Slovak University of Technology in Bratislava
Faculty of Electrical Engineering and Information Technology

Advisory Systems to Support Decision Making in Medical Cybernetics

Dissertation Thesis Abstract
to obtain the academic title of "philosophiae doctor", abbreviated as PhD, in the doctorate degree study programme Robotics and Cybernetics in the field of study Cybernetics

In Bratislava June 20, 2023 Ing. Martin Dodek

Thesis has been prepared at:	Institute of Robotics and Cybernetics Faculty of Electrical Engineering and Information Technology Slovak University of Technology in Bratislava Ilkovičova 3 812 19 Bratislava 1
Submitter:	Ing. Martin Dodek Institute of Robotics and Cybernetics Faculty of Electrical Engineering and Information Technology Slovak University of Technology in Bratislava Ilkovičova 3 812 19 Bratislava 1
Supervisor:	doc. Ing. Eva Miklovičová, PhD. Institute of Robotics and Cybernetics Faculty of Electrical Engineering and Information Technology Slovak University of Technology in Bratislava Ilkovičova 3 812 19 Bratislava 1
Reader:	prof. Ing. Miroslav Fikar, DrSc. Institute of Information Engineering, Automation, and Mathematics Faculty of Chemical and Food Technology Slovak University of Technology in Bratislava Radlinského 9 812 37 Bratislava
Reader:	Ing. Ivana Budinská, PhD. Institute of Informatics Slovak Academy of Sciences Dúbravská cesta 9 845 07 Bratislava 45
Dissertation Thesis Abstract was sent:	

Dissertation Thesis Defence will be held on ..... at ..... at Institute of Robotics and Cybernetics Faculty of Electrical Engineering and Information Technology Slovak University of Technology in Bratislava Ilkovičova 3 812 19 Bratislava 1.

.....
prof. Ing. Vladimír Kutiš, PhD.
Dean of Faculty of Electrical Engineering and Information Technology

Abstract

Faculty of Electrical Engineering and Information Technology of Slovak Technical University in Bratislava
Programme: Robotics and Cybernetics
Study field: Cybernetics
Author: Ing. Martin Dodek
Final thesis topic: Advisory Systems to Support Decision Making in Medical Cybernetics
Final thesis supervisor: doc. Ing. Eva Miklovičová, PhD.
Month, Year: May, 2023

Biocybernetics is a relatively new multidisciplinary scientific field offering interesting options for further research of originally engineering approaches being applied to processes in living organisms. Therefore, this thesis deals with the research of new mathematical, statistical, and cybernetic methods applied to biosystems and, particularly, to the dynamics of glycemia in subjects with type 1 diabetes. The proposed new methods address partial subproblems such as empirical modeling and system identification, optimal state estimation, estimation and analysis of statistical models, and last but not least, model predictive control and optimal impulsive control of glycemia. The main purpose of such advisory system is the system-based decision support of a patient with diabetes with regard to the application of insulin therapy. The ultimate practical aim of the advisory system is to increase overall quality of life by optimizing traditional insulin therapy in the terms of smart bolus calculator design.

Keywords: biocybernetics, diabetes mellitus, model predictive control, insulin therapy, system identification, parameter estimation, optimization

Anotácia dizertačnej práce

Slovenská technická univerzita v Bratislave Fakulta elektrotechniky a informatiky
Program: robotika a kybernetika
Študijný odbor: kybernetika
Autor: Ing. Martin Dodek
Téma záverečnej práce: Poradné systémy v lekárskej kybernetike
Vedúci záverečnej práce: doc. Ing. Eva Miklovičová, PhD.
Mesiac, rok odovzdania: Máj, 2023

Biokybernetika je relatívne nová multidisciplinárna vedná oblasť, ktorá ponúka zaujímavé možnosti pre ďalší výskumu pôvodne výlučne inžinierskych prístupov aplikovaných na procesy prebiehajúce v živých organizmoch. Táto práca sa preto venuje výskumu nových matematických, štatistických a kybernetických metód aplikovaných na biosystémy a to konkrétne na dynamiku glykémie u pacientov s diabetom prvého typu. Tieto nové metódy potom riešia jednotlivé čiastkové problémy ako je empirické modelovanie a identifikácia parametrov, optimálny odhad stavu systému, odhad a analýza štatistických modelov a v neposlednom rade prediktívne a optimálne impulzné riadenie glykémie. Principiálne je hlavnou úlohou takéhoto poradného systému práve systémovo založená podpora pacienta s diabetom pri rozhodovaní v súvislosti s aplikáciou inzulínovej terapie. Konečným praktickým cieľom poradného systému je celkové zvýšenie kvality života pacientov optimalizáciou klasickej inzulínovej terapie v zmysle inteligentného bolus kalkulátora.

Kľúčové slová: biokybernetika, diabetes mellitus, prediktívne riadenie, inzulínová terapia, identifikácia systémov, odhad parametrov, optimalizácia

Contents

Introduction

Biocybernetics is a multidisciplinary field that offers wide options for applicability and new opportunities for further research of originally engineering approaches such as mathematical modeling and control systems design on the processes happening in living organisms, i.e. biosystems. This thesis is dedicated to the problem of the metabolic disorder called diabetes mellitus, in particular its treatment, modeling, and finally automatic control of glycemia in the context of the so-called artificial pancreas.

Basically, the main purpose of an advisory system is system-oriented support to decision-making in regard to the application of insulin therapy in subjects with diabetes. This topic involves the design of new efficient strategies to execute insulin treatment based on systems identification, model predictive control, and optimization methods. The ultimate purpose of this advisory system is to improve the quality of life of a diabetic subject and to enhance the efficiency of insulin dosing management in terms of improving the chosen qualitative clinical metrics. However, this thesis also deals with many related partial sub-problems including system identification, optimal state estimation, stochastic modeling and estimation.

Thesis goals

Medical cybernetics is a field in which, among other things, systems theory and advanced automatic control methods are applied to the technical support of medical research and practice. An example would be the design of automated advisory systems that evaluate data to facilitate decision making in a given domain. The aim of the thesis is to design algorithms for a complex decision support system for the application of insulin therapy in patients with type 1 diabetes mellitus. Partial problems include empirical modelling of glycemia dynamics, identification of model parameters in compliance with physiology, design of model predictive control of glycemia and design of optimal bolus calculator algorithm.

Thesis goals:

1. To analyze the state of the art in the area of modeling and control of type 1 diabetes mellitus.
2. To propose algorithms for identifying diabetes model parameters that ensure consistency with basic physiology.
3. To propose new approaches to improve the performance and safety of predictive control of glycemia within the framework of artificial pancreas implementation.
4. To optimize conventional insulin therapy and design a smart model-based bolus calculator.
5. To validate the proposed algorithms and methods by the means of numerical simulations and to evaluate the achieved results.

Chapter 1

Diabetes and its Treatment

Glucose is an essential substance that provides the required energy for cells and tissues in the human body, and hence it is vital to maintain its blood concentration, i.e. glycemia, in a relatively narrow physiologically acceptable interval. In a healthy subject, glycemia is autonomously governed by a complex neurohormonal regulatory system where the pancreas plays the most significant role.

It is the hormone insulin excreted by pancreatic beta cells, which is the primary mediator maintaining glucose homoeostasis in the organism. The specific mechanism of insulin action allows some types of cells to absorb and consume glucose from the blood plasma, while insulin also has the ability to inhibit endogenous glucose production.

In addition to insulin, other hormones also participate in the glycemia regulatory system. From the class of insulin-antagonist, i.e. counterregulatory hormones, we can mention glucagon, which stimulates glucose production in the liver, also called endogenous glucose production. This response of the human body can ultimately be seen as a natural mechanism to prevent hypoglycemia, i.e. the state when the glycemia level is below a certain physiologically unacceptable and safe threshold. It is worth noting that the primary source of glucose inflow is related to food-induced carbohydrate intake and digestion.

Correctly and effectively working insulin-glucose regulatory system responses to sudden (usually meal-related) glycemia increase by insulin release according to pancreatic beta cell function while increasing glucose consumption by cells and tissues that are sensitive to insulin and by suppressing endogenous glucose production. This sequence ultimately results in a lowering of glycemia and its return to normal. On the contrary, in the case of hypoglycemia state, insulin excretion is virtually stopped while there is present glucagon release that stimulates glucose production in the liver.

Diabetes mellitus is a relatively common metabolic disorder characterized by a persistently and chronicly elevated glycemia level above normal values, also called hyperglycemia. In the long term, this state leads to a wide variety of symptoms and health complications for the patient.

Based on the underlying physiological mechanisms and phenomena, we distinguish two major types of diabetes. Type 1 diabetes can be characterized as absolute insulin deficiency. In that case, the damaged pancreas does not produce a sufficient amount of insulin or does not produce insulin at all. This pancreatic damage is permanent and irreversible, so patients face lifelong therapy in the form of synthetic insulin administration. In fact, this insulin therapy is vital to maintain glycemia within the clinically acceptable interval and therefore to slow the progression of the health complications.

1.1 Problems of Managing Diabetes and Controlling Glycemia

Maintaining glycemia within the physiologically acceptable interval is in fact a complex problem belonging primarily to the field of system and control theory. From a cybernetic point of view, the controlled variable is the blood glucose concentration, and the manipulated variable is the insulin administration rate. However, the controlled system is subject to various disturbances, while meal-related carbohydrate intake can be considered the most significant, which can actually be measured. Other disturbances include physical activity, or psychical/mental load and stress, yet these are hard to quantify and can be rather considered random affecting factors.

Conventional insulin therapy can be seen to be a very simple form of control. The disturbance represented by the carbohydrate intake is usually compensated by the corresponding insulin administration, which represents the feedforward component. Eventual fluctuations of glycemia determined by sparse finger-stick measurements (the so-called self monitoring blood glucose) are also considered when determining the insulin dose, what can be seen as a simple feedback control.

The invention of new minimally invasive subcutaneous glucose sensors provided opportunities for a widespread clinical application of continuous glucose monitoring devices, which offer densely sampled real-time data compared to finger-stick measurements. By an eventual fusions of available technologies such as the

insulin pump and the subcutaneous continuous glucose monitoring device together with an appropriate control algorithm, a rather complex device called the artificial pancreas would be augmented.

The most significant engineering problems and challenges related to the development of artificial pancreas can be stated as follows:

- lag character of insulin action dynamics
- non-negative nature of insulin administration
- non-linearity of insulin-glucose interaction
- time-varying physiological parameters
- effects of unmeasurable random disturbances
- high demands on the control performance

1.2 Conventional Insulin Therapy and Bolus Calculator

In clinical practice, a standardized approach was established to determine the suboptimal size of the insulin bolus, called the bolus calculator. Regarding the static response of glycemia to input excitation, a diabetic subject can be characterized with respect to both insulin administration and carbohydrate intake. The first of these characteristics is the insulin sensitivity IS [mmol/l/U] quantifying the glycemia lowering effect of the administered unit of insulin, whereas the carbohydrate sensitivity CS [mmol/l/g] describes the effect of unit carbohydrate intake on increasing glycemia. Then, the helper insulin-carbohydrate parameter ICR [g/U] is defined as

$$ICR = \frac{IS}{CS} \ , \tag{1.1}$$

and quantifies the ability of insulin to compensate for the carbohydrate intake.

The insulin bolus is then determined as [1]

$$B = \frac{G - G_w}{IS} + \frac{CHO}{ICR} \ , \tag{1.2}$$

where the insulin bolus B [U] is expressed using a dedicated insulin unit U, G [mmol/l] is the currently measured glycemia, G_w [mmol/l] is its target value, and CHO [g] is the carbohydrate content of the meal. Insulin bolus can be decomposed into the component compensating carbohydrate intake and the component correcting the deviation of glycemia from the target value.

Concerning the timing of insulin bolus administration, it is more appropriate and broadly recommended to apply the injection with a certain time advance before the corresponding carbohydrate intake [2].

In the context of this thesis, the bolus calculator represents the simplest and most widely used advisory algorithm to support decision making in the insulin therapy application.

Chapter 2

Physiology-Compliant Empirical Model for Glycemia Prediction

2.1 Introduction

This chapter is focused primarily on the problem of maximizing the accuracy of glycemia prediction in type 1 diabetic subjects, particularly if linear empirical models are used. The two-input Box-Jenkins model has been preferred since its structure reflects the actual differences in the dynamics of insulin administration and carbohydrate intake input. Furthermore, the basal state of the subject was integrated directly into the model structure. Since it is common to experience issues with the physiology compliance of empirical models, it was proposed to perform the multi-step-ahead predictive identification of the zero-pole-gain representation with applied constraints.

The main motivation for performing highly accurate predictions is to forecast severe states of hyper and hypo glycemia as they are the cardinal risks associated with diabetes and its treatment [3]. From the cybernetic point of view, a diabetic subject can be seen as a dynamic system with two measurable inputs and a single output. The inputs quantify the rate of insulin administration and the rate of carbohydrate intake, whereas the output of this system-based abstraction can be interpreted as the blood glucose concentration.

However, real diabetic patients are also subject to many unmeasurable or ambiguously affecting disturbances, such as the effect of exercise [4] or even stress factors. Additionally, if oversimplifying modeling of the actual complex glycemia dynamics, a significant plant-model mismatch may appear.

2.2 Proposed Model Structure

The following notation of signals is used:

- control input $u(t)$ [U/min] - alias the rate of insulin administration
- disturbance input $d(t)$ [g/min] - alias the rate of carbohydrate intake
- output $y(t)$ [mmol/l] alias the blood glucose concentration
- unmeasurable random disturbances $\epsilon(t)$ with the character of zero-mean white noise

The model dynamic is defined as follows:

$$y_{(k)} = \frac{B^u(z)}{A^u(z)}u_{(k)} + \frac{B^d(z)}{A^d(z)}d_{(k)} + \frac{C(z)}{D(z)}\epsilon_{(k)} + y_0 \tag{2.1}$$

2.3 Parameters Identification

Individualization of the T1DM model should ideally be performed using exclusively passively acquired clinical data from free-living conditions, i.e. CGM measurements and diabetic diary logs [5].

Regular bolus calculator algorithms are based on relatively simple rules [1] that utilize some clinical parameters of the subject, particularly the insulin and the carbohydrate sensitivity (see section 1.2). Assuming the insulin-carbohydrate ratio *ICR* parameter as fixed between the bolus calculations causes both inputs to be linearly dependent, which consequently results into poor excitation quality of the input signals.

Typical diabetic signals pose another specific challenge concerning the model identification. The input signals, i.e., the insulin administration and the carbohydrate intake do not prove to have persistent excitation properties, or in other words: the inputs are mostly inactive during the conventional insulin therapy. Therefore,

a novel approach based on preferential weighting of the postprandial samples is proposed. If the corresponding weights of the postprandial samples of the identification dataset are increased, the effect of the input part of the model can be emphasized this way.

The traditional nonpredictive identification techniques are based on the least squares minimization of the single-step-ahead prediction error. In order to improve the prediction performance, an alternative approach is proposed. This improvement consists of designing the cost function that involves the multi-step-ahead prediction error for each sample of the dataset.

The cost function of the multi-step-ahead predictive identification technique can be written as:

$$J(\hat{\theta}) = \sum_{k=1}^{N} \lambda_k^n \left\{ \sum_{i=0}^{n_e} \lambda_i^e \left[y_{(k+i)} - \hat{y}_{(k+i|k)} \right]^2 \right\} \tag{2.2}$$

where n_e is the prediction horizon, λ_k^n is the weight with respect to the sample number and λ_i^e is the weight with respect to the prediction horizon.

2.3.1 Identification in the Space of Zeros, Poles and Gains

Each empirical model of T1DM must meet a set of specific requirements to be considered physiologically-compliant. In detail, the particular properties demanded of the empirical model of T1DM are the following.

1. negative static gain of the insulin submodel corresponding to the glycemia-lowering effect of the insulin administration
2. positive static gain of the carbohydrate submodel corresponding to the glycemia-increasing effect of the carbohydrate intake
3. positive and reasonable basal glycemia
4. stable poles of all submodels
5. minimum-phase character of both submodels
6. aperiodic transient response of both sub-models

The proposed solution is based on performing the parameter estimation in a different but equivalent space of zeros, poles, and gains of the empirical model (2.1). It is also quite convenient that the cost function (2.2) remains virtually untouched by this modification. The new parameter vector can be divided into sections.
Vector of poles:

$$\theta_\rho = \begin{bmatrix} \rho_1^u \cdots & \rho_{n_{A^u}}^u & \rho_1^d \cdots & \rho_{n_{A^d}}^d & \rho_1^\epsilon \cdots & \rho_{n_D}^\epsilon \end{bmatrix}^{\mathrm{T}} \tag{2.3}$$

Vector of zeros:

$$\theta_\zeta = \begin{bmatrix} \zeta_1^u \cdots & \zeta_{n_{B^u}-1}^u & \zeta_1^d \cdots & \zeta_{n_{B^d}-1}^d & \zeta_1^\epsilon \cdots & \zeta_{n_C}^\epsilon \end{bmatrix}^{\mathrm{T}} \tag{2.4}$$

Vector of gains:

$$\theta_\gamma = \begin{bmatrix} \gamma^u & \gamma^d \end{bmatrix}^{\mathrm{T}} \tag{2.5}$$

Finally, the full parameter vector:

$$\theta_{\rho\zeta\gamma} = \begin{bmatrix} \theta_\rho^{\mathrm{T}} & \theta_\zeta^{\mathrm{T}} & \theta_\gamma^{\mathrm{T}} & y_0 \end{bmatrix}^{\mathrm{T}} \tag{2.6}$$

The estimates of the sensitivities parameters are now directly related to the identified static gains:

$$\hat{IS} = -\frac{\gamma^u}{T_s} \tag{2.7}$$

$$\hat{CS} = \frac{\gamma^d}{T_s} \tag{2.8}$$

One can realize that the demand for a model with pure aperiodic transient response is implicitly satisfied since the poles (2.3) are estimated as the real ones, i.e they all have zero imaginary part. Other physiology-based criteria can be implemented as simple constraints of the parameter vector (2.6).

In order to meet the stability criterion, all poles have to lie within the unit circle.

$$\rho_i^u \in \langle -0.99, 0.99 \rangle \qquad \rho_i^d \in \langle -0.99, 0.99 \rangle \tag{2.9}$$

The demand for a minimum phase can be applied for both submodels:

$$\zeta_i^u \in \langle -1, 1 \rangle \qquad \zeta_i^d \in \langle -1, 1 \rangle \tag{2.10}$$

Concerning the restrictions of the static gains, the physiology of insulin administration and carbohydrate intake can be represented by the following bounds:

$$\gamma^u \in \langle -500, -50 \rangle \quad \gamma^d \in \langle 10, 100 \rangle \tag{2.11}$$

The output constant term y_0 should theoretically correspond to the glycemia when no basal insulin is being administrated, so its bounds have to be set empirically:

$$y_0 \in \langle 7, 12 \rangle \tag{2.12}$$

2.4 Prediction

Applying the matrix calculation method one can write the following set of equations:

$$\hat{x}^u_f = M^y_f(A^u)^{-1}\left(-M^y_p(A^u)\hat{x}^u_p + M^u_f(B^u)u_f + M^u_p(B^u)u_p\right) \tag{2.13}$$

$$\hat{x}^d_f = M^y_f(A^d)^{-1}\left(-M^y_p(A^d)\hat{x}^d_p + M^u_f(B^d)d_f + M^u_p(B^d)d_p\right) \tag{2.14}$$

$$\hat{x}^\epsilon_f = M^y_f(D)^{-1}\left(-M^y_p(D)\hat{x}^\epsilon_p + \left(M^u_f(C)\begin{bmatrix} I & 0 \end{bmatrix} + \begin{bmatrix} 0 & I \end{bmatrix}\right)\hat{\epsilon}_f + M^u_p(C)\hat{\epsilon}_p\right) \tag{2.15}$$

However, signal ϵ cannot be directly measured, but it is possible to use its estimate based on the error of single-step-ahead prediction:

$$\hat{\epsilon}_{(k|k)} = y_{(k)} - \hat{y}_{(k|k-1)} \ . \tag{2.16}$$

Concerning the prediction, the future values of the unmeasurable disturbance will be considered unchanged.

$$\hat{\epsilon}_f = \hat{\epsilon}_{(k|k)}\begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^{\mathrm{T}} \tag{2.17}$$

The single-step-ahead prediction of the noise sub-model $\hat{x}^\epsilon_{(k|k)}$ has to be corrected by the current input noise term estimate $\hat{\epsilon}_{(k|k)}$, hereby the measurement-corrected prediction is made:

$$\hat{x}^\epsilon_{(k|k)} = {h^\epsilon_{(k)}}^{\mathrm{T}}\theta^\epsilon + \hat{\epsilon}_{(k|k)} \tag{2.18}$$

Finally, the output prediction $\hat{y}_f$ is calculated as the sum of the partial predictions (2.13),(2.14),(2.15):

$$\hat{y}_f = \hat{x}^u_f + \hat{x}^d_f + \hat{x}^\epsilon_f + \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^{\mathrm{T}} y_0 \tag{2.19}$$

2.5 In-silico Experiment

To validate the presented improvements, a comprehensive in-silico experiment has been carried out.

The comprehensive nonlinear simulation model published in [6, 7, 8, 9] has been chosen. The basal glycemia of the subject was chosen as $G_b = 7$ mmol/l together with the basal insulin administration rate $v_b = 0.01$ U/min. Virtual diabetic diary data comprising multiple meal events together with the corresponding insulin boluses have been generated. In this way, the emulation of a typical diabetic treatment routine was made over a period of two days (48 hours). The corresponding meal-compensating insulin boluses were calculated by the traditional bolus calculator (1.2), however, variable insulin-carbohydrate ratio was applied. The glycemia readings were samples with the sample time $T_s = 10$ min and distorted by the additive measurement noise.

