXPLORATION

The project is motivated by and aims at solving a nontrivial real-life problem. Following the Moore method, the students are expected to work independently (individually or in groups), freely explore, and solve the problem on their own. The students are guided implicitly by being given a sequence of open-ended problems *to solve*. They receive feedback, self-study materials, and directions for further work through discussions with peers and supervisors but no help with technical difficulties. Thus, the students are fully responsible. However, they fully own the results because they are their own achievement.

From a pedagogical perspective, working on challenging but feasible open-ended problems, the students gradually build the necessary experience and confidence to solve the ultimate real-life problem. Depending on their background and experience, they may need to search for and learn new theory and tools. In comparison with traditional teaching, which first offers lectures on theoretical material and then lets students use *this* theory in projects, the free exploration approach is unstructured. The structure imposed by the traditional approach, however, is constraining, leaving less room for creativity. Typically, problems are well-defined and students receive a sequence of steps *to follow*.

Switching the teacher's / student's mindset about teaching / learning from "tell / being told what to do" to "ask / answer questions" is the hallmark of the proposed project-based teaching.

RESULTS-DRIVEN VS PROCESS-DRIVEN MINDSET

A free-exploration project is never complete: there is always a next "What if . . . ?" question inspiring further investigation. At any stage, success is measured by the extent to which the students are engaged and stimulated to think creatively, not by accomplishment of predefined objectives. It is normal, even expected, for students to define, work on, and solve different problems. They are free to use any theory, solution approach, and tools (*e.g.*, programming languages). Although there are no constraints in the exploration stage, the ultimate test for the solutions found is the real-life testbed, see Section "Implementation and Validation of the Method". Thus, the performance on the application is the final criterion for judging the results. The pros and cons of the approaches followed by the students are then compared and discussed in a group session.

The learning outcome is the experience gained from the exploration rather than the result [S1]. A method not performing "well" is not a failure. It is a part of the learning process. Students learn more in "unsuccessful" attempts to develop their own methods than in "successful" reproduction of established methods following step-by-step instructions. Switching the teacher's / student's mindset about teaching / learning from "tell / being told what to do" to "ask / answer questions" is the hallmark of the proposed project-based approach to teaching.

REFERENCES

[S1] C. Dweck, *Mindset: The New Psychology of Success*. Random House, 2006.

formulations makes the students more involved, because they feel ownership of the problems that they solve. This higher level of involvement in turn leads to gaining deeper and longer-lasting knowledge [5]–[7].

This paper presents a project for learning systems theory and signal processing motivated by the Moore method, learning by exploration, and the importance of real-life applications. The high-level problem considered in the project is

> *Dynamic measurement:* improve the speed and accuracy of a sensor by real-time signal processing.

The project goes gradually from simple to complex, making at first simplifying assumptions that are relaxed later on, see Sidebar "From Simple to Complex". Section "Dynamic Temperature Measurement" presents a project of temperature measurement. It starts with a motivating exercise which is a simplified version of the problem in the project and thus prepares the students to tackle the coming difficulties.

Section "Extensions" presents follow-up projects of weight measurement, fusing measurements of multiple sensors, and measurement of a non-constant quantity. The follow-up projects can be linked to and solved by generalizations of the Kalman filter. The assumption that the sensor dynamics is a priori known, however, is unrealistic in these scenarios. Therefore, an alternative direct data-driven approach, which estimates the quantity of interest directly from the measured data is developed [12], [13]. The alternative approach is a data-driven version of the Kalman filter, which does not require a given model [14].

The project presented in Section "Dynamic Temperature Measurement" is suitable for undergraduate students with specialization in systems and control or signal processing. The minimum background requirements are basic linear algebra and linear time-invariant systems theory, as well as experience with Matlab (or a similar high-level programming language) for the numerical simulations and C for the implementation of the method on a digital signal processor. The project corresponds to one study point (25 hours of work). Section "Extensions" presents directions for further work that are suitable for advanced undergraduate and master level students.

DYNAMIC TEMPERATURE MEASUREMENT

The sample project presented in this section illustrates the motto of the paper—*learning by exploration*. It is split into three steps.

1) Understand the high-level problem statement "improve the speed and accuracy characteristics of a sensor by real-time signal processing" and formalize it by formulating a well defined mathematical problem.
2) Develop a method for solving the mathematical problem.
3) Implement and validate the method in practice.

Problem Understanding

As a motivating exercise, the students are presented with the following problem from [15, Page 53]:

> A thermometer reading 21°C, which has been inside a house for a long time, is taken outside. After one minute the thermometer reads 15°C; after two minutes it reads 11°C. What is the outside temperature? (According to Newton's law of cooling, an object of higher temperature than its environment cools at a rate that is proportional to the difference in temperature.)

The exercise is a simplified version of the project and shows the underlying idea that a "slow" processes can be made faster by data processing. In the project, the students explore a generalization of the exercise and its solution to higher order multivariable processes, noisy data, and real-time data acquisition.

The first task of the students is to understand the problem. They should be able to answer the questions [11]:

- What is the given data and what is to be found?
- How are the data and the to-be-found quantity related?

In the dynamic measurement setup, the given data are the sensor's readings. In practice, the sensor's readings are sampled in time, quantized, and collected in real-time. Initially, the students sidestep the complications resulting from the quantization and real-time data collection by assuming that the data are exact and available in batch. Later on, however, they remove these assumptions and eventually derive a practical solution method. In particular, the real-time processing of the data is an essential aspect for the implementation of the method in practice.

The to-be-found variable is the measured quantity. In metrology, it is natural to assume that it is constant over the measurement time. Students who are interested in relaxing this assumptions, *i.e.*, consider tracking of a time-varying quantity, may explore this in a follow-up project.