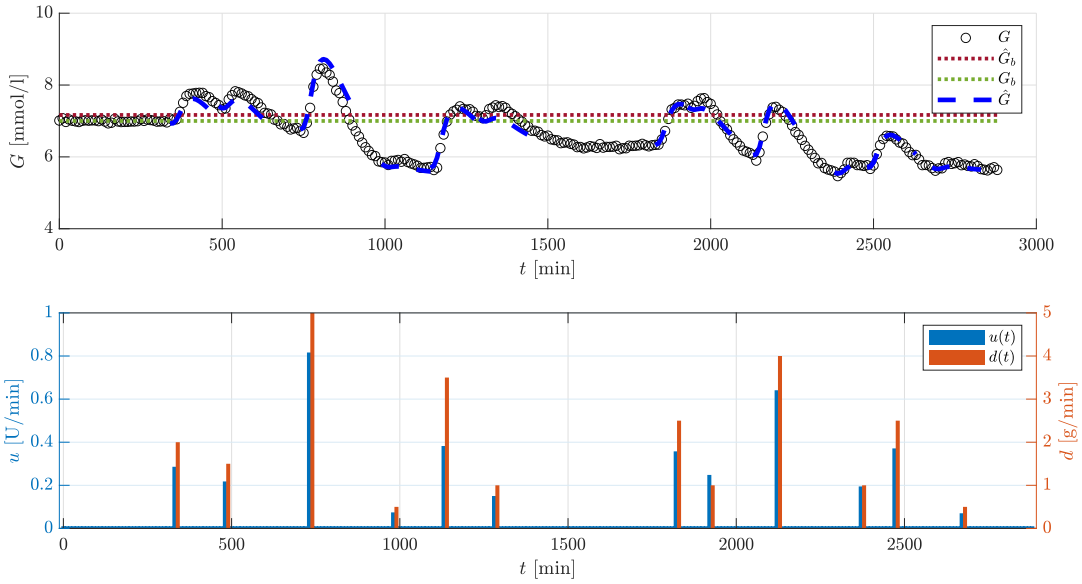
The starting points of the predictions were chosen to represent the postprandial periods while assuming the prediction horizon $n_e = 15$. For the proposed model structure (2.1) the polynomial degrees were chosen as $n_{A^u} = 5 \quad n_{A^d} = 5 \quad n_{B^u} = 5 \quad n_{B^d} = 5 \quad n_C = 2 \quad n_D = 2$. The weights λ^e were chosen as linearly decreasing $\lambda^e_i = 2 - \frac{i}{15}$.

The generated virtual diabetic dataset used for the identification is depicted in Figure 2.1a together with the postprandial prediction. However, the prediction performance for the validation dataset, which is shown in Figure 2.1b, is actually the important one.

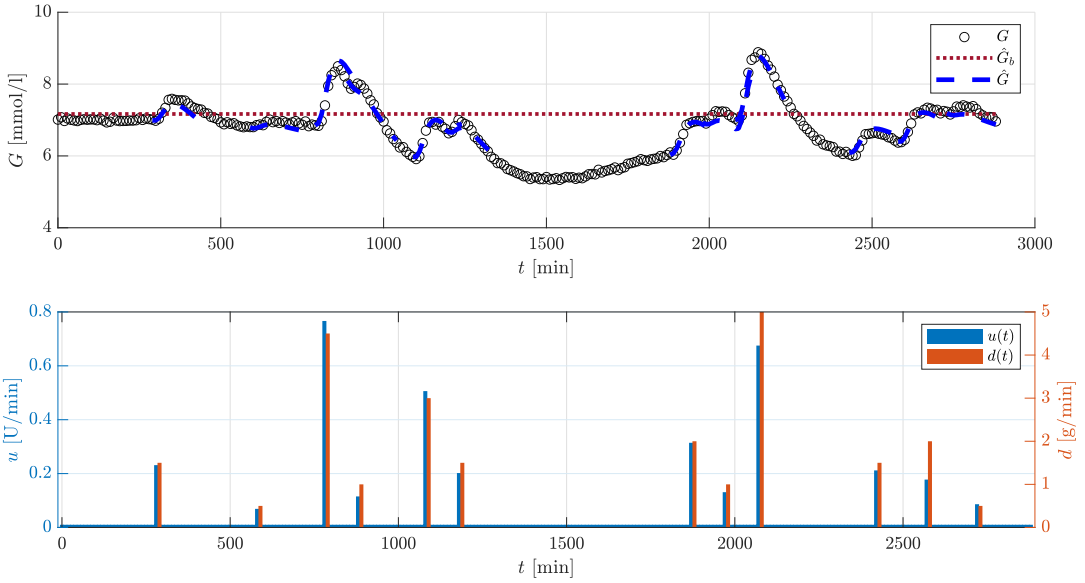
For a quick analysis of the identified model dynamics, the impulse responses are plotted in Figure 2.2.

2.6 Conclusions

The proposed linear empirical model with separate feedback dynamics for insulin administration and carbohydrate intake input turned out to be an appropriate structure for the prediction of blood glucose concentration in type 1 diabetic subjects. The multi-step-ahead predictive identification was proven to be a suitable approach for estimating the parameters of prediction-oriented models. A notable contribution is the parameter identification in the space of model zeros, poles, and gains with the consequent neat implementation of the physiology-based constraints. Moreover, to cope with the poor excitation caused by impulse character of the input signals, the weighting of samples related to the postprandial period was applied.

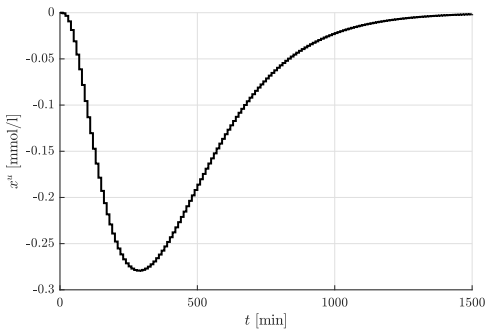


(a) Identification dataset

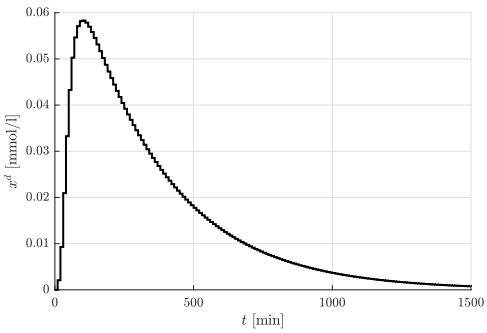


(b) Validation dataset

Figure 2.1: Prediction of glycemia



(a) Insulin administration sub-model



(b) Carbohydrate intake sub-model

Figure 2.2: Impulse responses of the identified model

Chapter 3

Maximizing Performance of Linear Model Predictive Control of Glycemia for T1DM Subjects

3.1 Introduction

The primary objective of this chapter is the custom design of an effective, yet relatively easy-to-implement, predictive control algorithm to maintain normoglycemia in patients with type 1 diabetes. In the introduced linear model predictive control, the constraints were applied to the manipulated variable in order to reflect the technical limitations of insulin pumps and the typical nonnegative nature of insulin administration. Similarly, inequalities constraints for the controlled variable were also assumed while anticipating suppression of hypoglycemia states during the automated insulin treatment. However, the problem of control infeasibility has emerged, especially if one uses too tight constraints of the manipulated and the controlled variable concurrently.

In this chapter, the two-input transfer function-based discrete-time empirical model (2.1) is adopted from chapter 2 for the MPC synthesis.

3.2 Predictive Equations

The predictive equations for the model (2.1) have to be briefly introduced before the control algorithm itself. The output prediction can be decomposed into the free and the forced response as:

$$\hat{y}_f = \hat{y}_f^{\text{free}} + \hat{y}_f^{\text{forc}} \tag{3.1}$$

where $\hat{y}_f$ $[n_e \times 1]$ gets:

$$\hat{y}_f = \begin{bmatrix} \hat{y}_{(k+1)} & \hat{y}_{(k+2)} & \cdots & \hat{y}_{(k+n_e-1)} & \hat{y}_{(k+n_e)} \end{bmatrix}^{\mathrm{T}} \tag{3.2}$$

The forced response represents the effect of future control changes Δu_f, while the vector $\hat{y}_f^{\text{forc}}$ $[n_e \times 1]$ gets the following linear form:

$$\hat{y}_f^{\text{forc}} = H_f \Delta u_f \tag{3.3}$$

where H_f $[n_e \times n_u]$ is the step-response matrix. The prediction horizon representing the length of vector $\hat{y}_f$ is denoted n_e.
The vector of future control changes Δu_f $[n_u \times 1]$ is defined as:

$$\Delta u_f = \begin{bmatrix} \Delta u_{(k)} & \Delta u_{(k+1)} & \cdots & \Delta u_{(k+n_u-1)} \end{bmatrix}^{\mathrm{T}} \tag{3.4}$$

where n_u is the control horizon representing the number of changes of the manipulated variable.

The following matrix equation can be formed for the prediction of control submodel:

$$\hat{x}_f^u = M_f^y (A^u)^{-1} \left(-M_p^y (A^u) \hat{x}_p^u + M_f^u (B^u) u_f + M_p^u (B^u) u_p \right) \tag{3.5}$$

Predictive equation (3.5) can be modified to analogously represent the predictions of the remaining terms of the model (2.1).
The forced response is equal to:

$$\hat{y}_f^{\text{forc}} = H_f \Delta u_f = M_f^y (A^u)^{-1} M_f^u (B^u) M_\Sigma \Delta u_f \ , \tag{3.6}$$

and the free response gets the following form:

$$\begin{aligned}\hat{y}_f^{\text{free}} &= M_f^y(A^u)^{-1}\left(-M_p^y(A^u)\hat{x}_p^u + M_f^u(B^u)\begin{bmatrix}1 & 1 & \dots & 1\end{bmatrix}^{\mathrm{T}} u_{(k)} + M_p^u(B^u)u_p\right)\\ &+ M_f^y(A^d)^{-1}\left(-M_p^y(A^d)\hat{x}_p^d + M_f^u(B^d)d_f + M_p^u(B^d)d_p\right)\\ &+ M_f^y(D)^{-1}\left(-M_p^y(D)\hat{x}_p^\epsilon + \left(M_f^u(C)\begin{bmatrix}I & \mathbf{0}\end{bmatrix} + \begin{bmatrix}\mathbf{0} & I\end{bmatrix}\right)\hat{\epsilon}_f + M_p^u(C)\hat{\epsilon}_p\right)\end{aligned} \tag{3.7}$$

whence I $[n_e \times n_e]$ is the unit matrix, $\mathbf{0}$ $[n_e \times 1]$ is the zero vector and matrix M_Σ $[n_e \times n_u]$ is defined as lower triangular:

$$M_\Sigma = \begin{pmatrix}1 & 0 & \cdots & 0\\ 1 & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 1 & 1 & \cdots & 1\\ \vdots & \vdots & \ddots & \vdots\\ 1 & 1 & \cdots & 1\end{pmatrix} \tag{3.8}$$

The output prediction $\hat{y}_f$ is finally determined as (2.19).

3.3 Control Algorithm

For the incremental control law, the decision vector of the future control changes Δu_f is related to the vector of future manipulated variable u_f using matrix M_Σ (3.8) as:

$$u_f = \begin{bmatrix}1 & 1 & \dots & 1\end{bmatrix}^{\mathrm{T}} u_{(k-1)} + M_\Sigma \Delta u_f \tag{3.9}$$

The general cost function of the predictive control can be written as [10]:

$$J(\Delta u_f) = \sum_{i=1}^{n_e} \lambda_i^y \left[\bar{y}_{(k+i)} - \hat{y}_{(k+i)}\right]^2 + \sum_{j=1}^{n_u} \lambda_j^u \Delta u_{(k+j-1)}^2 \tag{3.10}$$

The weighting vector λ^u $[n_u \times 1]$ for penalizing the squared changes of the manipulated variable can be interpreted as the factor affecting the aggressiveness of control:

$$\lambda^u = \begin{bmatrix}\lambda_1^u & \lambda_2^u & \dots & \lambda_{n_u-1}^u & \lambda_{n_u}^u\end{bmatrix}^{\mathrm{T}} \tag{3.11}$$

Vector λ^y $[n_e \times 1]$ is the counter-weighting vector for the control error penalty:

$$\lambda^y = \begin{bmatrix}\mathbf{0} & \lambda_{n_e^\star}^y & \lambda_{n_e^\star+1}^y & \dots & \lambda_{n_e-1}^y & \lambda_{n_e}^y\end{bmatrix}^{\mathrm{T}} , \tag{3.12}$$

where $\mathbf{0}$ $[1 \times n_e^\star]$ is the zero vector and parameters $n_e^\star$ and n_e represent the beginning and the end of the optimized prediction horizon, respectively.

The elements of reference vector $\bar{y}$ $[n_e \times 1]$ can be assumed equal to a constant G_t representing the target glycemia.

$$\bar{y} = \begin{bmatrix}\bar{y}_{(k+1)} & \bar{y}_{(k+2)} & \dots & \bar{y}_{(k+n_e-1)} & \bar{y}_{(k+n_e)}\end{bmatrix}^{\mathrm{T}} \tag{3.13}$$

The target glycemia G_t can be chosen according to the physician's recommendation, from relatively narrow interval $5.0 < G_t < 6.0$ mmol/l.

Cost function (3.10) can be reshaped into the equivalent quadratic form:

$$J(\Delta u_f) = \left(\bar{y} - \hat{y}_f\right)^{\mathrm{T}} \Lambda^y \left(\bar{y} - \hat{y}_f\right) + \Delta u_f^{\mathrm{T}} \Lambda^u \Delta u_f , \tag{3.14}$$

where Λ^u $[n_u \times n_u]$ and $\Lambda^y [n_e \times n_e]$ are positive definite diagonal matrices with diagonal vectors λ^u (3.11) and λ^y (3.12), respectively [11].

$$\Lambda^u = \mathrm{diag}(\lambda^u) \tag{3.15}$$

$$\Lambda^y = \mathrm{diag}(\lambda^y) \tag{3.16}$$

Substituting the output prediction (3.1) and the forced response (3.3) into cost function (3.14), the quadratic form with respect to the decision variable Δu_f gets [12]:

$$J(\Delta u_f) = \Delta u_f^{\mathrm{T}} \mathbf{A} \Delta u_f + 2\mathbf{b}^{\mathrm{T}} \Delta u_f + c , \tag{3.17}$$

where matrix $\mathbf{A}$ $[n_u \times n_u]$, vector $\mathbf{b}$ $[n_u \times 1]$, and scalar c are defined as:

$$\mathbf{A} = H_f^{\mathrm{T}} \Lambda_y H_f + \Lambda_u \tag{3.18a}$$

$$\mathbf{b}^{\mathrm{T}} = -\left(\bar{y} - \hat{y}_{\mathrm{free}}\right)^{\mathrm{T}} \Lambda_y H_f \tag{3.18b}$$

$$c = \left(\bar{y} - \hat{y}_{\mathrm{free}}\right)^{\mathrm{T}} \Lambda_y \left(\bar{y} - \hat{y}_{\mathrm{free}}\right) \tag{3.18c}$$

Using the receding-horizon strategy, only the first element of solution Δu_f is actually applied:

$$u_{(k)} = u_{(k-1)} + \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix} \Delta u_f \tag{3.19}$$

3.3.1 Constraining the Manipulated Variable

The greatest weakness of the traditional concept of artificial pancreas is the non-negative nature of insulin administration. In addition to the lower bound, insulin pumps also have technological limits for the maximal insulin delivery rate, so the manipulated variable lies within the interval:

$$u_{\min} \leq u \leq u_{\max} \tag{3.20}$$

The minimal infusion rate $u_{\min}$ is zero while typical $u_{\max}$ was reported in [13] or [14]:

$$u_{\min} = 0 \text{ U/min} \quad u_{\max} = \frac{1}{15} \text{ U/min} \tag{3.21}$$

It is desired to express the manipulated value constraints (3.20) as an equivalent system of linear inequalities with respect to the decision vector Δu_f using matrix M_Σ (3.8) as [11]:

$$-M_\Sigma \Delta u_f \leq -\begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^{\mathrm{T}} \left(u_{\min} - u_{(k-1)}\right) \tag{3.22}$$

$$+M_\Sigma \Delta u_f \leq +\begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^{\mathrm{T}} \left(u_{\max} - u_{(k-1)}\right) \tag{3.23}$$

3.3.2 Constraining the Controlled Variable

Constraining the controlled variable is especially fundamental in the case of artificial pancreas, since it could be the clue to reduce the risk of hypoglycemia during the automated insulin treatment. Formally, the controlled variable is supposed to be within the interval:

$$y_{\min} \leq y \leq y_{\max} \tag{3.24}$$

Relatively safe interval of glycemia is roughly:

$$y_{\min} = 4.5 \text{ mmol/l} \quad y_{\max} = 9 \text{ mmol/l} \tag{3.25}$$

In order to involve the constraints (3.24) in the optimization problem, the corresponding linear inequalities system with respect to the decision vector Δu_f has to be derived. Based on the output prediction decomposition (3.1) and the linear form of the forced response (3.3), one can write [11]:

$$-H_f \Delta u_f \leq -\begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^{\mathrm{T}} y_{\min} + \hat{y}^{\mathrm{free}} \tag{3.26}$$

$$+H_f \Delta u_f \leq +\begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^{\mathrm{T}} y_{\max} - \hat{y}^{\mathrm{free}} \tag{3.27}$$

3.3.3 Control Feasibility

Under some specific circumstances, particularly if applying too strict constraints concurrently, it may happen that there is no feasible solution for all the assumed inequalities (3.22), (3.23), (3.26), (3.27). In fact, the constraints of the manipulated variable basically cannot be violated since these are physically grounded, so it is the controlled variable, the constraints of which have to be corrupted after all.

To this end, the proposed strategy is to detect the infeasibility during the control, generate the corresponding alarm, and adapt the constraints of the controlled variable (3.26),(3.27) in order to recover the feasibility and optimality of the control. For a rigorous analysis of this problem, one can harness the Farkas lemma [15]. The following linear programming problem has to be solved, and the found minimum checked for meeting the feasibility condition defined by the Farkas lemma.

$$\begin{gathered} \min b^{\mathrm{T}} y \\ \text{subj. to: } A^{\mathrm{T}} y = \mathbf{0} \text{ and } y \geq \mathbf{0} \end{gathered} \tag{3.28}$$

The solution of the joint linear inequalities system formed by (3.22), (3.23), (3.26), (3.27) is feasible if the corresponding optimization problem (3.28) has a non-negative optimum i.e. if $\min b^{\mathrm{T}} y \geq 0$.

If there is no feasible solution, then the controlled variable constraints (3.24) have to be adapted according to Algorithm 1. This algorithm allows to choose either the upper or the lower bound of the controlled variable to be preserved via tuning the parameters β and γ. In the case of glycemia control, due to the serious consequences of hypoglycemia, the lower bound is much more important not to be violated, so we can choose $\beta = 0.1$ and $\gamma = 0.01$ to reflect this demand. Hence, restoring the feasibility and correcting the constraints can be considered non-symmetrical.

Parameters: β, γ, $u_{\max}$, $u_{\min}$
Data: $\hat{y}^{\text{free}}$, H_f, $u_{(k-1)}$
Result: $y_{\max}$, $y_{\min}$
begin
  $A = \begin{bmatrix} -M_\Sigma{}^{\mathrm{T}} & M_\Sigma{}^{\mathrm{T}} & -H_f{}^{\mathrm{T}} & H_f{}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$;
  repeat
    $$b \leftarrow \begin{bmatrix} -\begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^{\mathrm{T}} \left(u_{\min} - u_{(k-1)}\right) \\ +\begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^{\mathrm{T}} \left(u_{\max} - u_{(k-1)}\right) \\ -\begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^{\mathrm{T}} y_{\min} + \hat{y}^{\text{free}} \\ +\begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^{\mathrm{T}} y_{\max} - \hat{y}^{\text{free}} \end{bmatrix} ;$$
    $J \leftarrow \min b^{\mathrm{T}} y$ subj. to $A^{\mathrm{T}} y = \mathbf{0}$, $y \geq \mathbf{0}$;
    if $J < 0$ **then**
      $y_{\max} \leftarrow y_{\max}\,(1 + \beta)$;
      $y_{\min} \leftarrow y_{\min}\,(1 - \gamma)$;
    end
  until $J \geq 0$;
end

Algorithm 1: Controlled variable constraints adaptation

3.4 In-silico Experiment

The simulation model details were mentioned in chapter 2. The basal state was adjusted according to the basal glycemia $G_b = 7$ mmol/l and the corresponding basal insulin delivery rate $v_b = 0.01$ U/min. The virtual CGM measurements were distorted by the additive white noise with the standard deviation $\sigma = 0.1$ mmol/l. The experiment was designed to emulate the regular behavior of a subject with type 1 diabetes during the two-day period.

Applying the constraints to the manipulated variable in the terms of (3.22), (3.23) with bounds (3.21) yields the results that can be seen in Figure 3.1.

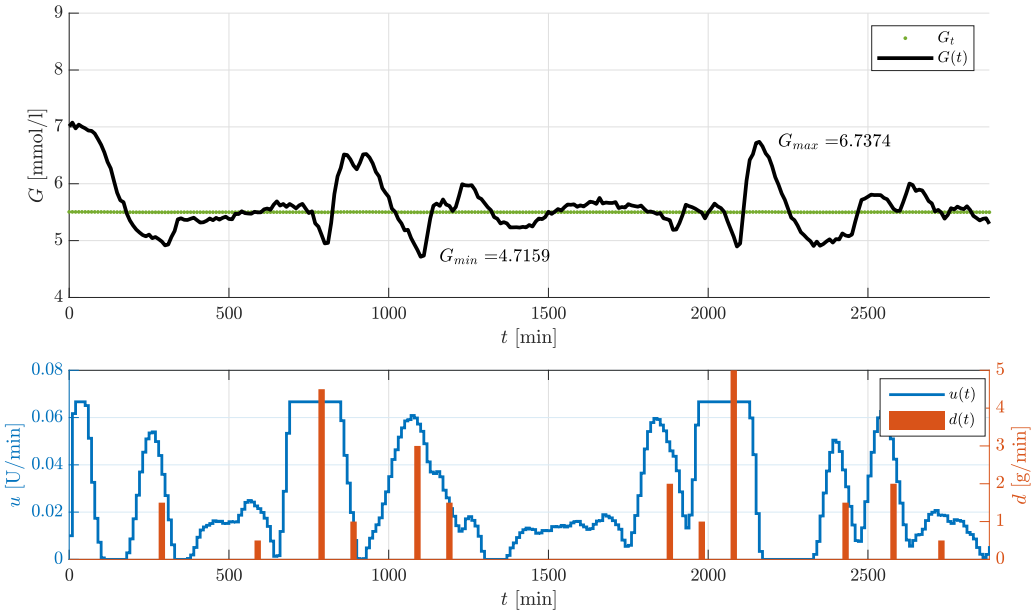


Figure 3.1: Predictive control in-silico experiment with constrained manipulated variable - meal scenario 1

Now the constraints are also applied to the controlled variable according to equations (3.26), (3.27) with bounds (3.25). In Figure 3.2 one can notice improved hypoglycemia management although it was apparently

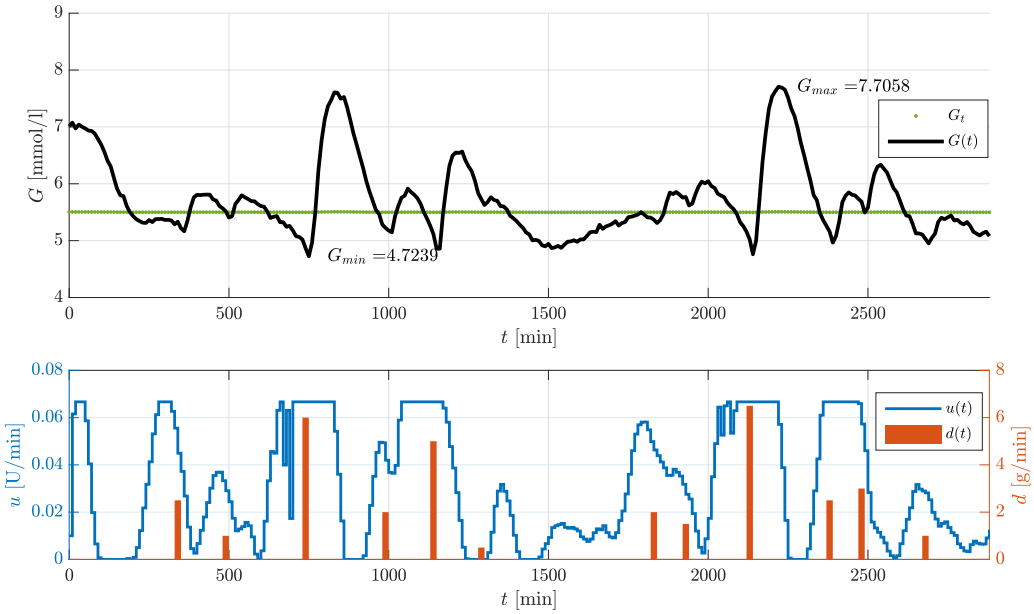


Figure 3.2: Predictive control in-silico experiment with constrained manipulated and controlled variable - meal scenario 2

achieved at the expense of higher maximal glycemia and worse performance.

3.5 Conclusions

In this chapter, the practical performance limits of the linear model predictive control were tested in such a complex and demanding application as the artificial pancreas to maintain normoglycemia in subjects with type 1 diabetes. In contrast to the traditional unconstrained predictive control law, linear inequalities were assumed for the constraints of the manipulated and the controlled variable. The emerging problem of control infeasibility was rigorously addressed by exploiting the properties of the Farkas lemma. In order to recover the feasibility and optimality of the control, an iterative algorithm was proposed to adapt the constraints of the controlled variable. However, this adaptation can be asymmetrical, what allowed to suppress hypoglycemia while tolerating mild hyperglycemia.

Chapter 4

Optimal State Estimation for the Artificial Pancreas

4.1 Introduction

The subject of this chapter is a novel approach to optimal state estimation of a discrete-time state-space model. Accordingly, the presented algorithm can be seen as an alternative to traditional state observers such as the prevailing Kalman filter. Our proposed solution considers the standard stochastic state-space model and the theoretical back iteration of the state vector with the estimation based on the generalized least squares method. Although calculating the estimate typically involves the matrix inversion operation, this can be conveniently precomputed offline, yielding online computations reduced to single matrix-vector multiplication. According to the theory of generalized least squares method, in order to obtain the minimum variance estimate, the covariance matrix of the stochastic output component has to be involved, and thereby the proposed algorithm could satisfy the criteria of the best linear unbiased estimator.

4.2 Preliminaries and Model Structure

The proposed structure of type 1 diabetes empirical model is the two-input linear discrete-time model defined by (4.1) as the sum of two transfer functions representing the insulin administration effect and the carbohydrate intake effect submodel.

$$y_{(k)} = \frac{B^u(z)}{A^u(z)} u_{(k)} + \frac{B^d(z)}{A^d(z)} d_{(k)} + y_0 + v_{(k)} \tag{4.1}$$

The output y [mmol/l] stands for glycemia, the input u [U/min] denotes the insulin administration rate (including the basal insulin) and d [g/min] is the signal of carbohydrate intake rate.

The input-output transfer function-based model (4.1) has to be transformed into the stochastic state-space model defined as:

$$x_{(k+1)} = Ax_{(k)} + B \begin{bmatrix} u_{(k)} \\ d_{(k)} \end{bmatrix} + w_{(k)} \quad y_{(k)} = Cx_{(k)} + v_{(k)} \tag{4.2}$$

where w is the process noise and v is the measurement noise, both representing uncorrelated stationary random processes with zero mean. The covariance matrix of the process noise and the variance of the measurement noise are equal to:

$$Q = \operatorname{cov}(w, w) = E\left\{ w_{(k)} w_{(k)}^{\mathrm{T}} \right\} \tag{4.3}$$

$$R = \operatorname{var}(v) = E\left\{ v_{(k)}^2 \right\} \tag{4.4}$$

Assuming the minimal state-space representation, the state vector x holds the canonical form:

$$x_{(k)} = \begin{bmatrix} x^u_{(k)} & \cdots & x^u_{(k-n_{A^u}+1)} & x^d_{(k)} & \cdots & x^d_{(k-n_{A^d}+1)} & y_0 \end{bmatrix}^{\mathrm{T}} \tag{4.5}$$

States x^u, x^d represent the partial effect of insulin administration and carbohydrate intake, respectively. The last state y_0 implements the output bias, which was constant in the original model (4.1), whereas in the case of state-space model (4.2) it is not affected by the state transition but only by the process noise in order to reflect the anticipated time variability of the basal glycemia.

The state transition matrix A for model (4.2) comprises the partial submatrices A^u, A^d and the zero matrix $\mathbf{0}$ of the conforming size.

$$A = \begin{pmatrix} A^u & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & A^d & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1 \end{pmatrix} \quad A^{u/d} = \begin{pmatrix} -a_1^{u/d} & \dots & -a_{n_{A^{u/d}}-1}^{u/d} & -a_{n_{A^{u/d}}}^{u/d} \\ 1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & 0 \end{pmatrix} \tag{4.6}$$

The input matrix is:

$$B = \begin{bmatrix} B^u & \mathbf{0} \\ \mathbf{0} & B^d \end{bmatrix} \qquad B^u = B^d = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}^{\mathrm{T}} \tag{4.7}$$

The output vector C comprises the coefficients of numerators:

$$C = \begin{bmatrix} C^u & C^d & 1 \end{bmatrix} \quad C^{u/d} = \begin{bmatrix} b_1^{u/d} & b_2^{u/d} & \dots & b_{n_{B^{u/d}}}^{u/d} \end{bmatrix} \tag{4.8}$$

4.3 Traditional State Observer Structure

The state vector (4.5) of model (4.2) is typically estimated using the traditional state observer with the following structure:

$$\hat{x}_{(k+1)} = A\hat{x}_{(k)} + B \begin{bmatrix} u_{(k)} \\ d_{(k)} \end{bmatrix} + K \left[y_{(k)} - C\hat{x}_{(k)} \right] \tag{4.9}$$

where $\hat{x}$ is the estimated state and K is the observer gain vector. Equation (4.9) represents the recursive observer where the state prior estimate is corrected by the output measurement $y_{(k)}$.

The most widespread state estimator is the Kalman filter based on the stochastic state-space model (4.2). The steady-state Kalman filter is basically equivalent to the state estimator (4.9), yet the gain vector K is designed in the way the trace of the state estimate covariance matrix $P_{(k)} = \mathrm{cov}\left(x_{(k)} - \hat{x}_{(k)}, x_{(k)} - \hat{x}_{(k)}\right)$ is minimized [16]. The steady-state covariance matrix P_∞ is determined as the solution of the discrete algebraic Riccati equation [17].

4.4 Least-Squares-Based State Estimation

The deterministic state component $\bar{x}$ of the system state (4.5), i.e. the dynamics affected solely by the input activity and undistorted by the process noise, can theoretically be separated as:

$$\bar{x}_{(k+1)} = A\bar{x}_{(k)} + B \begin{bmatrix} u_{(k)} \\ d_{(k)} \end{bmatrix} \tag{4.10}$$

The actual state x is defined as the sum:

$$x_{(k)} = \bar{x}_{(k)} + \tilde{x}_{(k)} \ , \tag{4.11}$$

where the stochastic state component $\tilde{x}$ represents the effect of the process noise, as well as it reflects the error of the initial state estimate. According to the decomposed state (4.11), the output holds:

$$y_{(k)} = C\left(\bar{x}_{(k)} + \tilde{x}_{(k)}\right) + v_{(k)} = \bar{y}_{(k)} + \tilde{y}_{(k)} + v_{(k)} \tag{4.12}$$

The dynamics of the stochastic state $\tilde{x}$ can be expressed as:

$$\tilde{x}_{(k+1)} = A\tilde{x}_{(k)} + w_{(k)} \tag{4.13}$$

Based on equation (4.13), $\tilde{x}$ can theoretically be iterated back in time:

$$\tilde{x}_{(k-i)} = A^{-i}\tilde{x}_{(k)} - \sum_{j=1}^{i} A^{j-i-1} w_{(k-j)} \tag{4.14}$$

Equations (4.12), (4.14) can be written for n recent samples in the vector form as

$$\begin{pmatrix} \tilde{y}_{(k)} \\ \tilde{y}_{(k-1)} \\ \vdots \\ \tilde{y}_{(k-n)} \end{pmatrix} = \begin{pmatrix} y_{(k)} \\ y_{(k-1)} \\ \vdots \\ y_{(k-n)} \end{pmatrix} - \begin{pmatrix} \bar{y}_{(k)} \\ \bar{y}_{(k-1)} \\ \vdots \\ \bar{y}_{(k-n)} \end{pmatrix} = \begin{pmatrix} C \\ CA^{-1} \\ \vdots \\ CA^{-n} \end{pmatrix} \tilde{x}_{(k)} + \begin{pmatrix} v_{(k)} \\ v_{(k-1)} \\ \vdots \\ v_{(k-n)} \end{pmatrix} - \begin{pmatrix} \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ CA^{-1} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ CA^{-n} & CA^{-n+1} & \dots & CA^{-1} \end{pmatrix} \begin{pmatrix} w_{(k-1)} \\ w_{(k-2)} \\ \vdots \\ w_{(k-n)} \end{pmatrix} \ , \tag{4.15}$$

while the summation in (4.14) has been transformed into the equivalent matrix form.

Equation system (4.15) may get a more compact notation:

$$\tilde{\mathcal{Y}} = \mathcal{Y} - \bar{\mathcal{Y}} = \Gamma \tilde{x}_{(k)} + \mathcal{V} - \Sigma \mathcal{W} \tag{4.16}$$

The covariance matrix of random vector $\mathcal{V} - \Sigma\mathcal{W}$ can be derived as follows.

$$\mathcal{P} = \mathcal{R} + \Sigma \mathcal{Q} \Sigma^{\mathrm{T}} \ , \tag{4.17}$$

where the covariance matrix of the joint process noise vector $\mathcal{W}$ is:

$$\mathcal{Q} = \operatorname{cov}(\mathcal{W}, \mathcal{W}) = E\left\{\mathcal{W}\mathcal{W}^{\mathrm{T}}\right\} = \begin{pmatrix} Q & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & Q \end{pmatrix} \tag{4.18}$$

The joint measurements noise vector $\mathcal{V}$ is uncorrelated so its covariance matrix is diagonal:

$$\mathcal{R} = \operatorname{cov}(\mathcal{V}, \mathcal{V}) = E\left\{\mathcal{V}\mathcal{V}^{\mathrm{T}}\right\} = RI \tag{4.19}$$

Equation system (4.16) can be solved for $\tilde{x}_{(k)}$ in the terms of generalized least squares method [18], while the corresponding quadratic cost function gets:

$$J(\hat{\tilde{x}}_{(k)}) = \frac{1}{2}\left(\mathcal{Y} - \bar{\mathcal{Y}} - \Gamma\hat{\tilde{x}}_{(k)}\right)^{\mathrm{T}} \mathcal{P}^{-1}\left(\mathcal{Y} - \bar{\mathcal{Y}} - \Gamma\hat{\tilde{x}}_{(k)}\right) \ , \tag{4.20}$$

where $\mathcal{P}$ denotes the covariance matrix of the combined noise vector $\mathcal{V} - \Sigma\mathcal{W}$ as derived in (4.17).

In order to satisfy the optimality condition $\nabla_{\hat{\tilde{x}}_{(k)}} J(\hat{\tilde{x}}_{(k)}) = \mathbf{0}$, the state estimate $\hat{\tilde{x}}_{(k)}$ must be equal to:

$$\hat{\tilde{x}}_{(k)} = \left(\Gamma^{\mathrm{T}}\mathcal{P}^{-1}\Gamma\right)^{-1}\Gamma^{\mathrm{T}}\mathcal{P}^{-1}\left(\mathcal{Y} - \bar{\mathcal{Y}}\right) \tag{4.21}$$

It is also convenient that the matrix inverses as well as the remaining matrix multiplications in (4.21) can be fully pre-computed offline, thus the state estimation reduces to simple multiplication:

$$\Omega = \left(\Gamma^{\mathrm{T}}\mathcal{P}^{-1}\Gamma\right)^{-1}\Gamma^{\mathrm{T}}\mathcal{P}^{-1} \ \rightarrow \ \hat{\tilde{x}}_{(k)} = \Omega\left(\mathcal{Y} - \bar{\mathcal{Y}}\right) \tag{4.22}$$

Finally, the true state estimate $\hat{x}_{(k)}$ can be calculated as defined in (4.11).

4.5 Simulation Experiment

The first part of the experiment was designed to emulate the typical behavior of a diabetic patient during the two-day period while subjected to multiple meal disturbances with various carbohydrates content and the application of standard insulin treatment. The second part of the experiment concerns the fully automatic insulin dosing managed by the artificial pancreas control algorithm.

The orders of empirical model (4.1) were chosen as $n_{A^u} = n_{B^u} = 4$ and $n_{A^d} = n_{B^d} = 3$. The prediction horizon was equal to $n_e = 15$ while the sample time of the virtual continuous glucose monitoring readings was $T_s = 10$ min. The statistical properties of the measurement and the process noise, represented by the variance (4.4) and the covariance matrix (4.3), were empirically guessed as:

$$Q = \operatorname{diag}\begin{pmatrix} 1 & 0 & \dots & 0 & 0.2 & 0 & \dots & 0 & 0.5 \end{pmatrix} \times 10^{-2} \quad R = 0.01 \tag{4.23}$$

Finally, the number of past output measurements n for the optimal state estimator (4.21) was chosen as $n = 30$.

4.6 Conclusions

The chapter presented a novel state estimator based on the generalized least squares method, which can be seen as an alternative to commonly used Kalman filter. The most significant contribution of this chapter was the derivation of the optimal state estimator for the linear stochastic state-space model, featuring the theoretical back iteration of the state vector and separation of the deterministic state component. Thanks to the generalized least squares method, the variance of the state estimate could be minimized, and thus the proposed algorithm satisfied the criteria of the best linear unbiased estimator. In order to reduce the on-line computational load, the algorithm was designed in the way that the corresponding matrix inversion was separated and precalculated, ultimately resulting in a single matrix-vector multiplication required to be performed at each iteration in real time.

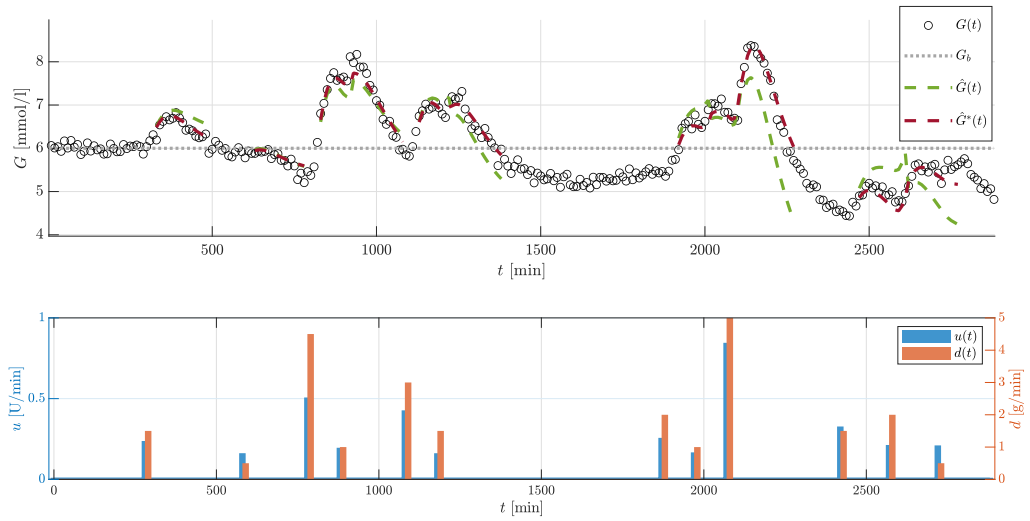


Figure 4.1: Glycemia prediction with the Kalman filter and the proposed optimal state estimator (marked by $*$)

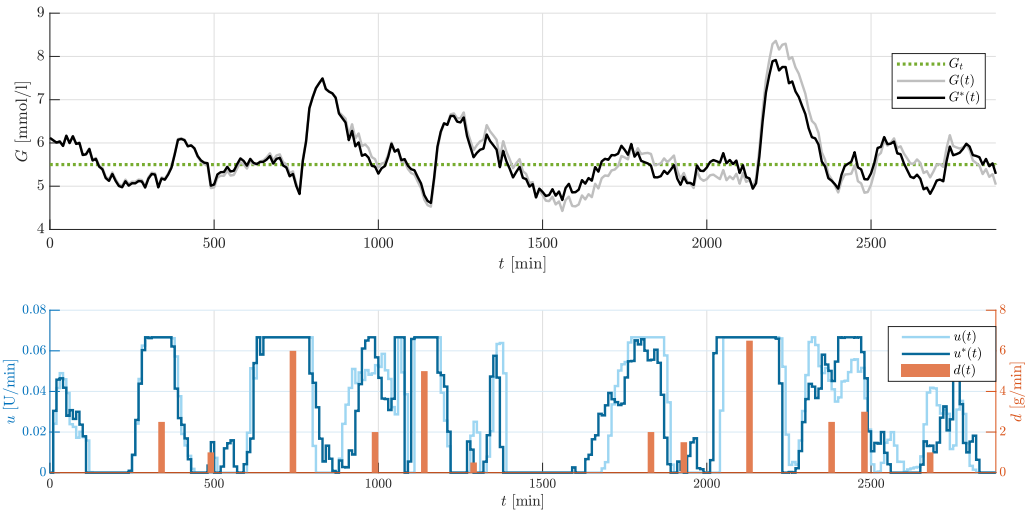


Figure 4.2: Glycemia predictive control with the Kalman filter and the proposed optimal state estimator (marked by $*$)

Chapter 5

Predicting the Output Error of the Suboptimal State Observer to Improve the Control Performance of the MPC-Based Artificial Pancreas

5.1 Introduction

The single step-ahead prediction error of the model output is typically used to correct the state estimate in the traditional state observer while exploiting the new measurement of the system output. However, its dynamics and statistical properties can be further studied, and be even exploited in other useful ways.

The widely used state estimators, such as the Kalman filter [16], are algorithms based on the prior state prediction and correction by the new output measurement. It means that the estimated state is corrected according to the error of the model single step-ahead output prediction calculated at each iteration. It is known from the theory of Kalman filtering that for an optimal filter, the sequence of the output error of the state observer (abbr. OESO), which is also called the innovation sequence, has the properties of Gaussian white noise. However, for a suboptimal filter, the innovation sequence is correlated [19, 20] and thus can be predicted.

The main motivation for studying the dynamics of the output error of the suboptimal state observer is to predict it in real time and correct the predictions of the output variable accordingly, yet the ultimate aim is to involve it in the model predictive control in order to improve the control performance.

5.2 Model Structure and Preliminaries

Discrete-time stochastic state-space empirical model of glycemia dynamics in subjects with type 1 diabetes as postulated in chapter 4 will be adopted. The model is defined by equations (4.2), (4.6), (4.7), (4.8), with the state vector (4.5) and the noise model (4.3), (4.4).

Consider that the process noise w in model (4.2) represents the effect of the input uncertainties, so one can write

$$w_{(k)} = \begin{bmatrix} \gamma_{(k)} & 0 & \dots & 0 & \delta_{(k)} & 0 & \dots & 0 \end{bmatrix}^{\mathrm{T}} . \tag{5.1}$$

The first random input $\gamma_{(k)} \sim \mathcal{N}\left(0, \sigma_\gamma^2\right)$ reflects various unmeasurable disturbances, including physiological changes in insulin absorption and action. The second random input $\delta_{(k)} \sim \mathcal{N}\left(0, \sigma_\delta^2\right)$ represents the uncertainty of the meal announcing.