We assume that the measurement process starts at the moment of time when the measured quantity exhibits a step change. Moreover, without loss of generality, we assume that the step change happens at time zero. It initiates then a transient process. The transient is the essential object that is analyzed in the dynamic measurement problem.

The connection between the data and the measured variable is the sensor dynamics. The sensor is based on a physical process, *e.g.*, heat exchange between the environment and the thermometer in the temperature measurement case. This is a link between engineering and physics, which leads to modeling using the underlying physical laws.

Once the problem is understood, the students are given time to work on the solution. This is important irrespective of whether or not they succeed. Even unsuccessful attempts lead to deeper understanding. After the students have spent sufficient time on independent work, the instructor summarizes the proposed solutions and attempted ideas in a group session. This is an opportunity for the instructor to demonstrate in practice that the best way to come up with a "good" solution is to come up with many solutions. A good solution is conceptually simple (a proxy of simplicity is conciseness) and has the potential to be generalized. The generalizations of interest for the project are:

- higher order dynamical processes,
- more measurements, which may be perturbed by noise,
- computational methods that are suitable for real-time implementation.

Sidebar "Sample Solution to the Motivating Exercise" presents two solutions, which expose the difficulty of the problem. The second solution uses an idea of differencing the measurements, which is used further on in the project.

Mathematical Formalization of the Problem

After having successfully solved the problem in the motivating exercise, the students are confronted with the generalization of their solutions to multiple noisy measurements, higher order processes, and real-time data processing. In doing this students should employ a more general setting of systems theory. The link between the problem at hand—dynamic measurement—and systems theory is done by modeling the sensor as a dynamical system. Students familiar with the input/output framework of systems theory may think of the sensor as a map that takes as an input the measured variable and produce as an output the sensor's reading, see Figure 1.

The input/output framework imposes a cause-and-effect relation—the input causes the output. This is an opportunity for the instructor to call for critical thinking. From the solution of the motivating exercise (see Sidebar "Sample Solution to the Motivating Exercise"), the students know that the actual process behind the temperature measurement is the heat exchange between the environment and the thermometer defined by Newton's law of cooling. It does not have a cause-and-effect relation. Thus, the input/output model does not faithfully reflect the physical reality—it adds a causal relation that does not exist.

Problem formulation is an important but often ignored skill. Like any other skill it can be developed by deliberate practice.

FROM SIMPLE TO COMPLEX

A series of open-ended problems implicitly guides the students' work. The problems reveal essential features of the general problem considered and build on each other. There are several aspects of the high-level dynamic measurement problem that allow us to go gradually from simple to complex.

Model of the Sensor's Dynamics

A basic assumption made from the start of the project is that the sensor's dynamics is linear time-invariant. Still, there are three important aspects to be considered.

Is the model given? Assuming that a model is given significantly simplifies the problem. The key observation is that the dynamic measurement problem is equivalent then to a state estimation problem for an autonomous system. Dealing with unknown sensor's dynamics requires model identification or direct data-driven methods.

Model order The project starts with an example of a first order process—temperature measurement—which has an exponential response. As another physical example the students may be presented with a weight measurement process, which is second order. Seeing concrete physical examples helps building intuition for the abstract general case of an nth order system.

Scalar or multivariable? Another aspect of the sensor's dynamics that allows transition from simple to complex is the number of variables. Initially, the problem is aimed at measurement of one variable via one sensor. Later on students may consider a generalization to measurement of multiple variables using multiple sensors, which in the metrology application has a *"data fusion"* interpretation.

Measured Data

Exact or noisy? The students are instructed to solve the problem first under the simplifying assumption that the data are exact. Although this is an unrealistic assumption in practice, it leads to a solution method that can be used later on for solving the problem with noisy data. It turns out that the modification needed is solving a least-squares approximation problem instead of a system of linear equations—a minor change of the solution method.

Sampling and quantization Apart from being noisy, in practice the data will be sampled and quantized. The students consider a real-valued discrete-time process, *i.e.*, the sampling is already done and there are no quantization errors. The choice of the sampling frequency and the effect of the quantization errors are considered, however, at the stage of the practical implementation and testing of the method.

Off-line or real-time data processing? The goal of dynamic measurement is real-time data processing. Initially, however, the students should solve the problem assuming that all the data are given in batch (off-line processing). Then they should consider the problem of making the batch method recursive, so that it can be used for real-time data processing.

Constant or non-constant measured variable? The starting assumption is that the to-be-measured variable is constant in time. The generalization to a non-constant measured variable can be done later on in two different ways. First, assuming a known model of the time variation, *e.g.*, linear, the problem is again reduced to a state estimation problem of an augmented system that includes the model of the time-variation and the sensor dynamics. Second, without prior knowledge of the time-variation, adaptive data-driven methods can be used.

The best way to come up with a good solution is to come up with many solutions. A "good" solution is conceptually simple (a proxy of simplicity is conciseness) and has the potential to be generalized.

SOLUTION OF THE MOTIVATING EXERCISE

Let $y(t)$ be the thermometer's reading at time t and $\bar{u}$ the environmental temperature. The cooling process is described by the first order linear constant coefficients differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}y(t) = \alpha\big(\bar{u} - y(t)\big), \qquad \text{for } t \geq 0,$$

where $\alpha > 0$ is an unknown constant (dependent on the environment).