Since all stochastic terms in (4.2) were defined as uncorrelated stationary random processes with zero mean, the covariance matrix $\mathcal{Q}$ of the process noise (5.1) and the variance of the measurement noise $\mathcal{R}$ are equal to

$$\begin{aligned} \mathcal{Q} &= \operatorname{cov}(w, w) = E\left\{ w_{(k)} w_{(k)}^{\mathrm{T}} \right\} \\ &= \operatorname{diag}\begin{pmatrix} \sigma_\gamma^2 & 0 & \dots & 0 & \sigma_\delta^2 & 0 & \dots & 0 \end{pmatrix} , \end{aligned} \tag{5.2}$$

$$\mathcal{R} = \operatorname{var}(v) = E\left\{ v_{(k)}^2 \right\} . \tag{5.3}$$

The state vector x (4.5) of the n-th order model (4.2) is usually estimated using the state observer (4.9). The state estimate residual e will be defined as

$$e_{(k)} = x_{(k)} - \hat{x}_{(k)} . \tag{5.4}$$

The output residual ϵ can be seen as the error of the single step-ahead prediction

$$\epsilon_{(k)} = y_{(k)} - C\hat{x}_{(k)} \ . \tag{5.5}$$

5.3 Dynamics of the State Observer Output Error

The dynamics of the output residual ϵ holds

$$\epsilon_{(k)} = C\left(zI - A + KC\right)^{-1}\left[w_{(k)} - Kv_{(k)}\right] \ . \tag{5.6}$$

Equation (5.6) implies that the dynamics of the OESO in the case of suboptimal estimator is represented by a stochastic multiple-input single-output system. One may realize that the term $C\left(zI - A + KC\right)^{-1}$ results in a row vector of rational functions $\frac{p_i(z)}{s(z)}$ with the common denominator $s(z)$ as the characteristic polynomial of this system. This consideration yields the following transfer function model of (5.6)

$$\epsilon_{(k)} = \begin{bmatrix} \frac{p_1(z)}{s(z)} & \frac{p_2(z)}{s(z)} & \dots & \frac{p_n(z)}{s(z)} \end{bmatrix}\left[w_{(k)} - Kv_{(k)}\right] \ , \tag{5.7}$$

where the characteristic polynomial $s(z)$ equals to

$$s(z) = \det\left(zI - A + KC\right) \ . \tag{5.8}$$

If $\mathcal{Q}$ is the covariance matrix of the process noise according to (5.2), then $w_{(k)}$ can be written as

$$w_{(k)} = \begin{pmatrix} \sqrt{\mathcal{Q}} & \mathbf{0} \end{pmatrix} \boldsymbol{\eta}_{(k)} \ , \tag{5.9}$$

where $\boldsymbol{\eta}$ $[n+1 \times 1]$ is the vector of uncorrelated noise inputs with the unit variance i.e. $\mathrm{cov}(\boldsymbol{\eta}, \boldsymbol{\eta}) = I$

$$\boldsymbol{\eta}_{(k)} = \begin{bmatrix} \eta_{1(k)} & \eta_{2(k)} & \dots & \eta_{n(k)} & \eta_{n+1(k)} \end{bmatrix}^{\mathrm{T}} \ , \tag{5.10}$$

and $\sqrt{\mathcal{Q}}$ is the Cholesky decomposition satisfying $\mathcal{Q} = \sqrt{\mathcal{Q}}\left(\sqrt{\mathcal{Q}}\right)^{\mathrm{T}}$ [21]. Similarly, the measurement noise $v_{(k)}$ can be replaced by

$$v_{(k)} = \begin{pmatrix} \mathbf{0}^{\mathrm{T}} & \sqrt{\mathcal{R}} \end{pmatrix} \boldsymbol{\eta}_{(k)} \ , \tag{5.11}$$

where $\mathcal{R}$ is the variance of the measurement noise according to (5.3).

Finally, the model (5.7) can be generalized as the sum of $n+1$ ARMA models by substituting $w_{(k)}$ in the terms of (5.9) and $v_{(k)}$ from (5.11) as

$$\begin{aligned} \epsilon_{(k)} &= \begin{bmatrix} \frac{p_1(z)}{s(z)} & \frac{p_2(z)}{s(z)} & \dots & \frac{p_n(z)}{s(z)} \end{bmatrix}\left[\begin{pmatrix} \sqrt{\mathcal{Q}} & \mathbf{0} \end{pmatrix} - K\begin{pmatrix} \mathbf{0}^{\mathrm{T}} & \sqrt{\mathcal{R}} \end{pmatrix}\right]\boldsymbol{\eta}_{(k)} \\ &= \begin{bmatrix} \frac{r_1(z)}{s(z)} & \frac{r_2(z)}{s(z)} & \dots & \frac{r_{n+1}(z)}{s(z)} \end{bmatrix}\boldsymbol{\eta}_{(k)} \\ &= \sum_{i=1}^{n+1} \frac{r_i(z)}{s(z)} \eta_{i(k)} \ . \end{aligned} \tag{5.12}$$

Since the process noise vector (5.1) has only two components and the covariance (5.2), the model (5.12) can be reduced to

$$\epsilon_{(k)} = \frac{p_1(z)}{s(z)}\sigma_\gamma \eta_{1(k)} + \frac{p_{n_u+1}(z)}{s(z)}\sigma_\delta \eta_{n_u+1(k)} - \frac{\sum_{i=1}^{n} K_i p_i(z)}{s(z)}\sqrt{\mathcal{R}}\eta_{n+1(k)} \ . \tag{5.13}$$

However, the structure (5.12) cannot be directly used to predict the OESO, since the random input vector η as well as the partial outputs $\frac{r_i(z)}{s(z)}$ are unmeasurable in practice. Due to the aforementioned reasons, two reduced single-input single-output stochastic model structures, particularly the autoregressive and the moving average model, will be considered.

5.4 Autoregressive Model

In this section, the dynamics of the OESO (5.12) will be approximated by the single-input single-output autoregressive model defined as

$$\epsilon_{(k)} = \frac{1}{q(z)}\eta_{(k)} \ , \tag{5.14}$$

where $\eta \sim \mathcal{N}\left(0, \sigma_\eta^2\right)$ is a random process with the properties of white noise signal. The polynomial $q(z)$ of this n_q-th order model gets

$$q(z) = 1 + q_1 z^{-1} + q_2 z^{-2} + \dots q_{n_q} z^{-n_q} \ . \tag{5.15}$$

The equivalent difference equation of model (5.14) holds

$$\epsilon_{(k)} = \eta_{(k)} - \sum_{i=1}^{n_q} q_i \epsilon_{(k-i)} \ . \tag{5.16}$$

The parameter vector $\mathbf{q}$ can be estimated as $\hat{\mathbf{q}}$ in a straightforward way using the least squares method with the optimal parameter estimate determined analytically as [22].

For model (5.14), the explicit prediction formula can be derived based on the difference equation (5.16). The future values of the white noise input are unknown, so assuming that its statistically unbiased prediction is zero i.e. $E\left\{\eta_{(k+i)}\right\}=0$, the predictive form for the prediction horizon n_e gets

$$\hat{\boldsymbol{\epsilon}}_f = -M_f^{\epsilon}(\mathbf{q})^{-1} M_p^{\epsilon}(\mathbf{q}) \boldsymbol{\epsilon}_p \ , \tag{5.17}$$

where the vectors $\boldsymbol{\epsilon}_p$ and $\hat{\boldsymbol{\epsilon}}_f$ are defined as

$$\boldsymbol{\epsilon}_p = \begin{bmatrix} \epsilon_{(k)} & \epsilon_{(k-1)} & \epsilon_{(k-2)} & \dots & \epsilon_{(k-n_q+1)} \end{bmatrix}^{\mathrm{T}} \ , \tag{5.18}$$

$$\hat{\boldsymbol{\epsilon}}_f = \begin{bmatrix} \hat{\epsilon}_{(k+1)} & \hat{\epsilon}_{(k+2)} & \hat{\epsilon}_{(k+3)} & \dots & \hat{\epsilon}_{(k+n_e)} \end{bmatrix}^{\mathrm{T}} \ , \tag{5.19}$$

and the matrices M_f^{ϵ}, M_p^{ϵ} comprise the elements of vector $\mathbf{q}$.

$$M_f^{\epsilon}(\mathbf{q}) = \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ q_1 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ q_{n_q} & q_{n_q-1} & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & q_{n_q} & \cdots & 1 \end{pmatrix} \tag{5.20}$$

$$M_p^{\epsilon}(\mathbf{q}) = \begin{pmatrix} q_1 & q_2 & \cdots & q_{n_q-1} & q_{n_q} \\ q_2 & q_3 & \cdots & q_{n_q} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ q_{n_q} & 0 & \cdots & 0 & 0 \end{pmatrix} \tag{5.21}$$

Since the noise input in (5.14) is unmeasurable, by reshaping equation (5.16), signal $\eta_{(k)}$ can be estimated as

$$\hat{\eta}_{(k)} = \epsilon_{(k)} + \sum_{i=1}^{n_q} q_i \epsilon_{(k-i)} \ . \tag{5.22}$$

5.5 Moving Average Model

In this section, the analytical stochastic model of the OESO (5.12) is approximated by the moving average structure

$$\epsilon_{(k)} = g(z)\eta_{(k)} \ , \tag{5.23}$$

where $\eta \sim \mathcal{N}\left(0, \sigma_\eta^2\right)$ is the white noise input and ϵ represents the colored noise. The polynomial $g(z)$ in the n_g-th order model (5.23) can be unwind as

$$g(z) = 1 + g_1 z^{-1} + g_2 z^{-2} + \dots g_{n_g} z^{-n_g} \ . \tag{5.24}$$

Difference equation of model (5.23) can be written as

$$\epsilon_{(k)} = \eta_{(k)} + \sum_{i=1}^{n_g} g_i \eta_{(k-i)} \ . \tag{5.25}$$

The parameter vector representing the coefficients of the model impulse response g_i gets

$$\mathbf{g} = \begin{bmatrix} g_1 & g_2 & \dots & g_{n_g} \end{bmatrix}^{\mathrm{T}} \ . \tag{5.26}$$

It is well known that estimating the parameters of moving average model is more difficult than estimating the autoregressive model [23]. Therefore, to estimate the coefficient vector (5.26) using only the available signal ϵ and the assumption that the input has the properties of white noise, the two-step Durbin's method [24, 23] is adopted.

Assuming that the statistically unbiased prediction of the input zero-mean white noise is zero, the predictive form of the moving average model (5.23) can be derived according to the difference equation (5.25) as

$$\hat{\boldsymbol{\epsilon}}_f = M_p^{\eta}(\mathbf{g})\hat{\boldsymbol{\eta}}_p \ , \tag{5.27}$$

where matrix M_p^{η} is formed by the elements of vector $\mathbf{g}$ defined by (5.26) such that

$$M_p^{\eta}(\mathbf{g}) = \begin{pmatrix} g_1 & g_2 & \cdots & g_{n_g-1} & g_{n_g} \\ g_2 & g_3 & \cdots & g_{n_g} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ g_{n_g} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix} \ , \tag{5.28}$$

and vector $\hat{\boldsymbol{\eta}}_p$ comprises the estimated past values of the noise input

$$\hat{\boldsymbol{\eta}}_p = \begin{bmatrix} \hat{\eta}_{(k)} & \hat{\eta}_{(k-1)} & \hat{\eta}_{(k-2)} & \cdots & \hat{\eta}_{(k-n_g+1)} \end{bmatrix}^{\mathrm{T}} \ . \tag{5.29}$$

In practice, the input noise signal η cannot be measured, so it has to be estimated based on the inverse filtering of the output signal ϵ according to the difference equation (5.25) as

$$\hat{\eta}_{(k)} = \epsilon_{(k)} - \sum_{i=1}^{n_g} g_i \hat{\eta}_{(k-i)} \ . \tag{5.30}$$

The corrected prediction of the output y of the state-space model (4.2) gets

$$\hat{y}_{(k+i)} = C\hat{x}_{(k+i)} + \hat{\epsilon}_{(k+i)} \ . \tag{5.31}$$

Notice that in (5.31), the predicted output error ϵ has been taken into account by correcting the output prediction $\hat{y}$.

5.6 Experimental Setup

The glycemia response for this experiment was obtained by in-silico approach using the complex physiology-based nonlinear simulation model that was described in [6, 7] and the references therein.

The orders of empirical model (4.2) were chosen as $n_{A^u} = n_{A^d} = 4$, so the overall order is $n = 8$. The prediction horizon was $n_e = 15$, while the sample time was chosen as $T_s = 10$ min. The variance of the measurement noise (5.3) and the variances of the process noise (5.2) were determined empirically as

$$\mathcal{R} = 0.01 \quad \sigma_\gamma^2 = 0.01 \quad \sigma_\delta^2 = 0.2 \ . \tag{5.32}$$

Concerning the tuning of the proposed empirical models of the OESO, the order of the autoregressive model (5.14) was set as $n_q = 4$, whereas the order of the moving average model (5.23) was chosen as $n_g = 12$.

The next comparison concerns the practical impact of correcting the output prediction by the predicted OESO. In Figure 5.1 one can see the uncorrected prediction of glycemia ($\hat{G}$), as well as the predictions that involved the corrections by the autoregressive ($\hat{G}^{AR}$) and the moving average ($\hat{G}^{MA}$) model.

The last part of the experiment is focused on the model predictive control of glycemia, where a positive effect of the proposed predictors on the control performance is anticipated. Figure 5.2 shows a visibly improved control performance characterized by a tighter control with reduced maximal and minimal observed glycemia.

5.7 Conclusions

This chapter stressed that the output error of the state observer in the case of a suboptimal state estimations can be effectively predicted and used to correct the output prediction and thus ultimately improve the performance of the model predictive control. We obtained theoretical results demonstrating that the dynamics of the output error of the state observer is analytically described as the sum of ARMA models, while this full structure can be approximated by simple autoregressive and moving average models in practice. Using these reduced predictors resulted in improved accuracy of the output variable prediction, as well as in a better performance of the model predictive control. It can be concluded that the actual effect of the proposed strategy depends primarily on the uncertainty of the process noise model.

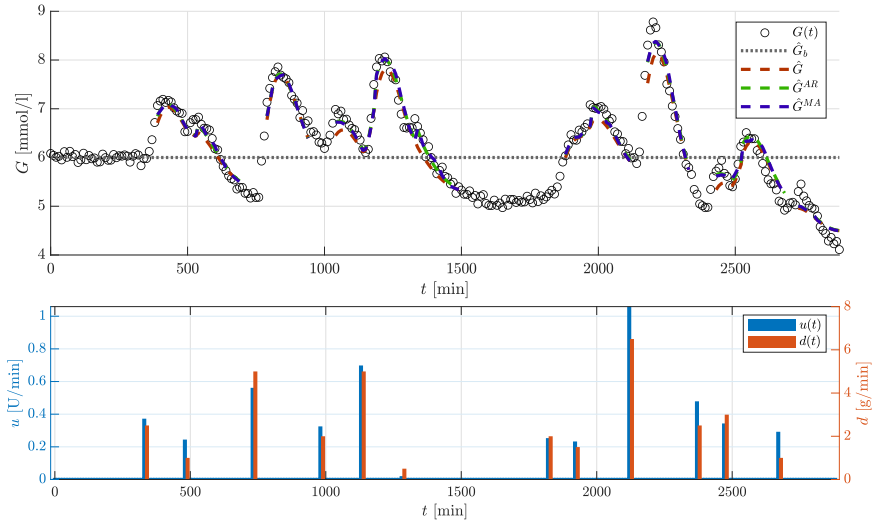


Figure 5.1: Prediction of glycemia $G(t)$ without the OESO compensation $\hat{G}$ compared to using the autoregressive model $\hat{G}^{AR}$ and the moving average model $\hat{G}^{MA}$

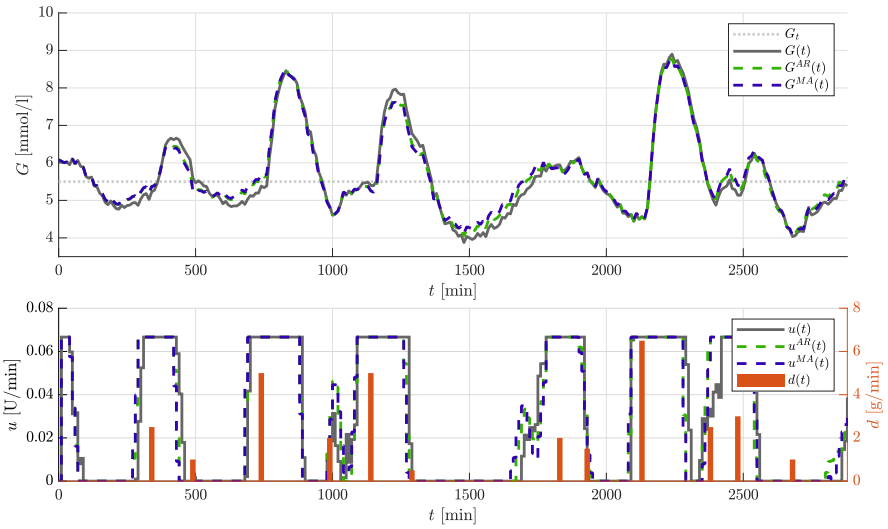


Figure 5.2: Predictive control of glycemia without the OESO compensation $G(t)$ compared to using the autoregressive model $G^{AR}(t)$ and the moving average model $G^{MA}(t)$

Chapter 6

Estimation of Process Noise Variances from the Measured Output Sequence with Application to the Empirical Model of Type 1 Diabetes

6.1 Introduction

This chapter presents a novel method for estimating a priori unknown variances of the process noise affecting a general discrete-time stochastic state-space model under the assumption that the elements of the process noise vector are not correlated. To evaluate the proposed method in practice, only the system model and a sufficiently long sequence of output measurements are required, so there is no need for the system state to be measured or the state observer to be involved. In detail, the design of the method is based on the autocorrelation function (abbr. ACF) of the stochastic output component, which is used to form the equivalent linear regression system with respect to the unknown vector of the process noise variances. Since the data-based estimate of the autocorrelation function has to be treated as uncertain and correlated, the generalized least squares method is proposed to solve the derived linear regression system, yielding an unbiased estimate with minimal variance.

6.2 Preliminaries

The basic preliminary is the following n-th order stochastic state-space model in the two-input single-output form

$$x_{(k+1)} = Ax_{(k)} + B\begin{bmatrix} u_{(k)} \\ d_{(k)} \end{bmatrix} + w_{(k)} \tag{6.1a}$$

$$y_{(k)} = Cx_{(k)} + v_{(k)} \ , \tag{6.1b}$$

where $x \in \mathbb{R}^{n\times 1}$ is the state vector, $y \in \mathbb{R}$ represents the output, whereas $u \in \mathbb{R}$ is the control input and $d \in \mathbb{R}$ is the disturbance input. Vector $w \in \mathbb{R}^{n\times 1}$ represents the process noise and $v \in \mathbb{R}$ is the measurement noise. Matrix $A \in \mathbb{R}^{n\times n}$ is the state transition matrix, $B \in \mathbb{R}^{n\times 2}$ is the input matrix and $C \in \mathbb{R}^{1\times n}$ is the output vector.

Assume that the mean values of the process noise and the measurement noise are zero, so one can write

$$E\left\{w_{(k)}\right\} = \mathbf{0} \tag{6.2}$$

$$E\left\{v_{(k)}\right\} = 0 \ . \tag{6.3}$$

The positive semidefinite symmetrical covariance matrix $\mathcal{Q} \in \mathbb{R}^{n\times n}$ of the process noise (abbr. CMPN) and the variance $\mathcal{R} \in \mathbb{R}$ of the measurement noise are formally defined using the expectancy operator as

$$\mathcal{Q} = \operatorname{cov}\left(w_{(k)}, w_{(k)}\right) = E\left\{w_{(k)} w_{(k)}^{\mathrm{T}}\right\} \ , \tag{6.4}$$

$$\mathcal{R} = \operatorname{var}\left(v_{(k)}\right) = E\left\{v_{(k)}^2\right\} \ . \tag{6.5}$$

Since the aim is to design a method for estimating the variances of the process noise, it will further be assumed

that the CMPN $\mathcal{Q}$ has a diagonal structure

$$\mathcal{Q} = \operatorname{diag}(\mathbf{q}) = \begin{pmatrix} q_1 & 0 & \dots & 0 \\ 0 & q_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & q_n \end{pmatrix} , \tag{6.6}$$

where $\mathbf{q} \in \mathbb{R}^{n\times 1}$ denotes the diagonal vector to be estimated, while its elements have to be non-negative i.e. $q_i \geq 0$ to ensure $\mathcal{Q}$ is positive semidefinite.

Moreover, we will suppose that the process noise is uncorrelated in time, so the individual samples are statistically independent according to (6.7).

$$E\left\{w_{(k)} w_{(k+i)}^{\mathrm{T}}\right\} = \mathbf{0} \quad \forall i \neq 0 \tag{6.7}$$

Also the measurement noise is assumed to be uncorrelated in time, what can be noted as

$$E\left\{v_{(k)} v_{(k+i)}\right\} = 0 \quad \forall i \neq 0 \ . \tag{6.8}$$

The process noise-induced stochastic state and the corresponding stochastic output can be noted as

$$\tilde{x}_{(k+1)} = A\tilde{x}_{(k)} + w_{(k)} \ , \tag{6.9a}$$
$$\tilde{y}_{(k)} = C\tilde{x}_{(k)} + v_{(k)} \ . \tag{6.9b}$$

The covariance matrix of the stochastic state component $\tilde{x}_{(k)}$ is actually essential for our concern. Dynamics of the stochastic state component $\tilde{x}_{(k)}$ defined by (6.9a) can be expressed with respect to the previous samples of the process noise while assuming zero initial state $\tilde{x}_{(0)} = \mathbf{0}$ as

$$\tilde{x}_{(k)} = \sum_{i=1}^{k} A^{k-i} w_{(i-1)} \ . \tag{6.10}$$

The covariance matrix of the stochastic state component $\tilde{x}_{(k)}$ can be derived as the expectancy

$$\Omega = E\left\{\tilde{x}_{(k)} \tilde{x}_{(k)}^{\mathrm{T}}\right\} = \lim_{k\to\infty} \sum_{i=0}^{k} A^i \mathcal{Q} {A^i}^{\mathrm{T}} \ . \tag{6.11}$$

In general, the ACF of a stationary continuous-time stochastic signal $z(t)$ is defined by the expectancy [25, 26]:

$$R_{zz}(\tau) = E\left\{z(t+\tau) z(t)\right\} \ , \tag{6.12}$$

where $\tau \in \mathbb{R}$ is the lag argument.

The discrete-time equivalent of (6.12), assuming the sampled signal $z_{(k)} = z(kT_s)$, gets

$$R_{zz}(nT_s) = E\left\{z_{(k+n)} z_{(k)}\right\} \ , \tag{6.13}$$

where $n \in \mathbb{N}$ is the integer lag argument and T_s is the sample time.

The ACF of the stochastic output component $\tilde{y}$ will be derived according to (6.9b) and (6.13) as

$$R_{\tilde{y}\tilde{y}}(nT_s) = C E\left\{\tilde{x}_{(k+n)} \tilde{x}_{(k)}^{\mathrm{T}}\right\} C^{\mathrm{T}} + E\left\{v_{(k)} v_{(k+n)}\right\} = \begin{cases} C\Omega C^{\mathrm{T}} + \mathcal{R} & n = 0 \\ CA^n \Omega C^{\mathrm{T}} & n \neq 0 \end{cases} . \tag{6.14}$$

6.3 Estimating the Variances of the Process Noise

Substituting different values of lag $n = 0, 1, 2 \dots P$ into the ACF (6.14) while assuming the covariance matrix Ω of the stochastic state as derived in (6.11), the equation system can be formed as

$$\begin{bmatrix} \sum_{i=0}^{\infty} \left(\left(CA^i\right) \odot \left(CA^i\right)\right) \\ \sum_{i=0}^{\infty} \left(\left(CA^{i+1}\right) \odot \left(CA^i\right)\right) \\ \sum_{i=0}^{\infty} \left(\left(CA^{i+2}\right) \odot \left(CA^i\right)\right) \\ \vdots \\ \sum_{i=0}^{\infty} \left(\left(CA^{i+P}\right) \odot \left(CA^i\right)\right) \end{bmatrix} \begin{pmatrix} q_1 \\ q_2 \\ q_3 \\ \vdots \\ q_n \end{pmatrix} = \begin{bmatrix} R_{\tilde{y}\tilde{y}}(0) - \mathcal{R} \\ R_{\tilde{y}\tilde{y}}(T_s) \\ R_{\tilde{y}\tilde{y}}(2T_s) \\ \vdots \\ R_{\tilde{y}\tilde{y}}(PT_s) \end{bmatrix} , \tag{6.15}$$

where $\odot$ denotes the Schur (element-wise) vector product.