Solution 1

The general solution of the cooling process differential equation

$$y(t) = \bar{u} + \underbrace{\big(y(0) - \bar{u}\big)e^{-\alpha t}}_{\text{transient}}, \qquad \text{for } t \geq 0,$$

leads to a nonlinear system of equations for $\bar{u}$ and $a := e^{-\alpha}$

$$\begin{aligned} y(1) &= ay(0) + (1-a)\bar{u} \\ y(2) &= a^2 y(0) + (1-a^2)\bar{u}. \end{aligned} \tag{$*$}$$

Although there is no general method for solving nonlinear systems, some students may notice that writing the system as

$$\begin{aligned} y(1) - ay(0) &= (1-a)\bar{u} \\ y(2) &= a^2 y(0) + (1+a)(1-a)\bar{u}. \end{aligned}$$

and substituting the first equation into the second equation eliminates $\bar{u}$. This leads us to a linear equation for the unknown a,

$$y(2) - y(1) = a\big(y(1) - y(0)\big).$$

Using the given data for $y(0), y(1), y(2)$, we find $a = 2/3$. With a known, $(*)$ is linear in $\bar{u}$ and we can solve it, finding $\bar{u} = 3^\circ$C.

Solution 2

The key observation from $(*)$ is that the problem is hard due to the unknown constant α. With α known, $(*)$ becomes linear in $\bar{u}$. Another solution to this problem is to use the difference operation $\Delta y(t) := y(t) - y(t-1)$ in order to eliminate the constant $\bar{u}$. The differenced signal Δy satisfies the homogeneous first order constant coefficients differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\Delta y(t) = \alpha \Delta y(t), \qquad \text{for } t \geq 0.$$

For $t = 0, 1, \ldots$, we obtain the difference equation

$$\Delta y(t+1) = a\Delta y(t). \tag{$**$}$$

Using the data, we find $\Delta y(0) = -6$ and $\Delta y(1) = -4$. Then, from $(**)$, we find $a = 2/3$. Finally, using $(*)$, we have $\bar{u} = 3^\circ$C.

Potential for Generalization

With more data $\big(y(0), y(1), \ldots, y(T)\big)$, $(*)$ becomes an overdetermined system of equations. If the data are noisy, least-squares approximate solution of $(*)$ allows us to estimate $\bar{u}$. The least-squares solution is also suitable for real-time implementation. Thus, the solution of the exercise is directly generalizable to more data, noisy measurements, and real-time implementation.

For processes defined by higher-order linear constant coefficients differential equations, a closed-form solution is still available (sum of polynomials-times-exponentials signal), however, the estimation of the model parameters (frequencies and dampings) becomes more involved. The key challenge in tackling higher-order processes is again the model's parameters estimation.

$$\begin{array}{c}\text{to-be-measured value } \bar{u} \\ \text{(constant during measurement)}\end{array} \xrightarrow{\text{sensor}} \begin{array}{c}\text{measured value } y \\ \text{(exhibits transient response)}\end{array}$$

FIGURE 1 The classical approach to systems theory suggests modeling the measurement process as an input-output map, where the input is the to-be-measured variable and the output is the value measured by the sensor.

The classical approach forces us to choose an input/output partitioning. An alternative approach that does not a priori impose an input/output structure of the model is the *behavioral approach* [16]. It defines the model as a set of trajectories

$$\mathscr{B} := \big\{\, y \mid \tfrac{\mathrm{d}}{\mathrm{d}t}y(t) = \alpha\big(\bar{u} - y(t)\big), \\ \text{for } t \geq 0 \text{ and } y(0) \in \mathbb{R} \,\big\}, \tag{DE}$$

where $\bar{u}$ is the to-be-measured value (refer to Sidebar "Sample Solution to the Motivating Exercise"). The notation $y \in \mathscr{B}$ expresses the fact that y is a trajectory of $\mathscr{B}$.

In addition to not enforcing an input/output partitioning, the behavioral approach separates the model from its numerous *representations* by equations. Indeed, a model $\mathscr{B}$ can be defined by different representations. For example, the model (DE) can be defined also by the state-space representation

$$\mathscr{B} = \Big\{\, y = \begin{bmatrix} 0 & 1 \end{bmatrix} x \mid \tfrac{\mathrm{d}}{\mathrm{d}t}x = \begin{bmatrix} 0 & 0 \\ \alpha & -\alpha \end{bmatrix} x \\ \text{for all } t \geq 0 \text{ and } x(0) = \begin{bmatrix} \bar{u} \\ y(0) \end{bmatrix} \Big\}. \tag{SS}$$

The choice of the representation is important for derivation of solution methods but should not play a role in problem formulations. For example, the statement $y \in \mathscr{B}$ is independent of how $\mathscr{B}$ is defined and how $y \in \mathscr{B}$ can be checked in practice. In the behavioral setting, problems (see Sidebar "The Kalman Filter/Smoother") are expressed in terms of the set $\mathscr{B}$ (the behavior) and are thus separated from subsequent methods used for solving them. This allows us to use different methods and helps us establish connections among (seemingly unrelated) methods. For example, the problem at hand can be solved using transfer function as well as state-space techniques. The resulting methods are different but ultimately give the same solution to the original problem.

The solution presented next uses the state-space representation (SS) (and its generalization (SS') to higher-order measurement processes). It allows us to employ existing methods for solving the dynamic measurement problem. The class of measurement process should satisfy the following assumptions.

1) *Bounded complexity LTI sensor dynamics:* The sensor is a bounded complexity linear time-invariant system. Equivalently, it is defined by a linear constant coefficients differential/difference equation in continuous/discrete-time [17, Theorem III.1].
2) *Constant to-be-measured quantity:* The to-be-measured variable $\bar{u}$ is constant during the measurement period.
3) *Calibrated sensor:* In steady-state, *i.e.*, when y is a constant, $y = G\bar{u}$, where G is a priori known.

The link to systems theory allows us to access generic methods that apply for a larger class of sensors. By assumptions 1–3, $y = G\bar{u} + y_{\text{trans}}$, where the transient y_{trans} is a sum of polynomials-times-exponentials signal. It can be written as

$$y_{\text{trans}}(t) = CA^t x_{\text{ini}},$$

for a vector $C \in \mathbb{R}^{1\times n}$ and a matrix $A \in \mathbb{R}^{n\times n}$ that describe the sensor and a vector $x_{\text{ini}} = x(0) \in \mathbb{R}^{n\times 1}$ that specifies the initial conditions. The measurements are then given by

$$\begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(T-1) \end{bmatrix} = \underbrace{\begin{bmatrix} G \\ G \\ \vdots \\ G \end{bmatrix}}_{\mathscr{G}} \bar{u} + \underbrace{\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{T-1} \end{bmatrix}}_{\mathscr{O}} x_{\text{ini}}. \qquad \text{(SE)}$$

Like $(*)$, (SE) is a system of linear equations for $\bar{u}$. With known sensor's model, *i.e.*, A and C known, (SE) can be solved in closed-form using the pseudo-inverse $(\cdot)^\dagger$

$$\begin{bmatrix} \widehat{u} \\ \widehat{x}_{\text{ini}} \end{bmatrix} = \begin{bmatrix} \mathscr{G} & \mathscr{O} \end{bmatrix}^\dagger y. \qquad \text{(SOL)}$$

When the observed data y is measured with additive noise $\widetilde{y}$, *i.e.*, $y = \overline{y} + \widetilde{y}$, where $\overline{y}$ is the noise free data, (SOL) gives the least-squares estimate of $\bar{u}$, which is a solution to the problem

$$\text{minimize} \quad \text{over } \widehat{u} \text{ and } \widehat{x}_{\text{ini}} \quad \left\| y - \begin{bmatrix} \mathscr{G} & \mathscr{O} \end{bmatrix} \begin{bmatrix} \widehat{u} \\ \widehat{x}_{\text{ini}} \end{bmatrix} \right\|. \qquad \text{(LS)}$$

Assuming moreover that the noise $\widetilde{y}$ is zero-mean, white, Gaussian, (SOL) is statistically optimal (maximum-likelihood).

Remark 1 (Internal model of the input)

The model defined by (DE) (or equivalently (SS)) is an *autonomous system*, *i.e.*, it has no inputs. The model of a sensor, depicted in Figure 1, on the other hand is an input-output system. The input u, however, is by assumption constant. Augmenting the input-output model of the sensor with the model $\dot{u} = 0$ (in continuous-time or $\Delta u = 0$ in discrete-time) of the constant input signal u leads to an autonomous system, see Figure 2. This shows that the classical and behavioral approaches for modeling the measurement process are closely related, as they should be since they describe the same underlying phenomenon.

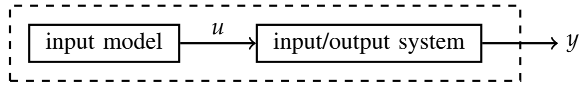

FIGURE 2 Including a model of the input to an input/output system results in an autonomous system.

Augmenting a system with a model of a signal is also used in control [18]. In the dynamic measurement problem, the augmented system has order $n+1$, where n is the order of the sensor. Moreover, the augmented system has an eigenvalue at 0 in continues-time and 1 in discrete-time due to the constant input. Indeed, both (DE) and (SS) have an eigenvalue at 0.

A Method for Solving the Problem

The solution (SOL) obtained using classical state-space methods gives us a block method for the estimation of $\bar{u}$. The next step is to derive a recursive method suitable for real-time implementation.

For the special case of temperature measurement, $G = 1$ and the transient process y_{trans} satisfies the equation $y_{\text{trans}}(t+1) = a y_{\text{trans}}(t)$, *i.e.*, in (SE) $A = a$ and $C = 1$. The solution (SOL) becomes then

$$\begin{bmatrix} \widehat{u} \\ \widehat{x}_{\text{ini}} \end{bmatrix} = \begin{bmatrix} T & \sum_{\tau=0}^{T-1} a^\tau \\ \sum_{\tau=0}^{T-1} a^\tau & \sum_{\tau=0}^{T-1} a^{2\tau} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{\tau=0}^{T-1} y(\tau) \\ \sum_{\tau=0}^{T-1} a^\tau y(\tau) \end{bmatrix}. \qquad \text{(RLS)}$$

The implementation of the algorithm requires about 20 floating point operations.

Although at this point the students have an algorithm for computing $\widehat{u}$, the problem is not solved, because the algorithm requires knowledge of the model's parameters A and C. The next task for the students is to estimate them. In the special case of temperature measurement, the sensor dynamics is described by Newton's law of cooling. Identifying the model is equivalent to determining the coefficient a using measurements of y. For the model identification, the data are available off-line. Existing methods, see [19], [20], can be used. As pointed out in Remark 1, however, the discrete-time system has an eigenvalue at 1. With noisy data, this prior information should be imposed as a constraint on the to-be-identified model. This leads to a nonstandard identification problem. Alternatively the data y can be preprocessed by the difference operator Δ, which removes the constant and results in a standard identification problem.

Implementation and Validation of the Method

Once the students have derived a real-time solution method, they can proceed with its implementation and testing in practice. They should first implement the method in a high-level programming language, such as Matlab, where bug fixes are faster and easier. Then, they can proceed by testing it on data simulated according to the hypothesis of a first order linear time-invariant dynamics and zero mean, white, Gaussian measurement noise.

THE KALMAN FILTER/SMOOTHER

The Kalman filter is one of the most important inventions of the 20th century. It provides an efficient real-time solution to the Wiener filtering problem, *i.e.*, the problem of separating signal from noise [S1]. The first attempts to solve the problem by Wiener and others resulted in solutions that are not suitable for on-line implementation. The success of Kalman's approach is due to the use of the new at the time state-space representation of the system. An alternative view of the Kalman filter is as an optimal state estimator for a linear stochastic system.