The shorthand notation will be used for equation system (6.15)

$$\Gamma \mathbf{q} = \mathbf{R}_{\tilde{y}\tilde{y}} \ , \tag{6.16}$$

where matrix $\Gamma \in \mathbb{R}^{P+1\times n}$ and vector $\mathbf{R}_{\tilde{y}\tilde{y}} \in \mathbb{R}^{P+1\times 1}$. Note that for the calculation of each row in (6.15), the infinite summation has to be truncated to a finite one.

Suppose that in practice, the ACF $R_{\tilde{y}\tilde{y}}(nT_s)$ is estimated as $\hat{R}_{\tilde{y}\tilde{y}}(nT_s)$ from the available experimental data. The data-based estimate of (6.13), assuming the ergodicity property and a finite number of samples N, can be obtained as [25]

$$\hat{R}_{zz}(nT_s) = \frac{1}{N-n} \sum_{i=1}^{N-n} z_{(i)} z_{(i+n)} \ , \tag{6.17}$$

where the lag argument must satisfy $n < N$.

Introducing the uncertainty of the ACF estimate as the random variable $e_n = \hat{R}_{\tilde{y}\tilde{y}}(nT_s) - R_{\tilde{y}\tilde{y}}(nT_s)$, the linear equation system (6.15) becomes the linear regression system

$$\begin{bmatrix} \sum_{i=0}^{\infty} \left(\left(CA^i\right) \odot \left(CA^i\right) \right) \\ \sum_{i=0}^{\infty} \left(\left(CA^{i+1}\right) \odot \left(CA^i\right) \right) \\ \sum_{i=0}^{\infty} \left(\left(CA^{i+2}\right) \odot \left(CA^i\right) \right) \\ \vdots \\ \sum_{i=0}^{\infty} \left(\left(CA^{i+P}\right) \odot \left(CA^i\right) \right) \end{bmatrix} \begin{pmatrix} q_1 \\ q_2 \\ q_3 \\ \vdots \\ q_n \end{pmatrix} = \begin{pmatrix} \hat{R}_{\tilde{y}\tilde{y}}(0) - \mathcal{R} \\ \hat{R}_{\tilde{y}\tilde{y}}(T_s) \\ \hat{R}_{\tilde{y}\tilde{y}}(2T_s) \\ \vdots \\ \hat{R}_{\tilde{y}\tilde{y}}(PT_s) \end{pmatrix} - \begin{pmatrix} e_0 \\ e_1 \\ e_2 \\ \vdots \\ e_P \end{pmatrix} \ . \tag{6.18}$$

The shorthand notation

$$\Gamma \mathbf{q} = \hat{\mathbf{R}}_{\tilde{y}\tilde{y}} - \mathbf{e} \tag{6.19}$$

will be used for (6.18)

To solve the linear regression system (6.19) and thus estimate the vector $\mathbf{q}$, the generalized least squares method will be used. The corresponding cost function with respect to the estimated vector $\hat{\mathbf{q}}$ gets [18, 27, 28]

$$J(\hat{\mathbf{q}}) = \frac{1}{2} \left(\hat{\mathbf{R}}_{\tilde{y}\tilde{y}} - \Gamma \hat{\mathbf{q}} \right)^{\mathrm{T}} \mathcal{P}^{-1} \left(\hat{\mathbf{R}}_{\tilde{y}\tilde{y}} - \Gamma \hat{\mathbf{q}} \right) \ , \tag{6.20}$$

where $\mathcal{P}$ is the covariance matrix of the random vector $\mathbf{e}$

$$\mathcal{P} = \mathrm{cov}\left(\mathbf{e}, \mathbf{e}\right) = \mathrm{cov}\left(\hat{\mathbf{R}}_{\tilde{y}\tilde{y}}, \hat{\mathbf{R}}_{\tilde{y}\tilde{y}} \right) \ , \tag{6.21}$$

which can also be interpreted as the covariance matrix of the ACF estimate vector $\hat{\mathbf{R}}_{\tilde{y}\tilde{y}}$.

The optimal parameter estimate minimizing the cost function (6.20) of the generalized least squares method can be obtained in the closed form as [18, 27, 28]

$$\hat{\mathbf{q}} = \left(\Gamma^{\mathrm{T}} \mathcal{P}^{-1} \Gamma \right)^{-1} \Gamma^{\mathrm{T}} \mathcal{P}^{-1} \hat{\mathbf{R}}_{\tilde{y}\tilde{y}} \ . \tag{6.22}$$

The estimate can be proven as unbiased by taking the expectancy operator to (6.22) while substituting $\hat{\mathbf{R}}_{\tilde{y}\tilde{y}} = \Gamma \mathbf{q} + \mathbf{e}$ according to (6.19), what results in

$$E\left\{\hat{\mathbf{q}}\right\} = \mathbf{q} \ . \tag{6.23}$$

The covariance matrix of the estimate $\hat{\mathbf{q}}$ based on equation (6.22) can be derived as [18, 27]

$$\mathrm{cov}\left(\hat{\mathbf{q}}, \hat{\mathbf{q}}\right) = E\left\{ \left(\hat{\mathbf{q}} - \mathbf{q}\right) \left(\hat{\mathbf{q}} - \mathbf{q}\right)^{\mathrm{T}} \right\} = \left(\Gamma^{\mathrm{T}} \mathcal{P}^{-1} \Gamma \right)^{-1} \ . \tag{6.24}$$

Thanks to the generalized least squares method approach applied to the regression problem (6.19), the variances of the estimate can be considered minimal [18, 28].

The m-th row element and the n-th column element of this covariance matrix can be derived as

$$\mathcal{P}_{mn} = \left[\frac{1}{(N-m)(N-n)} \sum_{i=1}^{N-m} \sum_{j=1}^{N-n} E\left\{ z_{(i)} z_{(i+m)} z_{(j)} z_{(j+n)} \right\} \right] - R_{zz}\left(mT_s\right) R_{zz}\left(nT_s\right) \ . \tag{6.25}$$

Moreover, the ergodicity property allows one to assume that the expectancy $E\left\{ z_{(i)} z_{(i+m)} z_{(j)} z_{(j+n)} \right\}$ in (6.25) is equivalent to

$$E\left\{ z_{(i)} z_{(i+m)} z_{(j)} z_{(j+n)} \right\} = E\left\{ z_{(i+q)} z_{(i+m+q)} z_{(j+q)} z_{(j+n+q)} \right\} \quad \forall q \in \mathbb{Z} \ . \tag{6.26}$$

Consider the third-order ACF $R_{zzzz}\left(p, q, r\right)$ as the following statistics [29, 30]

$$R_{zzzz}\left(p, q, r\right) = E\left\{ z_{(i)} z_{(i+p)} z_{(i+q)} z_{(i+r)} \right\} \ . \tag{6.27}$$

However, the analytical form (6.25) and the expectancy (6.27) are infeasible in practice since the ACFs $R_{\tilde{y}\tilde{y}}(mT_s)$, $R_{\tilde{y}\tilde{y}}(nT_s)$, $R_{\tilde{y}\tilde{y}\tilde{y}\tilde{y}}(p,q,r)$ cannot be determined without the prior knowledge of the process noise model parameters. Therefore, the data-based estimates of the corresponding statistics have to be used instead. The first-order ACFs are simply approximated as $\hat{R}_{\tilde{y}\tilde{y}}(mT_s)$, $\hat{R}_{\tilde{y}\tilde{y}}(nT_s)$ according to (6.17), whereas the third-order ACFs (6.27) have to be estimated as

$$\hat{R}_{zzzz}(p,q,r) = \frac{1}{N-\max(p,q,r)} \sum_{i=1}^{N-\max(p,q,r)} z_{(i)} z_{(i+p)} z_{(i+q)} z_{(i+r)} \ . \tag{6.28}$$

6.4 Simulation Experiment

In this section, virtual diabetic data will be used to estimate the variances of the process noise in a stochastic state-space empirical model of glycemia dynamics.

First, the linear input-output model (4.1) based on discrete-time transfer functions was identified pursuing the strategy presented in chapter 2, while the model orders were chosen as $n_{A^u}=n_{B^u}=n_{A^d}=n_{B^d}=3$ implying the overall model order $n=6$.

The deterministic nonlinear simulation model that was used to obtain the sequence of the virtual glycemia measurements was subject to input uncertainties, in particular the uncertainty in the insulin administration and the carbohydrate intake.

The variance $\mathcal{R}$ of the measurement noise was chosen with regard to the properties of the continuous glucose monitoring sensors as $\mathcal{R}=0.005$. The duration of the experiment was 30 days, implying the number of samples $N=4320$ if the sample time was $T_s=10$ min.

The estimate of the ACF $\hat{R}_{\tilde{y}\tilde{y}}(nT_s)$ based on the sequence was determined according to equation (6.17).

Estimating the process noise variances vector $\mathbf{q}$ according to (6.22) yields

$$\hat{\mathbf{q}} = \begin{bmatrix} 0.0155 & 0.0055 & 0.0003 & 0.0092 & 0.0069 & 0.0020 \end{bmatrix}^{\mathrm{T}} \ . \tag{6.29}$$

6.5 Conclusions

This chapter presented a novel approach to estimation of a priori unknown variances of the process noise in the general stochastic state-space model. In a nutshell, the original contribution to the state of the art is an estimation method based on the ACF of the stochastic output component, while the estimate of the process noise variances is obtained in terms of the linear regression and the generalized least squares method. The essential part of the presented mathematical rigor was the analysis of the statistical properties of the ACF estimate, resulting in the formulation of the covariance matrix of the uncertain estimate of the autocorrelation vector, which was actually mandatory for evaluating the generalized least squares method. The strategy of minimizing the variances of the estimate have led to an increased estimate confidence, thus the occurrence of invalid negative estimates could be suppressed.

Chapter 7

Correlation Method for Identification of a Nonparametric Model of Type 1 Diabetes

7.1 Introduction

This chapter describes a novel nonparametric identification method for estimating impulse responses of the general two-input single-output linear system with its target application to the individualization of an empirical model of type 1 diabetes. The proposed algorithm is based on correlation functions and the derived generalization of the Wiener-Hopf equation for systems with two inputs, while taking the stochastic properties of the output measurements into account. Ultimately, this approach to solving the deconvolution problem can be seen as an alternative to widely used prediction error methods. To estimate the impulse response coefficients, the generalized least squares method was used in order to reflect nonuniform variances and nonzero covariances of the stochastic estimate of the cross-correlation functions, hence yielding the minimum variance estimator. Estimate regularization strategies were also involved, while three different types of penalties were applied.

7.2 Model Structure and Preliminaries

The proposed nonparametric model defines the output $y(t)$ as

$$y(t)=\int_0^\infty g^u(\lambda)u(t-\lambda)\mathrm{d}\lambda+\int_0^\infty g^d(\lambda)d(t-\lambda)\mathrm{d}\lambda+\epsilon(t)\ , \tag{7.1}$$

where $g^u(t)$ is the impulse function of the insulin administration effect, $g^d(t)$ is the impulse function of the carbohydrate intake effect, both representing the convolution kernels.

The stochastic term $\epsilon(t)\sim\mathcal{N}(0,\sigma_\epsilon^2)$ stands for the uncorrelated zero-mean random process, which reflects the continuous glucose monitoring sensor noise [31, 3], as well as the effects of various unmeasurable disturbances.

7.3 Basic Algorithm

To derive the correlation-based identification method, the cross-correlation function has to be introduced. The cross-correlation function $R_{xz}(\tau)$ of two general continuous-time infinite-length signals $x(t)$, $z(t)$ is, under the assumption of signal ergodicity, defined by the following integral transform for the lag argument $\tau\in\mathbb{R}$ [26]

$$R_{xz}(\tau)=E\left\{x(t+\tau)z(t)\right\}=\lim_{\vartheta\to\infty}\frac{1}{2\vartheta}\int_{-\vartheta}^{\vartheta}x(t+\tau)z(t)\mathrm{d}t\ . \tag{7.2}$$

Recall that if $x(t)=z(t)$, then $R_{xx}(\tau)$ is called the autocorrelation function.

According to equation (7.2), the cross-correlation function $R_{yu}(\tau)$ can be derived as

$$R_{yu}(\tau)=\int_0^\infty g^u(\lambda)R_{uu}(\tau-\lambda)\ \mathrm{d}\lambda+\int_0^\infty g^d(\lambda)R_{du}(\tau-\lambda)\ \mathrm{d}\lambda+R_{\epsilon u}(\tau)\ . \tag{7.3}$$

Taking analogous steps, we derived the cross-correlation function $R_{yd}(\tau)$ for the second input as

$$R_{yd}(\tau)=\int_0^\infty g^d(\lambda)R_{dd}(\tau-\lambda)\ \mathrm{d}\lambda+\int_0^\infty g^u(\lambda)R_{ud}(\tau-\lambda)\ \mathrm{d}\lambda+R_{\epsilon d}(\tau)\ . \tag{7.4}$$

However, for a real experiment, a finite observation time $\vartheta \neq \infty$ is assumed instead. Consequently, all the cross-correlation and the autocorrelation functions in (7.3) and (7.4) have to be replaced with their estimates $\tilde{R}_{uu}(\tau)$, $\tilde{R}_{dd}(\tau)$, $\tilde{R}_{ud}(\tau)$, $\tilde{R}_{du}(\tau)$, $\tilde{R}_{yu}(\tau)$, $\tilde{R}_{yd}(\tau)$ respectively [26]

$$\tilde{R}_{xz}(\tau) = \frac{1}{\vartheta - \tau} \int_0^{\vartheta-\tau} x(t+\tau)z(t)\ \mathrm{d}t\ . \tag{7.5}$$

7.4 Discrete-Time Form

For practical data-driven identification, the input and output signals are sampled uniformly with the sample time T_s and therefore have a discrete-time nature. If $x(t)$ is a general continuous-time signal, then we introduce the notation $x_{(k)} = x(kT_s)$ where $k \in \mathbb{N}$ represents the sample index.

Moreover, we relate the coefficients of the discrete-time impulse function g_i to the continuous-time impulse function $g(t)$ such that

$$g_i = g(iT_s)T_s\ . \tag{7.6}$$

Accordingly, the continuous-time model (7.1) will be transformed into the corresponding discrete-time form. The convolution integrals in (7.1) can be approximated by the finite summations, while the infinitesimal element $\mathrm{d}\lambda$ is replaced by $T_s > 0$, which is being absorbed into g_i according to (7.6), yielding the finite impulse response model

$$y_{(k)} = \sum_{i=0}^{M_u} g_i^u u_{(k-i)} + \sum_{i=0}^{M_d} g_i^d d_{(k-i)} + \epsilon_{(k)}\ , \tag{7.7}$$

where M_u and M_d are the assumed lengths of the impulse response coefficients vectors g^u and g^d, respectively.

Similarly, the integral in the correlation function (7.5) can be approximated by the finite summation [25], [26]

$$\tilde{R}_{xz}(nT_s) \approx \hat{R}_{xz}(n) = \frac{T_s}{N-n} \sum_{k=1}^{N-n} x_{(k+n)} z_{(k)}\ , \tag{7.8}$$

where N denotes the number of samples of the processed time series x, z and $n \in \mathbb{Z}$ is the integer lag argument satisfying the condition $n < N$. In addition, the symmetry property

$$\hat{R}_{xz}(-n) = \hat{R}_{zx}(n) \tag{7.9}$$

holds for (7.8) [26]. The discrete-time form of the generalized Wiener-Hopf equations (7.3) and (7.4) can be derived as

$$\hat{R}_{yu}(n) = \sum_{i=0}^{M_u} g_i^u \hat{R}_{uu}(n-i) + \sum_{i=0}^{M_d} g_i^d \hat{R}_{du}(n-i) + \hat{R}_{\epsilon u}(n)\ , \tag{7.10}$$

$$\hat{R}_{yd}(n) = \sum_{i=0}^{M_d} g_i^d \hat{R}_{dd}(n-i) + \sum_{i=0}^{M_u} g_i^u \hat{R}_{ud}(n-i) + \hat{R}_{\epsilon d}(n)\ , \tag{7.11}$$

where $n = 0 \dots P$ is the lag argument. Note, that the maximum lag number P should satisfy the condition

$$P \ll N\ . \tag{7.12}$$

7.4.1 Statistical Properties of Cross-Correlation Functions

The cross-correlation functions $\hat{R}_{\epsilon u}(n)$, $\hat{R}_{\epsilon d}(n)$ in equations (7.10), (7.11) cannot be directly estimated in practice because the noise term ϵ is unmeasurable. However, we may at least analyze the statistical properties of these cross-correlation functions. To this end, we will introduce the random vectors $\zeta^{\epsilon u} \in \mathbb{R}^{P+1\times 1}$, $\zeta^{\epsilon d} \in \mathbb{R}^{P+1\times 1}$ comprising the theoretical $\hat{R}_{\epsilon u}(n)$, $\hat{R}_{\epsilon d}(n)$ for different values of the argument n as

$$\zeta^{\epsilon u} = \begin{bmatrix} \hat{R}_{\epsilon u}(0) & \hat{R}_{\epsilon u}(1) & \dots & \hat{R}_{\epsilon u}(P) \end{bmatrix}^{\mathrm{T}}\ , \tag{7.13a}$$

$$\zeta^{\epsilon d} = \begin{bmatrix} \hat{R}_{\epsilon d}(0) & \hat{R}_{\epsilon d}(1) & \dots & \hat{R}_{\epsilon d}(P) \end{bmatrix}^{\mathrm{T}}\ . \tag{7.13b}$$

The vectors $\zeta^{\epsilon u}$ and $\zeta^{\epsilon d}$ can be joint into

$$\zeta = \begin{bmatrix} \zeta^{\epsilon u} \\ \zeta^{\epsilon d} \end{bmatrix}\ . \tag{7.14}$$

Taking the expectancy operator to $\hat{R}_{\epsilon u}(n)$ in the terms of equation (7.8) yields

$$E\left\{\hat{R}_{\epsilon u}(n)\right\} = \frac{T_s}{N-n} E\left\{\sum_{k=1}^{N-n} \epsilon_{(k+n)} u_{(k)}\right\} = \frac{T_s}{N-n} \sum_{k=1}^{N-n} E\left\{\epsilon_{(k+n)}\right\} u_{(k)} = 0\ . \tag{7.15}$$

Note, that the above property holds also for $\hat{R}_{\epsilon d}(n)$. With regard to this finding, it can be deduced that the mean of the vector ζ defined by (7.14) is the zero vector, hence $E\{\zeta\}=\mathbf{0}$.

The covariance matrix $\mathcal{Q}\in\mathbb{R}^{2(P+1)\times 2(P+1)}$ of the vector ζ can be divided into four block submatrices and is defined as

$$\mathcal{Q}=\begin{pmatrix}\mathcal{Q}^{\epsilon\epsilon uu} & \mathcal{Q}^{\epsilon\epsilon ud}\\ \mathcal{Q}^{\epsilon\epsilon du} & \mathcal{Q}^{\epsilon\epsilon dd}\end{pmatrix}=E\left\{\begin{pmatrix}\zeta^{\epsilon u}\zeta^{\epsilon u\mathrm{T}} & \zeta^{\epsilon u}\zeta^{\epsilon d\mathrm{T}}\\ \zeta^{\epsilon d}\zeta^{\epsilon u\mathrm{T}} & \zeta^{\epsilon d}\zeta^{\epsilon d\mathrm{T}}\end{pmatrix}\right\} \tag{7.16}$$

The i-th row and the j-th column element of the covariance matrix $\mathcal{Q}^{\epsilon\epsilon ud}\in\mathbb{R}^{P+1\times P+1}$ can be derived according to the correlation function estimate (7.8) and the definitions (7.13a), (7.13b) of the vectors $\zeta^{\epsilon u}$, $\zeta^{\epsilon d}$ as

$$\mathcal{Q}_{ij}^{\epsilon\epsilon ud}=\frac{{T_s}^2}{(N-i)\,(N-j)}\sigma_\epsilon^2\sum_{k=1}^{N-i}u_{(k)}d_{(k+i-j)}\ . \tag{7.17}$$

In the case $i<j$ complementary to (7.17), the formula can be obtained as

$$\mathcal{Q}_{ij}^{\epsilon\epsilon ud}=\frac{{T_s}^2}{(N-i)\,(N-j)}\sigma_\epsilon^2\sum_{l=1}^{N-j}d_{(l)}u_{(l+j-i)}\ . \tag{7.18}$$

One may notice that $\mathcal{Q}_{ji}^{\epsilon\epsilon ud}=\mathcal{Q}_{ij}^{\epsilon\epsilon du}$ what implies that $\mathcal{Q}^{\epsilon\epsilon du}=\left(\mathcal{Q}^{\epsilon\epsilon ud}\right)^{\mathrm{T}}$. The above steps can be taken to derive the remaining submatrices $\mathcal{Q}^{\epsilon\epsilon uu}$, $\mathcal{Q}^{\epsilon\epsilon dd}$ of (7.16). Since the covariance matrix (7.16) is symmetric, the submatrices $\mathcal{Q}^{\epsilon\epsilon uu}$, $\mathcal{Q}^{\epsilon\epsilon dd}$ are also symmetric, implying that $\mathcal{Q}_{ji}^{\epsilon\epsilon uu}=\mathcal{Q}_{ij}^{\epsilon\epsilon uu}$ and $\mathcal{Q}_{ji}^{\epsilon\epsilon dd}=\mathcal{Q}_{ij}^{\epsilon\epsilon dd}$.