Dynamic Measurement by Kalman Filtering

A state-space representation of the measurement process, augmented with the input model, is

$$x(t+1) = \underbrace{\begin{bmatrix} 1 & \\ & A \end{bmatrix}}_{A_{\text{aug}}} x(t), \quad x(0) = \begin{bmatrix} \bar{u} \\ x_{\text{ini}} \end{bmatrix}$$
$$y(t) = \underbrace{\begin{bmatrix} G & C \end{bmatrix}}_{C_{\text{aug}}} x(t) + \tilde{y}. \tag{SS'}$$

Assuming that the measurement noise $\tilde{y}$ is a zero-mean, white, Gaussian with covariance V, the Kalman filter solving the least-squares estimation problem (SOL) in real-time is given by [S1, Theorem 9.2.1]:

$$\begin{aligned} K(t) &= \left(A_{\text{aug}} P(t) C_{\text{aug}}^{\top}\right)\left(V + C_{\text{aug}} P(t) C_{\text{aug}}^{\top}\right)^{-1}, \\ \widehat{x}(t+1) &= A_{\text{aug}} \widehat{x}(t) + K(t)\left(y(t) - C_{\text{aug}} x(t)\right), \\ P(t+1) &= A_{\text{aug}} P(t) A_{\text{aug}}^{\top} - K(t)\left(V + C_{\text{aug}} P(t) C_{\text{aug}}^{\top}\right) K^{\top}(t). \end{aligned}$$

The estimate $\widehat{u}$ of $\bar{u}$ is obtained then from the first component of the state vector $\widehat{x}$.

A Deterministic Interpretation of the Kalman Smoother

In filtering, only past data are used. In smoothing, both future as well as past data are used. The optimal smoothing problem is solved recursively by a variant of the Kalman filtering algorithm, called the Kalman smoother. The smoothing problem has a deterministic interpretation as a projection on the behavior:

$$\widehat{y}^* := \arg\min_{\widehat{y} \in \mathscr{B}} \|y - \widehat{y}\|. \tag{KS}$$

The minimizer $\widehat{y}^*$ or, equivalently, the projection of y on $\mathscr{B}$ is the smoothed signal. The Kalman smoother is thus an efficient algorithm for solving the least-squares minimization problem (KS).

REFERENCES

[S1] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall, 2000.

With the testing on simulated data successfully completed, the students can proceed with a low-level implementation of the algorithm and its testing on the experimental setup (see Figure 3). For the DSP implementation we recommend the C-like language, called NXC (Not eXactly C) [21]. It is simple to use and comes with a convenient GUI development environment.

FIGURE 3 The experimental setup consists of the Lego NXT brick (the DSP) and a digital Lego temperature sensor.

Sample results on data obtained from human temperature measurement and the corresponding optimal fit by the identified model are shown in Figure 4. The real-time prediction on the same data, obtained with the Kalman filter designed for the identified model, is shown in Figure 5. It converges to the measured temperature faster than the natural response of the sensor. This demonstrates the sensor speed-up by the method.

Since the prediction of the Kalman filter is tested on the data that was used for identification, the results are not representative for the actual performance of the method. In Figure 6 we verify the robustness of the Kalman filter, by applying it on new data—the temperature of another subject. The results show similar performance as on the identification data.

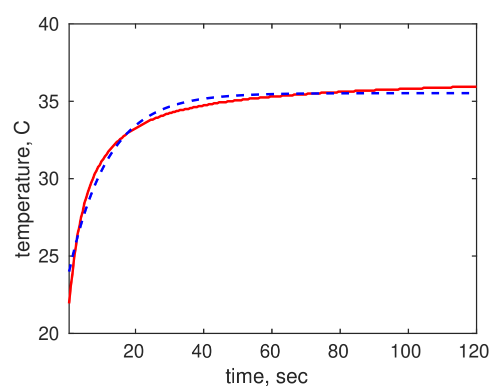

FIGURE 4 The model's output (dashed line) fits the data (solid line).

The project has theoretical as well as practical implementation aspects. In particular, the implementation of the method requires programming in high-level languages, such as MATLAB, for validation on simulation examples, and low-level languages, such as C, for the implementation on the digital signal processor. Writing well documented computer code is therefore important for the successful completion of the project. Software implementation and reproducibility of computational results is an increasingly important issue that needs special

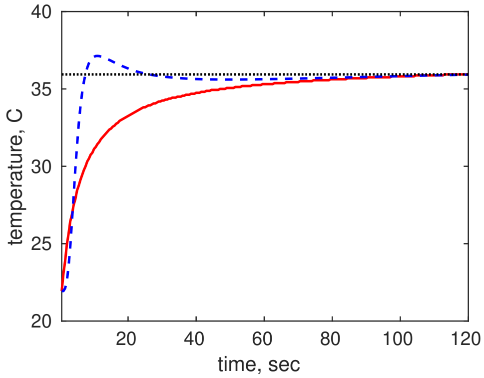

FIGURE 5 The prediction of the Kalman filter (dashed line) converges to the measured temperature (dotted line) faster than the natural response of the sensor (solid line).

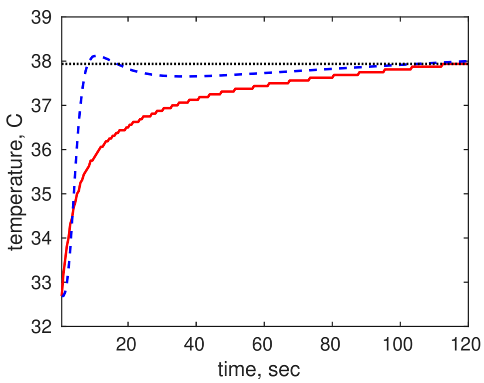

FIGURE 6 Prediction of the Kalman filter (dashed line) on different data than the one used for identification still converges to the measured temperature (dotted line) faster than the natural response of the sensor (solid line). This demonstrates the robustness of the method to model uncertainty.

attention in teaching systems and control. The code development in the context of a scientific project is tightly linked with the associate theoretical development. The classical approaches for code development and documentation however are not well suited for this sort of integration. A solution is provided by the literate programming style, which integrates code and explanation into one document, see Sidebar "Literate Programming and Reproducibility of Computational Results".