7.5 Estimate of Impulse Response Coefficients

For estimating the impulse response coefficients of the nonparametric model (7.7) we will utilize the generalized least squares method [18], [27], [28] applied to the derived discrete-time Wiener-Hopf equations (7.10), (7.11). The parameter vector $g\in\mathbb{R}^{M_u+M_d+2\times 1}$ can be formally noted as

$$g=\begin{bmatrix}g^u\\ g^d\end{bmatrix}\ , \tag{7.19}$$

where the subvectors $g^u\in\mathbb{R}^{M_u+1\times 1}$ and $g^d\in\mathbb{R}^{M_d+1\times 1}$ are defined as

$$g^u=\begin{bmatrix}g_0^u & g_1^u & g_2^u & \dots & g_{M_u}^u\end{bmatrix}^{\mathrm{T}}\ , \tag{7.20a}$$

$$g^d=\begin{bmatrix}g_0^d & g_1^d & g_2^d & \dots & g_{M_d}^d\end{bmatrix}^{\mathrm{T}}\ . \tag{7.20b}$$

The cross-correlation functions $\hat{R}_{yu}(n)$, $\hat{R}_{yd}(n)$ from the Wiener-Hopf equations (7.10),(7.11), respectively, have to be reshaped into the equivalent matrix form (7.21) assuming the lag argument $n=0\dots P$.

$$\begin{pmatrix}\hat{R}_{yu}(0)\\ \hat{R}_{yu}(1)\\ \vdots\\ \hat{R}_{yu}(P)\\ \hat{R}_{yd}(0)\\ \hat{R}_{yd}(1)\\ \vdots\\ \hat{R}_{yd}(P)\end{pmatrix}=\begin{pmatrix}\hat{R}_{uu}(0) & \hat{R}_{uu}(1) & \dots & \hat{R}_{uu}(M_u) & \hat{R}_{du}(0) & \hat{R}_{ud}(1) & \dots & \hat{R}_{ud}(M_d)\\ \hat{R}_{uu}(1) & \hat{R}_{uu}(0) & \dots & \hat{R}_{uu}(M_u-1) & \hat{R}_{du}(1) & \hat{R}_{du}(0) & \dots & \hat{R}_{ud}(M_d-1)\\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots\\ \hat{R}_{uu}(P) & \hat{R}_{uu}(P-1) & \dots & \hat{R}_{uu}(P-M_u) & \hat{R}_{du}(P) & \hat{R}_{du}(P-1) & \dots & \hat{R}_{du}(P-M_d)\\ \hat{R}_{ud}(0) & \hat{R}_{du}(1) & \dots & \hat{R}_{du}(M_u) & \hat{R}_{dd}(0) & \hat{R}_{dd}(1) & \dots & \hat{R}_{dd}(M_d)\\ \hat{R}_{ud}(1) & \hat{R}_{ud}(0) & \dots & \hat{R}_{du}(M_u-1) & \hat{R}_{dd}(1) & \hat{R}_{dd}(0) & \dots & \hat{R}_{dd}(M_d-1)\\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots\\ \hat{R}_{ud}(P) & \hat{R}_{ud}(P-1) & \dots & \hat{R}_{ud}(P-M_u) & \hat{R}_{dd}(P) & \hat{R}_{dd}(P-1) & \dots & \hat{R}_{dd}(P-M_d)\end{pmatrix}\begin{pmatrix}g_0^u\\ g_1^u\\ \vdots\\ g_{M_u}^u\\ g_0^d\\ g_1^d\\ \vdots\\ g_{M_d}^d\end{pmatrix}+\begin{pmatrix}\hat{R}_{\epsilon u}(0)\\ \hat{R}_{\epsilon u}(1)\\ \vdots\\ \hat{R}_{\epsilon u}(P)\\ \hat{R}_{\epsilon d}(0)\\ \hat{R}_{\epsilon d}(1)\\ \vdots\\ \hat{R}_{\epsilon d}(P)\end{pmatrix} \tag{7.21}$$

The symmetry property (7.9) was also accounted in (7.21). Equation system (7.21) is overdetermined if condition

$$M_u+M_d<2P \tag{7.22}$$

holds, while condition (7.12) must be satisfied as well.

The full equation system (7.21) can be written in a compact form

$$\begin{pmatrix}\hat{\mathcal{R}}_{yu}\\ \hat{\mathcal{R}}_{yd}\end{pmatrix}=\begin{pmatrix}\hat{\mathcal{R}}_{uu} & \hat{\mathcal{R}}_{du}\\ \hat{\mathcal{R}}_{ud} & \hat{\mathcal{R}}_{dd}\end{pmatrix}\begin{pmatrix}g^u\\ g^d\end{pmatrix}+\begin{pmatrix}\zeta^{\epsilon u}\\ \zeta^{\epsilon d}\end{pmatrix}\ , \tag{7.23}$$

and even more simplified as

$$\hat{\mathcal{R}}_{\mathcal{Y}}=\hat{\mathcal{R}}_{\mathcal{U}}g+\zeta\ . \tag{7.24}$$

The structures of submatrices $\hat{\mathcal{R}}_{uu} \in \mathbb{R}^{P+1\times M_u+1}$, $\hat{\mathcal{R}}_{dd} \in \mathbb{R}^{P+1\times M_d+1}$, $\hat{\mathcal{R}}_{ud} \in \mathbb{R}^{P+1\times M_u+1}$, $\hat{\mathcal{R}}_{du} \in \mathbb{R}^{P+1\times M_d+1}$ and vectors $\hat{\mathcal{R}}_{yu} \in \mathbb{R}^{P+1\times 1}$, $\hat{\mathcal{R}}_{yd} \in \mathbb{R}^{P+1\times 1}$ in (7.23), as well as matrix $\hat{\mathcal{R}}_{\mathcal{U}} \in \mathbb{R}^{2(P+1)\times M_u+M_d+2}$ and vector $\hat{\mathcal{R}}_{\mathcal{Y}} \in \mathbb{R}^{2(P+1)\times 1}$ in (7.24) result from the full equation system (7.21), while the random vectors $\zeta^{\epsilon u}$, $\zeta^{\epsilon d}$ and ζ were defined in (7.14).

For the least squares-based estimation of the impulse response coefficients vector $\hat{g}$, the residuals vector $e \in \mathbb{R}^{2(P+1)\times 1}$ has to be introduced as

$$e = \begin{bmatrix} e^{yu} \\ e^{yd} \end{bmatrix} = \hat{\mathcal{R}}_{\mathcal{Y}} - \hat{\mathcal{R}}_{\mathcal{U}}\hat{g} = \begin{bmatrix} \hat{\mathcal{R}}_{yu} - \hat{\mathcal{R}}_{uu}\hat{g}^u - \hat{\mathcal{R}}_{du}\hat{g}^d \\ \hat{\mathcal{R}}_{yd} - \hat{\mathcal{R}}_{dd}\hat{g}^d - \hat{\mathcal{R}}_{ud}\hat{g}^u \end{bmatrix} . \tag{7.25}$$

The cost function of the generalized least squares method modified by adding the regularization of the estimate is defined by the quadratic form

$$J(\hat{g}) = \frac{1}{2}\left[\left(\hat{\mathcal{R}}_{\mathcal{Y}} - \hat{\mathcal{R}}_{\mathcal{U}}\hat{g}\right)^{\mathrm{T}} \mathcal{Q}^{-1}\left(\hat{\mathcal{R}}_{\mathcal{Y}} - \hat{\mathcal{R}}_{\mathcal{U}}\hat{g}\right) + \hat{g}^{\mathrm{T}}\Lambda\hat{g}\right] \tag{7.26}$$

with respect to the estimated parameter vector $\hat{g}$. In the above equation, $\mathcal{Q}$ is the covariance matrix of the noise vector ζ and $\Lambda \in \mathbb{R}^{M_u+M_d+2\times M_u+M_d+2}$ is a positive-definite symmetric regularization matrix, the design of which shall be clarified later.

According to the optimality condition $\nabla_{\hat{g}} J(\hat{g}) = \mathbf{0}$, the optimal parameter estimate $\hat{g}$ can be obtained in a closed form

$$\hat{g} = \left(\hat{\mathcal{R}}_{\mathcal{U}}^{\mathrm{T}}\mathcal{Q}^{-1}\hat{\mathcal{R}}_{\mathcal{U}} + \Lambda\right)^{-1}\hat{\mathcal{R}}_{\mathcal{U}}^{\mathrm{T}}\mathcal{Q}^{-1}\hat{\mathcal{R}}_{\mathcal{Y}} . \tag{7.27}$$

7.5.1 Estimate Regularization

Regularization is applied in order to involve some prior knowledge of the system being identified in the estimate and also to lower the estimate variance.

The particular effect of regularization depends on the linear transform operator $L \in \mathbb{R}^{n_\Gamma\times M_u+M_d+2}$ and the diagonal scaling matrix $\Gamma \in \mathbb{R}^{n_\Gamma\times n_\Gamma}$, so one can theoretically decompose the matrix Λ as [32]

$$\hat{g}^{\mathrm{T}}\Lambda\hat{g} = (L\hat{g})^{\mathrm{T}}\Gamma L\hat{g} = \hat{g}^{\mathrm{T}}L^{\mathrm{T}}\Gamma L\hat{g} , \tag{7.28}$$

where Γ is a diagonal scaling matrix

$$\Gamma = \mathrm{diag}(\gamma) . \tag{7.29}$$

If multiple types of penalty are combined, the regularization matrix Λ results from the sum $\Lambda = \Lambda^A + \Lambda^B + \Lambda^C$. In the framework of this chapter, three penalties are assumed. In particular, the smoothing operation and regularization to provide asymptotic stability and causality are applied to the estimated impulse response coefficients vector $\hat{g}$.

7.6 Case Study

In this section, virtual diabetic data generated by a complex physiology-based simulation model will be used for the validation of the proposed identification algorithm.

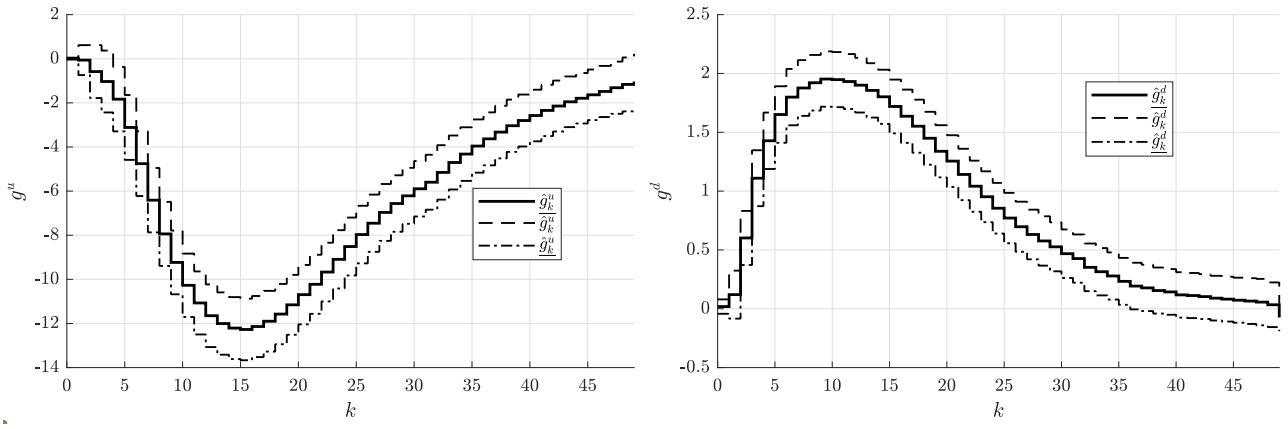
The glycemia response for this experiment was obtained in-silico, using the complex physiology-based nonlinear simulation model, discussed in [7] and [6]. The data acquisition experiment was designed to mimic the regular insulin treatment of a type 1 diabetic subject during the 6-day period with an overall number of 25 meals and a total carbohydrate amount of 433 g. The virtual continuous glucose monitoring readings were sampled with the sample time $T_s = 20$ min and the total length of the experiment was $6\times 60\times 24$ min resulting in the number of samples $N = 433$. The glycemia measurements were distorted by the additive white noise with the standard deviation $\sigma_\epsilon = 0.1$ mmol/l.

The impulse response coefficients vectors of the model (7.7) were estimated in terms of equation (7.27). In addition to the optimal solution $\hat{g}$, the confidence interval $\underline{\hat{g}}$, $\overline{\hat{g}}$ for $\alpha = 0.05$ was determined while the obtained results are summarized in a graphical form in Figure 7.1.

These long-term predictions will be performed for the nonparametric model (7.7) with the impulse responses from Figure 7.1 and also for its parametric approximation can be seen in Figure 7.2.

7.7 Conclusions

This chapter presented a correlation-based identification algorithm to estimate impulse response coefficients of the two-input linear nonparametric empirical model of type 1 diabetes. This algorithm deals with the deconvolution problem and represents an alternative to the traditional identification techniques based on the least squares minimization of the model single step-ahead prediction error.



(a) Insulin administration effect submodel

(b) Carbohydrate intake effect submodel

Figure 7.1: Estimated impulse response coefficients of the insulin administration effect submodel $\hat{g}^u$ and the carbohydrate intake effect submodel $\hat{g}^d$ together with the corresponding confidence interval $\overline{\hat{g}}$,$\underline{\hat{g}}$

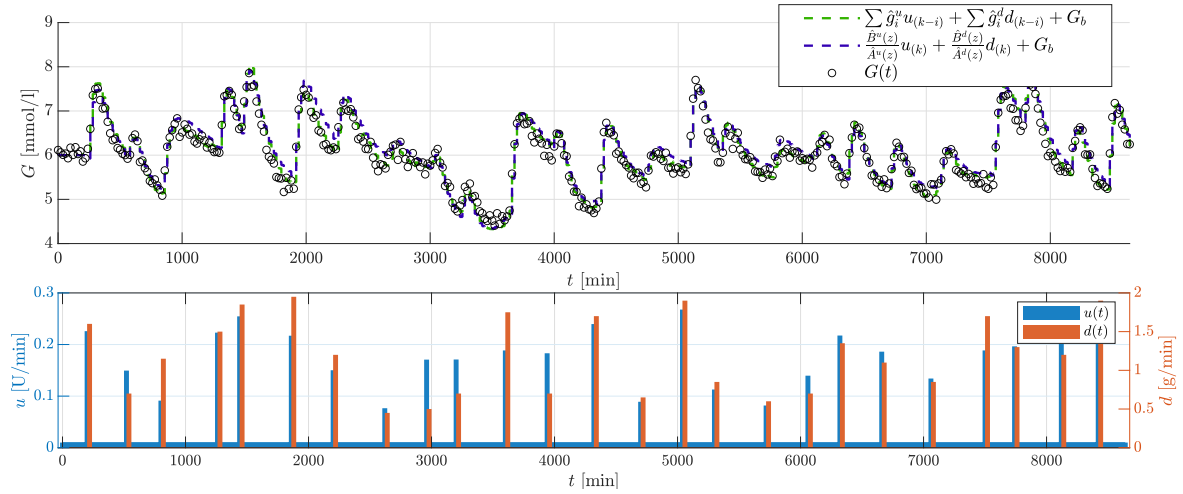


Figure 7.2: Prediction of the glycemia response for the validation dataset using the nonparametric model and the approximate parametric model

The most significant original contributions include the derivation of the generalized form of the Wiener-Hopf equation for the continuous-time model with two inputs. Based on the discrete-time equivalent of the Wiener-Hopf equation, we featured the linear equation system that resulted in formulation of the equivalent regression system. By solving this regression system in the least squares sense, the coefficients of impulse responses were estimated. Moreover, three types of regularization were applied to obtain a smoother impulse response and also to provide causality and asymptotic stability of the identified model.

Chapter 8

Robust Online Correlation Method for Identification of a Nonparametric Model of Type 1 Diabetes

8.1 Introduction

The chapter presents an improved online version of the identification method for estimating the impulse responses of the general two-input single-output linear empirical model of type 1 diabetes that allows to adapt the model parameters to intra-subject time variability in real time. The presented theory builds on and augments chapter 7 by providing important enhancements concerning the online parameter estimation, recursive formulation of the essential equations, improved regularization strategies and the last, but not least, new effective approaches to numerically solve the estimation problem. Features to robustify the estimate were also involved, as the optimal regularization strategies based on the inverse of the covariance matrix of the true parameter vector distribution and the inter-sample parameter drift were applied. Since the diabetic subject is expected to be influenced by significant intra-subject time variability of the physiology-based characteristics, there is a need to deploy online estimation algorithms that can ensure necessary adaptation of the model parameters in real time.

8.2 Model Structure and Preliminaries

Consider the two-input single-output linear nonparametric model with the finite impulse response-based structure (7.7). However, in contrast to chapter 7, here we consider the parameter-varying structure of the original time-invariant model (7.7). To this end, we introduce the bracket notation $\cdot[\ ]$ for particular objects in order to disambiguate their instances in time and make the notation of crucial recursive relations neater throughout the chapter. The output of the parameter-varying model holds

$$y_{(k)}[N] = \sum_{i=0}^{M_u} g_i^u[N] u_{(k-i)} + \sum_{i=0}^{M_d} g_i^d[N] d_{(k-i)} + \epsilon_{(k)} \ . \tag{8.1}$$

8.2.1 Exponentially Weighted Estimate of the Cross-Correlation Function

The cross-correlation function $R_{xz}(n)$ of two general discrete-time infinite-length signals $x_{(k)}$, $z_{(k)}$ is, under the assumption of ergodicity, defined by the expectancy (7.2). It is important to note that the property

$$E\left\{x_{(k+n)} z_{(k)}\right\} = E\left\{x_{(l+n)} z_{(l)}\right\} \quad \forall k, l \in \mathbb{Z} \tag{8.2}$$

holds for (7.2) if $x_{(k)}$, $z_{(k)}$ are stationary and ergodic.

Now we introduce the sample-based exponentially weighted estimate of the cross-correlation function (7.2) obtained by processing N available samples while assuming the forgetting factor $0<\lambda<1$ as

$$\hat{R}_{xz}(n)[N] = \frac{1}{\sum_{k=1}^{N-n} \lambda^{(N-n-k)}} \sum_{k=1}^{N-n} \lambda^{(N-n-k)} x_{(k+n)} z_{(k)} \ , \tag{8.3}$$

where $n \in \mathbb{Z}$ is the integer lag argument satisfying the condition $n<N$.

The sum $\sum_{k=1}^{N-n} \lambda^{(N-n-k)}$ in (8.3) will be further denoted as

$$s(n)[N] = \sum_{k=1}^{N-n} \lambda^{(N-n-k)} \ . \tag{8.4}$$

More importantly, the recursive formula for effective updating of summation (8.4) can be derived as

$$s(n)[N+1] = \sum_{k=1}^{N+1-n} \lambda^{(N+1-n-k)} = \lambda \sum_{k=1}^{N-n} \lambda^{(N-n-k)} + \lambda^0 = \lambda s(n)[N] + 1 \ . \tag{8.5}$$

Finally, considering the notation (8.4), the recursive relation to update the estimate (8.3) can be derived as

$$\hat{R}_{xz}(n)[N+1] = \frac{1}{s(n)[N+1]} \left(\lambda s(n)[N] \hat{R}_{xz}(n)[N] + x_{(N+1)} z_{(N+1-n)} \right) \ . \tag{8.6}$$

8.3 Estimate of the Impulse Response Coefficients

To estimate the impulse response coefficients, the cross-correlation functions of the system output with the inputs are essential. In this chapter, we will adopt the concerned equations (7.10),(7.11). Equations (7.10),(7.11) were transformed into the equivalent regression system (7.21). We will further use the shorthand notation (7.23)

$$\hat{\mathcal{R}}_{\mathcal{Y}} = \hat{\mathcal{R}}_{\mathcal{U}} g + \zeta \tag{8.7}$$

for (7.24).

The vectors $\hat{\mathcal{R}}_{yu} \in \mathbb{R}^{P+1 \times 1}$, $\hat{\mathcal{R}}_{yd} \in \mathbb{R}^{P+1 \times 1}$ get

$$\hat{\mathcal{R}}_{yu} = \begin{bmatrix} \hat{R}_{yu}(0) & \hat{R}_{yu}(1) & \dots & \hat{R}_{yu}(P) \end{bmatrix}^{\mathrm{T}} \ , \tag{8.8a}$$

$$\hat{\mathcal{R}}_{yd} = \begin{bmatrix} \hat{R}_{yd}(0) & \hat{R}_{yd}(1) & \dots & \hat{R}_{yd}(P) \end{bmatrix}^{\mathrm{T}} \ . \tag{8.8b}$$

8.3.1 Statistical Properties of the Parameters

In order to reflect the presence of inter-subject parametric variability typical for the diabetic patient population, we will assume that the actual parameter vector g follows the multivariate normal distribution. Therefore, the mean-population parameter vector $g_\mu \in \mathbb{R}^{M_u+M_d+2 \times 1}$ will be defined as

$$g_\mu = E\{g\} \ . \tag{8.9}$$

The covariance matrix $\Psi \in \mathbb{R}^{M_u+M_d+2 \times M_u+M_d+2}$, representing the statistical model of the inter-subject parametric variability, will be defined as

$$\Psi = \mathrm{cov}(g, g) = E\left\{ (g - E\{g\}) (g - E\{g\})^{\mathrm{T}} \right\} \ . \tag{8.10}$$

Furthermore, in order to reflect the presence of intra-subject parametric time variability, we will assume that the actual parameter vector is time-varying, so the bracket notation $g[i]$ will be used to distinguish its individual occurrences. Therefore, the actual parameter vector can be seen as a random process, while anticipating that its drift leads to no permanent bias, what can be noted as

$$E\{g[i]\} = g \ \forall i \ . \tag{8.11}$$

According to (8.11) the expectancy

$$E\{g[N+1] - g[N]\} = E\{g[N+1]\} - E\{g[N]\} = \mathbf{0} \ . \tag{8.12}$$

holds for the difference of two successive parameter vectors representing the inter-sample parameter change. The covariance matrix $\Phi \in \mathbb{R}^{M_u+M_d+2 \times M_u+M_d+2}$ of the inter-sample parameter change $g[N+1] - g[N]$ will be defined as

$$\Phi = \mathrm{cov}(g[N+1] - g[N]) = E\left\{ (g[N+1] - g[N]) (g[N+1] - g[N])^{\mathrm{T}} \right\} \ . \tag{8.13}$$

8.3.2 Generalized Least Squares Method

To solve the linear regression system (8.7) and thus estimate the parameter vector (7.19), i.e. the impulse response coefficients of the model (8.1), the generalized least squares method [18, 27, 28] will be adopted.