EXTENSIONS

The sample project presented in Section "Dynamic Temperature Measurement" is a starting point for follow-up projects on adaptive signal processing, direct data-driven methods, nonlinear and time-varying systems theory. A limitation of the Kalman filter solution is that a model of the sensor must be known a priori. Such a model can be identified off-line from measured data. However, in some cases, *e.g.*, the mass measurement setup described in Section "Dynamic Weighing", the model depends on the unknown measured parameter, so that knowing it is unrealistic. One can use instead adaptive filtering, *i.e.*, identify the model on-line while collecting and filtering the measured data [22]. Alternatively, one can use direct data-driven methods [14]. Adaptive and direct data-driven methods are also needed in situations where the measured value is non-constant and thus the measurement process dynamics becomes time-varying.

The approach based on systems theory yields a method that is directly applicable to higher-order multivariate processes. The possibility to deal with multivariate processes implies that the method can use measurements of multiple sensors. This is explored in section "Fusing Data from Multiple Sensors", which poses the questions of how to build an accurate sensor using several inaccurate sensors and what the limit of achievable performance is. The improved performance is achieved by software rather than hardware. Indeed, the method is applicable to any type of sensor.

Dynamic Weighing

In weight measurement the sensor is a scale, the to-be-estimated parameter is the mass M, and the scale's reading y is the position, see Figure 7. The scale is modeled as a mass-spring-damper system, which has second order linear time-invariant dynamics

$$(M+m)\frac{\mathrm{d}^2}{\mathrm{d}t^2}y = -ky - d\frac{\mathrm{d}}{\mathrm{d}t}y - Mg.$$

In dynamic weighing the assumption that the process dynamics is known is unrealistic because it depends on the unknown mass M. In order to deal with the issue of the unknown process dynamics, an adaptive method that performs simultaneously on-line model identification and filtering is used in [22]. Another approach for dealing with unknown process dynamics is based on a data-driven representation of the model [12]–[14].

The main idea of the data-driven method is to replace the extended observability matrix $\mathscr{O}$ in (LS) with the Hankel matrix

$$\mathscr{H}_L(\Delta y) := \begin{bmatrix} \Delta y(1) & \Delta y(2) & \cdots & \Delta y(T-L+1) \\ \Delta y(2) & \Delta y(3) & \cdots & \Delta y(T+L+2) \\ \vdots & \vdots & & \vdots \\ \Delta y(L) & \Delta y(L+1) & \cdots & \Delta y(T) \end{bmatrix}$$

constructed from the difference signal

$$\Delta y(t) := y(t) - y(t-1).$$

This substitution is possible because under the persistency of excitation assumption

$$\operatorname{rank} \mathscr{H}_{T-n}(\Delta y) = n, \tag{PE}$$

the image of the Hankel matrix coincides with the image of the extended observability matrix

$$\operatorname{image} \mathscr{O} = \operatorname{image} \mathscr{H}_{T-n}(\Delta y).$$

The resulting equation

$$\begin{bmatrix} \mathscr{G} & \mathscr{H}_{T-n}(\Delta y) \end{bmatrix} \begin{bmatrix} \bar{u} \\ \ell \end{bmatrix} = \begin{bmatrix} y(1) \\ \vdots \\ y(T-n) \end{bmatrix} \tag{DD-SOL}$$

LITERATE PROGRAMMING

Literate programming was conceived by Don Knuth during his work on the TeX project [S1].

> At first, I thought programming was primarily analogous to musical composition—to the creation of intricate patterns, which are meant to be performed. But lately I have come to realize that a far better analogy is available: Programming is best regarded as the process of creating *works of literature*, which are meant to be read. Don Knuth [S2]

There are different software systems for literate programming, *e.g.*, the org-mode of the Emacs text editor [S4]. The main advantage of using literate programming is that it allows to integrate code into a document that has text, formulas, figures, and results of the execution of the code. The automation of the process of including figures and numerical results in the document is essential for making the results reproducible in the sense of [S5]. Reproducibility is an important aspect of computational science and engineering where the results are often based on empirical evidence.

REFERENCES

[S1] D. Knuth, *Computers & Typesetting, Volume B: TeX: The Program.* Addison-Wesley, 1986.

[S2] ——, *Literate programming.* Cambridge University Press, 1992.

[S4] C. Dominik, *The Org Manual.* Free Software Foundation, 2024.

[S5] J. Buckheit and D. Donoho, *Wavelets and statistics.* Springer-Verlag, 1995, ch. Wavelab and reproducible research.

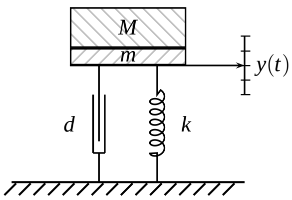

FIGURE 7 In dynamic weighing, *i.e.*, weight measurement taking into account the dynamic properties of the scale, the scale is modeled as a mass-spring-damper system. The model is second order linear time-invariant, however, the unknown mass M affects the sensor's dynamics. This implies that a model of the sensor is not known a priori.

does not involve model parameters and can be solved for the parameter of interest $\bar{u}$. The corresponding method is a recursive least-squares algorithm, which is a data-driven version of the Kalman filter, see Sidebar "Data-Driven Kalman Smoother".

Fusing Data from Multiple Sensors

Consider measuring temperature by two thermometers. The question occurs of how to combine the two measurements into one that is more accurate than either of the individual measurements. The simple idea of averaging (point-wise in time) is problematic. First, the thermometers may start measuring at different moments of time, which means that their transients are not synchronized in time. Second, the initial conditions (*i.e.*, thermometers' initial temperatures) may not be the same, which again leads to a synchronization problem. Finally, the time constants of the thermometers may be different.

Consider for example measuring temperature by a thermometer that is slow but accurate and another thermometer that is fast but inaccurate. The questions of how to combine the two measurements in an optimal way and how to compute the estimate efficiently in real-time are nontrivial. As another example, consider a closed container with constant volume. By Boyle's law, the pressure and the temperature in the container are related. This means that we can fuse data from a thermometer and a pressure sensor.

Assuming that the noise is zero-mean white Gaussian, the solution (SOL) presented in Section "Dynamic Temperature Measurement" gives optimal (in a statistical sense) fusion of the measurements. The data-driven subspace method, described in Section "Dynamic Weighing" is not optimal because (DD-SOL) is a structured errors-in-variables problem whose maximum-likelihood solution is given by the structured total least squares estimator [23]. For computational reasons, however, we use recursive the ordinary least squares method, which gives a suboptimal solution. This makes the scaling of the Kalman filtering and data-driven subspace methods to a large number of sensors computationally feasible.

Non-Constant Measured Value

In Section "Dynamic Temperature Measurement" we assumed that the parameter of interest $\bar{u}$ is a constant. The method described then augments the sensor's model with a model of a constant (see (SS')). The constant value is encoded in the initial condition and is estimated by the Kalman filter. When $\bar{u}$ is non-constant, we can use an autonomous linear time-invariant system as a model for its evolution. In particular, an autonomous linear time-invariant system can describe a linear trend. In the formalization where $\bar{u}$ is modeled by an autonomous linear time-invariant system, the model is given and fixed but the initial condition that correspond to $\bar{u}$ is unknown. The dynamic measurement problem is then solved by a model-based or data-driven Kalman filter, designed for the augmented system—the sensor's model augmented with a model for the input $\bar{u}$. Alternatively, adaptive methods can be used [12], [22]. An easy modification that makes the direct data-driven method adaptive is to introduce in the solution of the recursive least-squares problem windowing and exponential forgetting.

CONCLUSIONS

The motivation for project-based teaching is the gap between the classical university education and the need of hands-on

DATA-DRIVEN KALMAN SMOOTHER

The interpretation of the Kalman smoother as a method for computing the projection

$$\widehat{w}^* := \arg\min_{\widehat{w}\in\mathscr{B}} \|w-\widehat{w}\| \tag{KS}$$

of the data w on the system $\mathscr{B}$ suggests an alternative direct data-driven smoothing method. Next, we consider a general linear time-invariant system $\mathscr{B}$ with m inputs of order n. We aim to solve the smoothing problem (KS) with the system $\mathscr{B}$ unknown but implicitly specified by a trajectory w_{d}.

Let $\mathscr{B}|_T$ be the restriction of $\mathscr{B}$ to the interval $[1,T]$. Under the generalized persistency of excitation condition

$$\operatorname{rank}\mathscr{H}_T(w_{\mathrm{d}}) = Tm+n, \tag{GPE}$$

the *data-driven representation*

$$\mathscr{B}|_T = \operatorname{image}\mathscr{H}_T(w_{\mathrm{d}}) \tag{DD-REPR}$$

holds true, *i.e.*, the finite-horizon behavior $\mathscr{B}|_T$ of $\mathscr{B}$ is fully specified by the data w_{d} [S3, Corollary 21]. The representation (DD-REPR) is inspired by the *fundamental lemma* [S1], which gives alternative *sufficient conditions* for (DD-REPR) that are suitable for *input design* [S2]. The conditions of the fundamental lemma require 1) an input/output partitioning of the data, 2) controllability of the data generating system $\mathscr{B}$, and 3) persistency of excitation of an input component of w_{d}.

Using (DD-REPR), problem (KS) becomes

$$\text{minimize} \quad \text{over } g \quad \|w-\mathscr{H}_T(w_{\mathrm{d}})g\|$$

The smoothed signal $\widehat{w}^*$ is then given by the pseudo-inverse:

$$\widehat{w}^* = \mathscr{H}_T(w_{\mathrm{d}})\big(\mathscr{H}_T(w_{\mathrm{d}})\big)^{\dagger} w, \tag{KS-SOL}$$

and involves only the given data w_{d} and w. More generally, problems involving a mixture of exact, missing, and noisy data can be solved using the data-driven representation [S4].

REFERENCES

[S1] J. C. Willems, P. Rapisarda, I. Markovsky, and B. De Moor, “A note on persistency of excitation,” *Systems & Control Lett.*, vol. 54, no. 4, pp. 325–329, 2005.

[S2] I. Markovsky, E. Prieto-Araujo, and F. Dörfler, “On the persistency of excitation,” *Automatica*, p. 110657, 2023.

[S3] I. Markovsky and F. Dörfler, “Identifiability in the behavioral setting,” *IEEE Trans. Automat. Contr.*, vol. 68, pp. 1667–1677, 2023.

[S4] ——, “Data-driven dynamic interpolation and approximation,” *Automatica*, vol. 135, p. 110008, 2022.

knowledge in solving practical problems. A major factor for the existence of the gap is a lack of experience in translating ill-posed “real-life” problems into well defined mathematical ones. This lack of experience is due to insufficient project oriented work where students are presented with open-ended problems for which they are encouraged to explore alternative solutions.