The corresponding cost function customized by adding two regularization terms gets the quadratic form

$$J(\hat{g}) = \frac{1}{2}\left[\left(\hat{\mathcal{R}}_{\mathcal{Y}} - \hat{\mathcal{R}}_{\mathcal{U}}\hat{g}\right)^{\mathrm{T}} \mathcal{Q}^{-1}\left(\hat{\mathcal{R}}_{\mathcal{Y}} - \hat{\mathcal{R}}_{\mathcal{U}}\hat{g}\right) + (\hat{g}-\bar{g})^{\mathrm{T}}\,\alpha\Psi^{-1}\,(\hat{g}-\bar{g}) + (\hat{g}-\hat{g}[N])^{\mathrm{T}}\,\beta\Phi^{-1}\,(\hat{g}-\hat{g}[N])\right] , \quad (8.14)$$

where $\mathcal{Q} \in \mathbb{R}^{2(P+1)\times 2(P+1)}$ is the covariance matrix of the noise vector ζ, Ψ^{-1} (8.10) and Φ^{-1} (8.13) are regularization matrices accompanied by $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}$ as scalar scaling factors, the mean-population parameter vector $\bar{g}$ and the parameter estimate from the last sample $\hat{g}[N]$.

The optimal parameter estimate that minimizes the cost function (8.14) can be obtained in the closed form assuming the optimality condition $\mathbf{g}(\hat{g}) = \mathbf{0}$ as

$$\hat{g} = \left(\hat{\mathcal{R}}_{\mathcal{U}}^{\mathrm{T}}\mathcal{Q}^{-1}\hat{\mathcal{R}}_{\mathcal{U}} + \alpha\Psi^{-1} + \beta\Phi^{-1}\right)^{-1}\left(\hat{\mathcal{R}}_{\mathcal{U}}^{\mathrm{T}}\mathcal{Q}^{-1}\hat{\mathcal{R}}_{\mathcal{Y}} + \alpha\Psi^{-1}\bar{g} + \beta\Phi^{-1}\hat{g}[N]\right) . \quad (8.15)$$

8.3.3 Regularization Strategies

One way of involving the prior knowledge in the regularization is via the inverse Ψ^{-1} of the covariance matrix (8.10) of the actual parameter vector distribution, which can be considered the "optimal" regularization matrix [33]. Therefore the first regularization term $(\hat{g}-\bar{g})^{\mathrm{T}}\,\alpha\Psi^{-1}\,(\hat{g}-\bar{g})$ in cost function (8.14) penalizes the deviation of the parameter estimate $\hat{g}$ from the mean population value $\bar{g}$.

Recall that the online parameter estimation was proposed primarily in order to reflect the time variability of the actual parameter vector, also referred to as intra-subject variability. Therefore, to robustify the identification algorithm and hence prevent unwanted excessive changes of the parameter estimate throughout the iterations due to the presence of outliers and deteriorated quality of input-output data, a dedicated regularization term $(\hat{g}-\hat{g}[N])^{\mathrm{T}}\,\beta\Phi^{-1}\,(\hat{g}-\hat{g}[N])$ is considered in the cost function (8.14). This regularization strategy penalizes the inter-sample change $\hat{g}[N+1]-\hat{g}[N]$ of the estimate using the inverse Φ^{-1} of the covariance matrix (8.13) of the drift $g[N+1]-g[N]$ of the actual parameter vector. To adjust the strength of both regularization terms, the scalar parameters $\alpha>0$ and $\beta>0$ can be tuned.

8.4 Covariance Matrix of the Cross-Correlation Function Estimate

The i-th row element and the j-th column element of the submatrix $\mathcal{Q}^{\epsilon\epsilon ud}$ can be derived according to definitions (7.13a), (7.13b) of vectors $\zeta^{\epsilon u}$, $\zeta^{\epsilon d}$ as

$$\mathcal{Q}_{ij}^{\epsilon\epsilon ud}[N] = \frac{1}{s[N](i)s[N](j)}\sigma_\epsilon^2 \sum_{k=1}^{N-i} \lambda^{2(N-i-k)} u_{(k)} d_{(k+i-j)} . \quad (8.16)$$

In the case $i<j$ complementary to (8.16), the corresponding formula can be obtained as

$$\mathcal{Q}_{ij}^{\epsilon\epsilon ud}[N] = \frac{1}{s[N](i)s[N](j)}\sigma_\epsilon^2 \sum_{l=1}^{N-j} \lambda^{2(N-j-l)} d_{(l)} u_{(l+j-i)} . \quad (8.17)$$

One may notice that $\mathcal{Q}_{ji}^{\epsilon\epsilon ud} = \mathcal{Q}_{ij}^{\epsilon\epsilon du}$ what implies that $\mathcal{Q}^{\epsilon\epsilon du} = \left(\mathcal{Q}^{\epsilon\epsilon ud}\right)^{\mathrm{T}}$. The above steps can be taken to derive the remaining submatrices $\mathcal{Q}^{\epsilon\epsilon uu}$, $\mathcal{Q}^{\epsilon\epsilon dd}$. Since the covariance matrix is symmetric, the submatrices $\mathcal{Q}^{\epsilon\epsilon uu}$, $\mathcal{Q}^{\epsilon\epsilon dd}$ are also symmetric, implying that $\mathcal{Q}_{ji}^{\epsilon\epsilon uu} = \mathcal{Q}_{ij}^{\epsilon\epsilon uu}$ and $\mathcal{Q}_{ji}^{\epsilon\epsilon dd} = \mathcal{Q}_{ij}^{\epsilon\epsilon dd}$.

Recursive relations need to be derived to effectively update the covariance matrix $\mathcal{Q}$ based on the new input data, but more importantly, to update its inverse $\mathcal{Q}^{-1}$.

Suppose that the elements of covariance submatrix $\mathcal{Q}^{\epsilon\epsilon ud}$ defined $\forall i \geq j$ by (8.16) can be obtained from $N+1$ available samples. The recursive formula can be obtained as

$$\mathcal{Q}_{ij}^{\epsilon\epsilon ud}[N+1] = \frac{1}{s[N+1](i)s[N+1](j)}\left(\lambda^2 s[N](i)s[N](j)\mathcal{Q}_{ij}^{\epsilon\epsilon ud}[N] + \sigma_\epsilon^2 u_{(N+1-i)} d_{(N+1-j)}\right) . \quad (8.18)$$

To derive the recursive formula for updating the inverse $\mathcal{Q}^{-1}$ of the covariance matrix $\mathcal{Q}$, the Sherman–Morrison formula [34] will be exploited.

Assuming column vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{\mathbf{N}\times 1}$ and matrices $\mathbf{D}, \mathbf{U}, \mathbf{V}, \mathbf{W} \in \mathbb{R}^{\mathbf{N}\times\mathbf{N}}$, we have the inversion lemma

$$\left(\mathbf{V}^{-1}\mathbf{DAUW}^{-1} + \mathbf{V}^{-1}\mathbf{u}\mathbf{v}^{\mathrm{T}}\mathbf{W}^{-1}\right)^{-1} = \mathbf{WU}^{-1}\left(\mathbf{A} + \mathbf{D}^{-1}\mathbf{u}\mathbf{v}^{\mathrm{T}}\mathbf{U}^{-1}\right)^{-1}\mathbf{D}^{-1}\mathbf{V} . \quad (8.19)$$

According to (8.19) and assuming $\mathbf{p} = \mathbf{D}^{-1}\mathbf{u}$, $\mathbf{r} = \left(\mathbf{U}^{-1}\right)^{\mathrm{T}}\mathbf{v}$ and $\mathbf{r}^{\mathrm{T}} = \mathbf{v}^{\mathrm{T}}\mathbf{U}^{-1}$, the Sherman–Morrison formula can be customized such that

$$\left(\mathbf{V}^{-1}\mathbf{DAUW}^{-1} + \mathbf{V}^{-1}\mathbf{u}\mathbf{v}^{\mathrm{T}}\mathbf{W}^{-1}\right)^{-1} = \mathbf{WU}^{-1}\mathbf{A}^{-1}\mathbf{D}^{-1}\mathbf{V} - \mathbf{WU}^{-1}\frac{\mathbf{A}^{-1}\mathbf{D}^{-1}\mathbf{u}\mathbf{v}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{A}^{-1}}{1 + \mathbf{v}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{A}^{-1}\mathbf{D}^{-1}\mathbf{u}}\mathbf{D}^{-1}\mathbf{V} \ . \tag{8.20}$$

For the inversion problem of the covariance matrix $\mathcal{Q}$, matrix $\mathbf{A}$ will represent the old covariance matrix $\mathcal{Q}[N]$ implying the dimension $\mathbf{N} = 2(P+1)$, and vectors $\mathbf{u} \in \mathbb{R}^{2(P+1)\times 1}$, $\mathbf{v} \in \mathbb{R}^{2(P+1)\times 1}$ have to get

$$\mathbf{v} = \mathbf{u} = \begin{pmatrix} u_p \\ d_p \end{pmatrix} \sigma_\epsilon \ , \tag{8.21}$$

where vectors $u_p \in \mathbb{R}^{P+1\times 1}$ and $d_p \in \mathbb{R}^{P+1\times 1}$ will be defined as

$$u_p = \begin{bmatrix} u_{(N+1)} & u_{(N)} & u_{(N-1)} & \dots & u_{(N+1-i)} & u_{(N+1-P)} \end{bmatrix}^{\mathrm{T}} \ , \tag{8.22}$$

$$d_p = \begin{bmatrix} d_{(N+1)} & d_{(N)} & d_{(N-1)} & \dots & d_{(N+1-i)} & d_{(N+1-P)} \end{bmatrix}^{\mathrm{T}} \ . \tag{8.23}$$

Matrices $\mathbf{D}, \mathbf{U}, \mathbf{V}, \mathbf{W}$ in (8.20) must have the diagonal structure

$$\mathbf{D} = \mathbf{U} = \lambda \ \mathrm{diag}\left(\mathbf{s}[N]\right) \ , \tag{8.24}$$

$$\mathbf{V} = \mathbf{W} = \mathrm{diag}\left(\mathbf{s}[N+1]\right) = \lambda\mathbf{D} + I \ , \tag{8.25}$$

where the vectors $\mathbf{s}[N]$, $\mathbf{s}[N+1]$ get

$$\mathbf{s}[N] = \begin{bmatrix} s[N](0) & s[N](1) & s[N](2) & \dots & s[N](P-1) & s[N](P) \end{bmatrix}^{\mathrm{T}} \ , \tag{8.26}$$

$$\mathbf{s}[N+1] = \begin{bmatrix} s[N+1](0) & s[N+1](1) & s[N+1](2) & \dots & s[N+1](P-1) & s[N+1](P) \end{bmatrix}^{\mathrm{T}} \ . \tag{8.27}$$

Since $\mathbf{D}$ is diagonal, its inverse is easy to calculate.

Formula (8.5) for recursive updating the sum $s[N+1](i)$ can be generalized to update the whole vector $\mathbf{s}[N+1]$ (8.27) as

$$\mathbf{s}[N+1] = \mathbf{s}[N]\lambda + \mathbf{1} \ , \tag{8.28}$$

where $\mathbf{1} \in \mathbb{R}^{\mathbf{N}\times 1}$ is the vector of ones.

The final formula for updating the inverse $\mathcal{Q}^{-1}$ can be derived according to (8.20) as

$$\mathcal{Q}^{-1}[N+1] = \mathbf{VD}^{-1}\mathcal{Q}^{-1}[N]\mathbf{D}^{-1}\mathbf{V} - \mathbf{VD}^{-1}\frac{\mathcal{Q}^{-1}[N]\mathbf{D}^{-1}\mathbf{v}\mathbf{v}^{\mathrm{T}}\mathbf{D}^{-1}\mathcal{Q}^{-1}[N]}{1 + \mathbf{v}^{\mathrm{T}}\mathbf{D}^{-1}\mathcal{Q}^{-1}[N]\mathbf{D}^{-1}\mathbf{v}}\mathbf{D}^{-1}\mathbf{V} \ . \tag{8.29}$$

8.5 Simulation Experiment

To validate the design and effectiveness of the proposed robust online identification algorithm, virtual patient diabetic data generated by a complex physiology-based simulation model will be considered. The procedure is similar to the procedure presented in chapter 7.

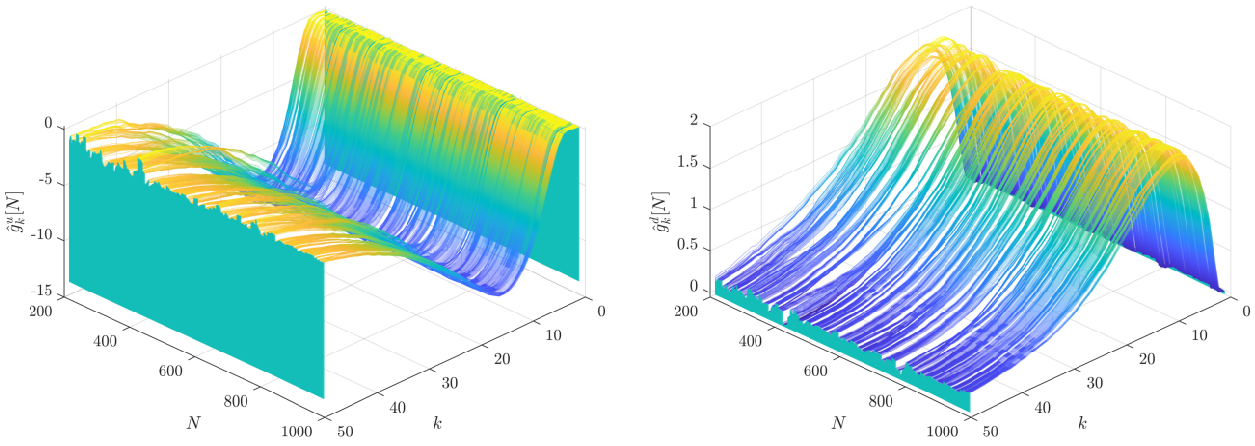
The forgetting factor for the first non-adaptive scenario was set to $\lambda = 1$ and for the second scenario that assumed the effect of parametric variability and online adaptation of impulse responses, it was chosen as $\lambda = 0.995$. The population of virtual diabetic subjects had to be generated randomly, assuming the normal distribution of all parameters of the simulation model. To involve the mechanism of intra-subject time variability into the in-silico experiment, the chosen parameters of the simulation model were considered and implemented as time-varying.

By applying the first regularization term, i.e. assuming $\alpha = 5 \times 10^2$ and omitting the second regularization term as $\beta = 0$, the sequence of impulse responses documented in Figure 8.1 was obtained by processing the parameter-invariant dataset.

Assuming both regularization strategies with the corresponding weights chosen as $\alpha = 5 \times 10^2$, $\beta = 1 \times 10^{-3}$ and the forgetting factor $\lambda = 0.995$, by processing the parameter-varying dataset we obtained the results summarized in Figure 8.2. Significant drift and the adaptation of estimated impulse responses can be observed since the chosen physiology-based parameters of the simulation model were varying.

8.6 Conclusions

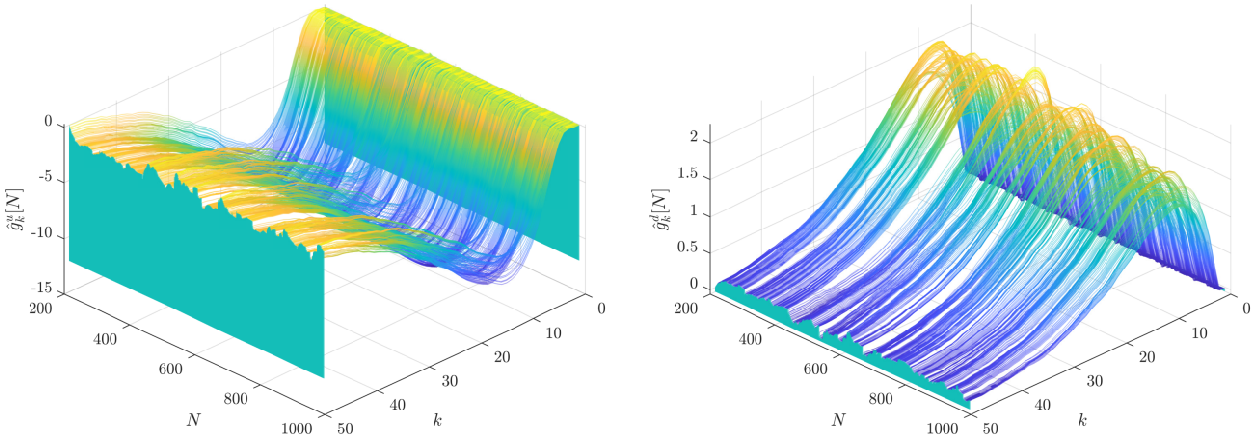
The chapter presented an extension and theoretical framework for the correlation-based method to estimate impulse responses in the case of a two-input single-output nonparametric empirical model of type 1 diabetes. The augmented method now allows effective online re-estimation and adaptation of the model parameters in



(a) Insulin administration effect submodel

(b) Carbohydrate intake effect submodel

Figure 8.1: Regularized $\alpha \neq 0$, $\beta = 0$ estimates of the impulse response coefficients as the functions of number of processed samples N, obtained by processing the parameter-invariant dataset



(a) Insulin administration effect submodel

(b) Carbohydrate intake effect submodel

Figure 8.2: Online adaptive estimates of the impulse response coefficients as the functions of number of processed samples N, obtained by processing the parameter-varying dataset

real time by considering the new available samples of input-output signals and recursively updated correlation functions. To this end, the exponentially weighted estimate of the correlation function was introduced to implement the forgetting of older samples, while the recursive formula for updating this estimate was derived to effectively exploit the new signals samples and the estimate obtained from the last iteration. One of the important findings to highlight is the recursive formula for updating the covariance matrix of the uncertainty of the correlation function estimate, but more importantly, the recursive formula for updating its inverse, which is actually essential to evaluate the generalized least squares method estimate.

Chapter 9

Optimal Model-Based Insulin Bolus Advisor for Subjects with Type 1 Diabetes

9.1 Introduction

Insulin bolus calculators are tools used by people with diabetes to help calculate the amount of insulin they need to take with meals or snacks. The goal is to help individuals maintain stable blood glucose levels throughout the day, which is important for avoiding complications of diabetes. A smart insulin bolus calculator represents an advanced advisory algorithm to support the decision making with regard to the diabetes mellitus, particularly to improve the insulin therapy. The ultimate purpose of the smart bolus advisor is to determine the time and the size of the administered insulin bolus such that the preprandial-postprandial glycemia response will be physiology-optimal while minimizing the short-term and long-term complications related to hyperglycemia or hypoglycemia. To achieve this goal, we propose a model-based optimal approach that assumes a personalized empirical model of glycemia dynamics, continuous glucose monitoring readings and a state estimator.

9.2 Model Structure and Preliminaries

Consider a deterministic linear empirical model of a subject with type 1 diabetes defined as the sum of two discrete-time transfer functions

$$y_{(k)} = \frac{B^u(z)}{A^u(z)} u_{(k)} + \frac{B^d(z)}{A^d(z)} d_{(k)} \ , \tag{9.1}$$

where the model output y [mmol/l] represents the deviation of glycemia from its basal value, i.e. the steady-state value G_b [mmol/l], the first input u [U/min] denotes the deviation of insulin administration rate from the basal insulin dosing rate u_b [U/min], and the second input d [g/min] stands for the carbohydrate intake rate. In further considerations, transfer function model (9.1) will be transformed into the equivalent minimal state-space form

$$x_{(k+1)} = Ax_{(k)} + B \begin{bmatrix} u_{(k)} \\ d_{(k)} \end{bmatrix} \quad y_{(k)} = Cx_{(k)} \ , \tag{9.2}$$

For model (9.2), the $i \in \mathbb{N}$ steps-ahead state prediction can be determined explicitly as

$$\hat{x}_{(k+i)} = A^i \hat{x}_{(k)} + \sum_{j=1}^{i} A^{i-j} B \begin{bmatrix} u_{(k+j-1)} \\ d_{(k+j-1)} \end{bmatrix} \ , \tag{9.3}$$

where $k \in \mathbb{N}$ is the current sample.

The output prediction is simply

$$\hat{y}_{(k+i)} = C\hat{x}_{(k+i)} \ . \tag{9.4}$$

Note, that in (9.3) the current state $x_{(k)}$, which is unmeasurable in practice, was replaced by its estimate $\hat{x}_{(k)}$, what implies that a state estimator (4.9) has to be considered in the structure of insulin bolus advisor.

It can be concluded that due to (4.9), the measurement-based correction of the state estimate is made, hence the decision on the insulin bolus administration is obtained online while taking the continuous glucose monitoring readings into the account.

9.2.1 Standard Bolus Calculator

Standard linear bolus calculators currently used in diabetic practice are based on a simple rule that requires the knowledge of chosen clinical parameters of the subject, particularly the insulin sensitivity and the carbohydrate sensitivity. The insulin sensitivity IS [mmol/l/U] quantifies the static effect of the administered insulin on lowering the glucose concentration, whereas the carbohydrate sensitivity parameter CS [mmol/l/g] represents the static effect of carbohydrate intake on raising glycemia. The conventional bolus calculator defines the meal compensating bolus B [U] simply as (1.2).

Concerning the timing of insulin administration, the meal intake and the corresponding insulin bolus typically occur either virtually simultaneously or the insulin administration precedes the meal intake with a constant time advance since the premeal bolus treatment strategy is broadly advised by diabetologists [2, 35].

It can be shown that empirically determined time advance and insulin-carbohydrate ratio ultimately lead to suboptimal insulin treatment as the cancellation of carbohydrate intake effect and insulin administration action is not best possible.

9.3 Optimal Bolus Calculator

The concept of model-based bolus calculator to optimize the insulin dosing strategy means that the personalized advices on when and in what quantity to administer the insulin are provided to the patient in real time in order to improve the quality and safety of therapy.

Compared to the traditional artificial pancreas [35, 36, 37, Auth4] where the insulin dosing is continuous and its administration rate is adjusted at each sample (see chapter 3), the bolus calculator assumes the control action constrained to the form of sparsely applied impulses to reject the impulse-like disturbances.