The project presented is motivated by the dynamic measurement problem in metrology. It involves data modeling, state-estimation, and Kalman filtering. It also requires practical skills, such as implementation and validation of an algorithm on simulated data in Matlab, and C programming for DSP implementation of the method. Nonstandard learning outcomes are: ability to identify and critically analyze assumptions, deterministic interpretation of Kalman smoother as a projection on the system’s behavior (KS), a direct data-driven method for Kalman smoothing (KS-SOL), and testing hypothesis in practice by performing experiments. Possible extensions of the project are: using multiple sensors (data fusion), dealing with unknown sensor dynamics and time-varying measured quantity (adaptive signal processing), derivation of confidence bounds (statistical analysis), and application of the method to different types of sensors. A long term goal is to develop other educational projects in the area of systems theory, signal processing, and control and collect them in a publicly available data-base.

ACKNOWLEDGMENTS

Alberto Padoan, Saverio Bolognani, Zhisen Jiang, and Anastasios Tsiamis read drafts of the paper and made suggestions for improvement.

The research leading to these results has received funding from: the Catalan Institution for Research and Advanced Studies (ICREA), the Fond for Scientific Research Vlaanderen (FWO) project G033822N, and the Spanish Ministry of Science Spanish State Research Agency MCIU/AEI/10.13039/501100011033, grant PID2023-148952OB-I00.

AUTHOR INFORMATION

Ivan Markovsky (imarkovsky@cimne.upc.edu) is an ICREA professor at the International Centre for Numerical Methods in Engineering, Barcelona. He received his Ph.D. degree in Electrical Engineering from the Katholieke Universiteit Leuven in February 2005. From 2006 to 2012 he was an Assistant Professor at the School of Electronics and Computer Science of the University of Southampton and from 2012 to 2022 an Associate Professor at the Vrije Universiteit Brussel. He is a recipient of an ERC starting grant "Structured low-rank approximation: Theory, algorithms, and applications" 2010–2015, Householder Prize honorable mention 2008, and research mandate by the Vrije Universiteit Brussel research council 2012–2022. His main research interests are computational methods for systems theory, identification, and data-driven control in the behavioral setting.

REFERENCES

[1] S. Papert, *Mindstorms: Children, Computers, And Powerful Ideas*. Basic Books, 1993.

[2] J. Parker, *R. L. Moore: Mathematician and Teacher*. Mathematical Association of America, 2005.

[3] F. Jones, “The Moore method,” *American Mathematical Monthly*, vol. 84, pp. 273–277, 1977.

[4] D. Zitarelli, "The origin and early impact of the Moore method," *American Mathematical Monthly*, vol. 111, pp. 465–486, 2004.

[5] J. A. Rossiter, C. G. Cassandras, J. Hespanha, S. Dormido, L. de la Torre, G. Ranade, A. Visioli, J. Hedengren, R. M. Murray, P. Antsaklis, F. Lamnabhi-Lagarrigue, and T. Parisini, "Control education for societal-scale challenges: A community roadmap," *Annual Reviews in Control*, vol. 55, pp. 1–17, 2023.

[6] J. A. Rossiter, J. Hedengren, and A. Serbezov, "Technical committee on control education: A first course in systems and control engineering [technical activities]," *IEEE Control Systems Magazine*, vol. 41, no. 1, pp. 20–23, 2021.

[7] D. Bernstein, "Enhancing undergraduate control education," *IEEE Control Systems Magazine*, vol. 19, pp. 40–43, 1999.

[8] E. Mazur, "Confessions of a converted lecturer," https://set.kuleuven.be/en/news/eric-mazur, 2019.

[9] P. Graham, "How to start Google," https://www.paulgraham.com/google.html, March 2024.

[10] I. Markovsky, "Dynamical systems and control mindstorms," in *Proc. of the 20th Mediterranean Conf. on Control and Automation*, Barcelona, 2012, pp. 54–59.

[11] G. Pólya, *How to Solve It*. Doubleday, 1957.

[12] I. Markovsky, "Comparison of adaptive and model-free methods for dynamic measurement," *IEEE Signal Proc. Lett.*, vol. 22, no. 8, pp. 1094–1097, 2015.

[13] ——, "An application of system identification in metrology," *Control Eng. Practice*, vol. 43, pp. 85–93, 2015.

[14] I. Markovsky, L. Huang, and F. Dörfler, "Data-driven control based on behavioral approach: From theory to applications in power systems," *IEEE Control Systems Magazine*, vol. 43, pp. 28–68, 2023.

[15] D. G. Luenberger, *Introduction to Dynamical Systems: Theory, Models and Applications*. John Wiley, 1979.

[16] J. C. Willems, "The behavioral approach to open and interconnected systems: Modeling by tearing, zooming, and linking," *IEEE Control Systems Magazine*, vol. 27, pp. 46–99, 2007.

[17] ——, "Paradigms and puzzles in the theory of dynamical systems," *IEEE Trans. Automat. Contr.*, vol. 36, pp. 259–294, 1991.

[18] B. Francis and W. M. Wonham, "The internal model principle of control theory," *Automatica*, vol. 12, no. 5, pp. 457–465, 1976.

[19] T. Söderström and P. Stoica, *System Identification*. Prentice Hall, 1989.

[20] L. Ljung, *System Identification: Theory for the User*. Prentice-Hall, 1999.

[21] J. Hansen, *LEGO Mindstorms NXT Power Programming: Robotics in C*. Variant Press, 2007.

[22] W. Shu, "Dynamic weighing under nonzero initial conditions," *IEEE Trans. Instrumentation Measurement*, vol. 42, no. 4, pp. 806–811, 1993.

[23] I. Markovsky and S. Van Huffel, "Overview of total least squares methods," *Signal Proc.*, vol. 87, pp. 2283–2302, 2007.