9.3.1 Input Signals

Suppose a single event of sparse impulse disturbance input signal $d_{(k)}$ as

$$d_{(k)} = a_d \delta(k-n) \ , \tag{9.5}$$

where $k \in \mathbb{N}$ is the current sample, $n \in \mathbb{N}$ is the sample number corresponding to the carbohydrate intake event, $a_d > 0 \in \mathbb{R}$ [g/min] is the magnitude related to the carbohydrate content of the meal CHO and the sample time $T_s > 0 \in \mathbb{R}$ as $a_d = CHO/T_s$, and $\delta(z)$ is the discrete approximation of the Dirac delta function defined as

$$\delta(z) = \begin{cases} 1 & z = 0 \\ 0 & z \neq 0 \end{cases} \quad \forall z \in \mathbb{Z} \ . \tag{9.6}$$

The carbohydrate intake event (9.5) will be considered known in advance as it will be announced p samples before its actual occurrence. Parameter $p \in \mathbb{N}$ thus represents the disturbance prediction horizon and can be related to n and k as

$$p = n - k \ . \tag{9.7}$$

The compensating insulin administration rate $u_{(k)}$, which is basically the subject of optimization, will also have impulse nature and will be defined as

$$u_{(k)} = \alpha \delta(k-n-\beta) \ , \tag{9.8}$$

where $\alpha > 0 \in \mathbb{R}$ [U/min] is the magnitude, and $\beta \in \mathbb{Z}$ is the time shift between the carbohydrate intake disturbance impulse and the corresponding insulin bolus administration impulse in the terms of integer multiple of the sample time T_s. It can be concluded that if $\beta > 0$ then the insulin administration is delayed, whereas if $\beta < 0$ then the insulin administration precedes the carbohydrate intake. We will further assume that β is constrained symmetrically as $-p \leq \beta \leq p$.

9.3.2 Output Prediction

Before formulating the optimization problem itself, the output prediction for model (9.2) has to be derived. Considering the preprandial-postprandial time interval, vector $Y \in \mathbb{R}^{p+M+1}$ will be defined as

$$Y = \begin{bmatrix} y_{(n-p)} & y_{(n-p+1)} & y_{(n-p+2)} & \cdots & y_{(n)} & y_{(n+1)} & \cdots & y_{(n+M-1)} & y_{(n+M)} \end{bmatrix}^{\mathrm{T}} \ , \tag{9.9}$$

where $M \in \mathbb{N}$ is the prediction horizon.

According to (9.9), vector Y can be interpreted as the preprandial-postprandial glycemia profile in the interval $-p \leq i \leq M$ samples around the carbohydrate event at sample n.

The free response vector $\tilde{Y} \in \mathbb{R}^{p+M+1}$ can be formally defined as

$$\tilde{Y} = \begin{bmatrix} \tilde{y}_{(n-p)} & \tilde{y}_{(n-p+1)} & \tilde{y}_{(n-p+2)} & \cdots & \tilde{y}_{(n)} & \tilde{y}_{(n+1)} & \cdots & \tilde{y}_{(n+M-1)} & \tilde{y}_{(n+M)} \end{bmatrix}^{\mathrm{T}} . \tag{9.10}$$

The free response $\tilde{y}$ represents the component of output evolution that depends on the current state estimate $\hat{x}_{(k)}$, the previously announced disturbances up to sample $n-1$ and their optimized compensations up to sample $n+p-1$, while excluding the effect of concerned disturbance event $d_{(n)}$ and its optimized compensation $u_{(n+\beta)}$.

According to the general prediction (9.3) and (9.4), while assuming $u_{(k)} = \delta(k)$, $d_{(k)} = 0$ and $d_{(k)} = \delta(k)$, $u_{(k)} = 0$ respectively, vectors G^u, G^d of impulse response coefficients can be written as

$$G^u = \begin{bmatrix} 0 & C^u B^u & C^u \left(A^u\right) B^u & C^u \left(A^u\right)^2 B^u & \cdots & C^u \left(A^u\right)^{p+M-1} B^u \end{bmatrix}^{\mathrm{T}} , \tag{9.11}$$

$$G^d = \begin{bmatrix} 0 & C^d B^d & C^d \left(A^d\right) B^d & C^d \left(A^d\right)^2 B^d & \cdots & C^d \left(A^d\right)^{p+M-1} B^d \end{bmatrix}^{\mathrm{T}} . \tag{9.12}$$

Introducing the linear operator $S \in \mathbb{R}^{p+M+1 \times p+M+1}$ for shifting the elements of impulse response vector one step forward in the top-bottom direction yields the matrix

$$S = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix} . \tag{9.13}$$

To obtain the operator that shift the elements of impulse response vector multiple (i) steps, the matrix power S^i applies.

Note that the operator S as defined by (9.13) is not invertible because of the first zero row, what implies that the shifting operation is not reversible. Therefore, the complementary operator S^{-1} for shifting the elements of impulse response vector one step backward has to be defined explicitly as

$$S^{-1} = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix} . \tag{9.14}$$

According to the superposition principle for linear system (9.2), the effects of input events related to the concerned disturbance and its rejection can be fully separated from the free response. As a result, the preprandial-postprandial output profile Y defined by (9.9) can be decomposed as the sum of disturbance response $S^p G^d a_d$ according to (9.5), compensating action response $S^{p+\beta} G^u \alpha$ according to (9.8), and the free response $\tilde{Y}$ as

$$Y\left(\alpha, \beta\right) = S^p G^d a_d + S^{p+\beta} G^u \alpha + \tilde{Y} , \tag{9.15}$$

where S^p, $S^{p+\beta}$ is the matrix power of operator S defined by (9.13) and G^u, G^d are the impulse response vectors defined by (9.11),(9.12) respectively.

9.3.3 Optimization

The problem considered is to optimize each insulin dose by manipulating its magnitude α and time shift β in the compensating insulin bolus administration rate impulse (9.8) in such way that the area of the resulting preprandial-postprandial deviation of the glycemia response from the target profile will be minimized.

Consider the reference vector $Y_r \in \mathbb{R}^{p+M+1}$ as

$$Y_r = \begin{bmatrix} y_{r(n-p)} & y_{r(n-p+1)} & y_{r(n-p+2)} & \cdots & y_{r(n)} & y_{r(n+1)} & \cdots & y_{r(n+M-1)} & y_{r(n+M)} \end{bmatrix}^{\mathrm{T}} . \tag{9.16}$$

The control error signal gets

$$e_{(k)} = y_{(k)} - y_{r(k)} . \tag{9.17}$$

The aim of the design is to minimize the area of the error profile $Y - Y_r$ in the least squares sense, while the corresponding cost function gets

$$\min . \; J\left(\alpha, \beta\right) = \left(Y\left(\alpha, \beta\right) - Y_r\right)^{\mathrm{T}} \Lambda \left(Y\left(\alpha, \beta\right) - Y_r\right) , \tag{9.18}$$

where Y is the output response profile (9.15), Y_r is the reference profile (9.16), α, β are the decision variables, and $\Lambda \in \mathbb{R}^{p+M+1 \times p+M+1}$

$$\Lambda = \operatorname{diag}(\lambda) \succeq 0 \tag{9.19}$$

is a positive-definite diagonal weighting matrix. The optimization problem (9.18) involves the following constraints

$$u_{\min} \leq \alpha \leq u_{\max} \ , \tag{9.20}$$

$$-p \leq \beta \leq p \ . \tag{9.21}$$

Substituting Y according to (9.15) into the cost function (9.18) yields

$$J(\alpha, \beta) = \left(S^p G^d a_d + S^{p+\beta} G^u \alpha + \tilde{Y} - Y_r\right)^{\mathrm{T}} \Lambda \left(S^p G^d a_d + S^{p+\beta} G^u \alpha + \tilde{Y} - Y_r\right) \ , \tag{9.22}$$

Cost function (9.22) represents a bivariate mixed integer-real optimization problem, which needs to be solved to determine the optimal α^* and β^*.

The optimal value of α can be determined analytically. The partial derivative $\frac{\partial J(\alpha,\beta)}{\partial \alpha}$ of (9.22) with respect to α is equal to

$$\frac{\partial J(\alpha, \beta)}{\partial \alpha} = 2\left(S^p G^d\right)^{\mathrm{T}} \Lambda S^{p+\beta} G^u a_d + 2\left(S^{p+\beta} G^u\right)^{\mathrm{T}} \Lambda S^{p+\beta} G^u \alpha + 2\tilde{Y}^{\mathrm{T}} \Lambda S^{p+\beta} G^u - 2Y_r^{\mathrm{T}} \Lambda S^{p+\beta} G^u \ . \tag{9.23}$$

Considering (9.23) and assuming the optimality condition $\frac{\partial J(\alpha,\beta)}{\partial \alpha} = 0$, the formula for optimal α gets

$$\alpha^* = -\frac{\left(a_d S^p G^d + \tilde{Y} - Y_r\right)^{\mathrm{T}} \Lambda S^{p+\beta} G^u}{\left(G^u\right)^{\mathrm{T}} \left(S^{p+\beta}\right)^{\mathrm{T}} \Lambda S^{p+\beta} G^u} \ . \tag{9.24}$$

Equation (9.24) implies that the optimal insulin administration rate α^* is linearly dependent on the magnitude of the disturbance a_d, which is related to the carbohydrate content of the meal.

Cost function (9.22) can be further simplified by substituting the optimal α^* according to (9.24) into (9.22) what yields the univariate cost function

$$J(\beta) = \left(S^p G^d a_d + \tilde{Y} - Y_r\right)^{\mathrm{T}} \Lambda \left(S^p G^d a_d + \tilde{Y} - Y_r\right) - \frac{\left(\left(S^p G^d a_d + \tilde{Y} - Y_r\right)^{\mathrm{T}} \Lambda S^{p+\beta} G^u\right)^2}{\left(G^u\right)^{\mathrm{T}} \left(S^{p+\beta}\right)^{\mathrm{T}} \Lambda S^{p+\beta} G^u} \ . \tag{9.25}$$

The original bivariate mixed integer-real optimization problem (9.22) thus reduces to univariate integer optimization problem, which is easy to solve iteratively since β is constrained as (9.21).

It is apparent that there exist no analytical formula for $\beta^\star$, hence a simple iterative Algorithm 2 is proposed. In Algorithm 2, variable $s \in \mathbb{Z}$ represents the search direction and sign $(\cdot)$ is the signum function.

9.4 Experiment

In order to evaluate the proposed improvements and assess the actual effectiveness of the model-based bolus advisor, a simulation-based experiment will be carried out and the results will be discussed in this section. Each simulation experiment is designed to mimic the multiple daily insulin injection treatment of a patient with type 1 diabetes during the ten-day period while being a subject to multiple meal disturbances per day with varying carbohydrate content.

The considered length of the experiment was 10 days including 5 meal disturbances per day. The chosen sample time $T_s = 20$ min implies the total number of samples $N = 720$. The target glycemia was set to 5.5 mmol/l, hence the reference vector (9.16) gets $Y_r = \begin{pmatrix}1 & 1 & \dots & 1\end{pmatrix}^{\mathrm{T}} \times (5.5 - G_b) = -\begin{pmatrix}1 & 1 & \dots & 1\end{pmatrix}^{\mathrm{T}} \times 0.5$ mmol/l.

Concerning the tuning of the optimal model-based bolus calculator, the carbohydrate intake disturbance will be considered known $p = 6$ samples i.e. 120 minutes before its occurrence, what appears to be a realistic requirement feasible in clinical practice. The optimized prediction horizon was chosen as as $M = 50$. This implies that the bounds (9.21) of the decision variable β are $-6 \leq \beta \leq 6$. The bounds (9.20) for the magnitude α were set to $0 < \alpha < 0.5$ U/min, in order to reflect the inherited positivity of the insulin administration.

Global performance of the insulin therapy will be quantified by the quadratic metric

$$Q = \sum_{k=1}^{N} e_{(k)}^2 = \sum_{k=1}^{N} \left(y_{(k)} - y_{r\,(k)}\right)^2 \ . \tag{9.26}$$

Simulation results obtained by applying suboptimal and personalized optimal insulin treatment strategy will be discussed in this subsection. The performance of insulin therapy will be quantified by the maximal $e_{\max}$ and

Parameters: $p, M, \Lambda, A^u, A^d, B^u, B^d, C^u, C^d, u_{\min}, u_{\max}$
Data: $a_d, \hat{x}_{(k)}, u_f, d_f, Y_r$
Result: α^*, β^*
Assume: $J(\beta) = -\frac{\left(\left(S^p G^d a_d + \tilde{Y} - Y_r\right)^{\mathrm{T}} \Lambda S^{p+\beta} G^u\right)^2}{(G^u)^{\mathrm{T}} (S^{p+\beta})^{\mathrm{T}} \Lambda S^{p+\beta} G^u}$, $S =$(9.13) , $S^{-1} =$(9.14)
begin

$$G^u = \begin{bmatrix} 0 & C^u B^u & C^u (A^u) B^u & C^u (A^u)^2 B^u & \dots & C^u (A^u)^{p+M-1} B^u \end{bmatrix}^{\mathrm{T}} ;$$

$$G^d = \begin{bmatrix} 0 & C^d B^d & C^d \left(A^d\right) B^d & C^d \left(A^d\right)^2 B^d & \dots & C^d \left(A^d\right)^{p+M-1} B^d \end{bmatrix}^{\mathrm{T}} ;$$

$$\tilde{Y} = \mathcal{A}\hat{x}_{(k)} + \mathcal{B}^u u_f + \mathcal{B}^d d_f ;$$

$$\beta^* \leftarrow 0 ;$$

$$s \leftarrow \operatorname{sign}(J(-1) - J(1));$$

while $-p < \beta^* < p$ **do**

  if $J(\beta^* + s) > J(\beta^*)$ **then**

    break;

  end

  $\beta^* \leftarrow \beta^* + s$;

end

$$\alpha^* \leftarrow -\frac{\left(a_d S^p G^d + \tilde{Y} - Y_r\right)^{\mathrm{T}} \Lambda S^{p+\beta^*} G^u}{(G^u)^{\mathrm{T}} \left(S^{p+\beta^*}\right)^{\mathrm{T}} \Lambda S^{p+\beta^*} G^u} ;$$

$$\alpha^* \leftarrow \max(\alpha^*, u_{\min}) ;$$

$$\alpha^* \leftarrow \min(\alpha^*, u_{\max}) ;$$

end

Algorithm 2: Algorithm to determine the optimal α and β for the model-based insulin bolus advisor

minimal $e_{\min}$ control error (9.17), i.e. the observed deviation of glycemia from the reference value, as well as by the global performance criterion Q determined according to (9.26).

Figure 9.1 shows an application of the traditional suboptimal bolus calculator (1.2) with $\beta = 0$ i.e. with simultaneous carbohydrate intake and administration of insulin, while the insulin-carbohydrate ratio and insulin sensitivity were determined suboptimally according the model static gains as $ICR = 3.9740$ g/U, $IS = 28.52$ mmol/l/U. Despite the suboptimal design, the obtained glycemia response is still relatively acceptable and safe from the clinical point of view, yet there can be observed poor management of postprandial hyperglycemia. Another issue is the the response did not follow the target glycemia 5.5 mmol/l as its mean value is significantly higher.

Optimizing the insulin dosing according to the proposed strategy and Algorithm 2 by considering variable α and β, much more effective rejection of the disturbances can be observed in Figure 9.2. There can also be seen a better trade-off between the preprandial hypoglycemia and postprandial hyperglycemia and a smaller area of the response curve with respect to the reference value.

The resulting overall performance comparison is summarized in Table 9.1.

Experiment	Q	$e_{\max}$	$e_{\min}$
suboptimal insulin dosing	448.1923	2.3857	-0.4308
optimal insulin dosing	117.4105	1.0740	-0.8863

Table 9.1: Performance metrics evaluated for all assumed insulin dosing strategies

9.5 Conclusions

In a nutshell, the original contributions to the state of the art presented in this chapter include the design of the cost function and the algorithm to determine the optimal personalized timing and the size of insulin bolus administration by minimizing the deviation of the preprandial-postprandial glycemia response from the target trajectory.

Compared to conventional suboptimal bolus calculators widely used in clinical practice, the proposed smart bolus calculator allows to optimally adjust the timing and the size of insulin bolus administration based on a personalized empirical model of type 1 diabetes and the continuous glucose monitoring measurements in such way that the resulting glycemia response will be “best possible” in the terms of physiological optimality while minimizing short-term risks and long-term complications related to hyperglycemia or hypoglycemia state.

Compared to the traditional concept of artificial pancreas, the proposed insulin bolus advisor considers a sparse and impulse-like insulin administration, which is related to the rejection of carbohydrate intake events.

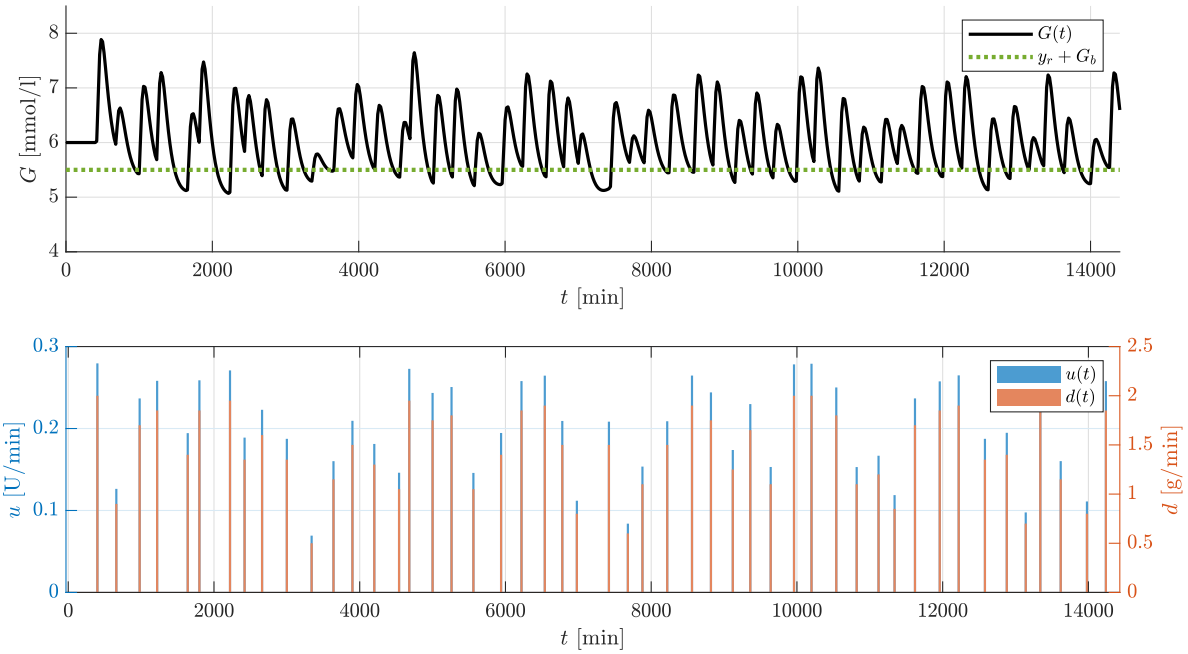


Figure 9.1: Simulation of the insulin treatment pursuing the suboptimal dosing strategy

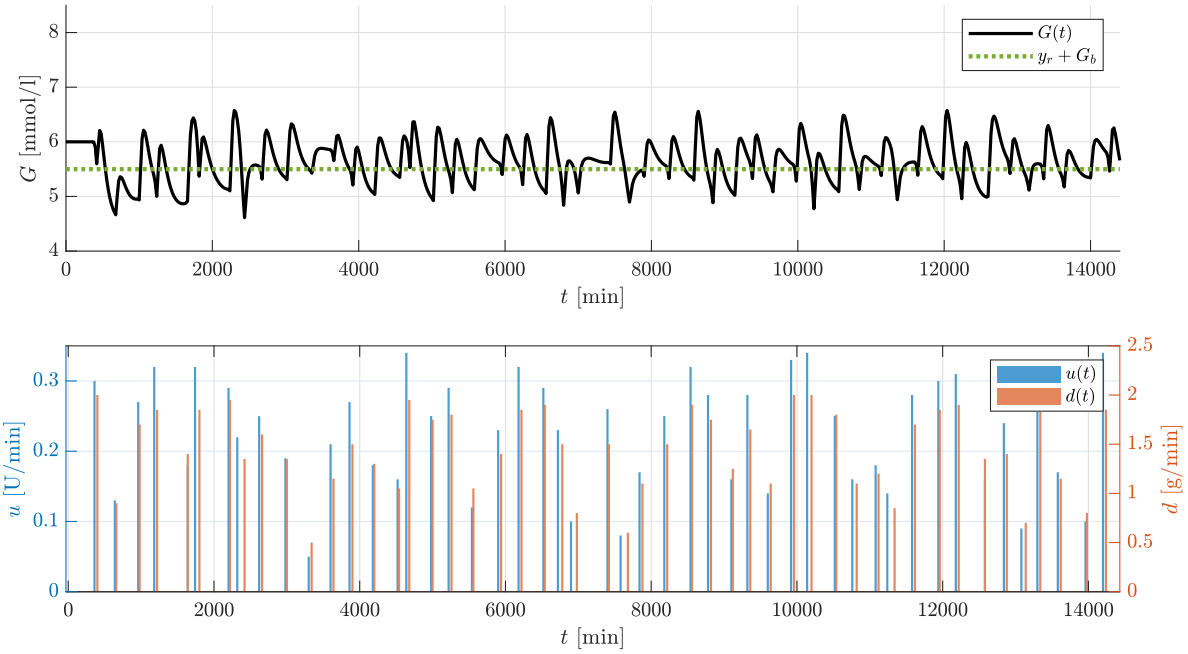


Figure 9.2: Simulation of the insulin treatment pursuing the optimal dosing strategy

Conclusions

This thesis dealt with the application of originally engineering and especially cybernetic approaches to model and control glycemia in type 1 diabetic patients. The ultimate goal was to design a complex advisory system to support decision making with regard to insulin therapy in diabetic subjects in the form of a set of algorithms for each of the emerging partial subproblems. The presented solution to this complex and holistic problem involved the design of multiple innovative algorithms and methods. First, we concerned with the empirical modeling of glycemia dynamics, where we proposed an improved transfer function-based model featuring different autoregressive dynamics for both inputs and a separate noise model as a better alternative to the ARX model. This topic also included the identification method based on multi-step-ahead predictive identification as a substitute to the tradition single-step-ahead prediction error criterion. Identification was performed using numerical optimization in the constrained parameter space of model zeros, poles, and gains to ensure compliance of the estimated model with the basic physiology.

The problem of model predictive control and the implementation of the so-called artificial pancreas to control glycemia in diabetic subjects in terms of automatic insulin administration was also addressed. We focused on studying the effect of constraints of the controlled variable to achieve a better management of the postprandial hyperglycemia and hypoglycemia. As the issue of potential control infeasibility has emerged if there were applied too tight constraints of the controlled and the manipulated variable concurrently, we proposed and validated an algorithm to adapt the constraints of the controlled variable and thus regain the feasibility and optimality.

In the case of physiology-based modeling or empirical modeling of glycemia dynamics, the actual internal state of the subject is completely unknown. We proposed a novel state estimation algorithm, which is based on the generalized least squares method. This state estimator can be seen as an alternative to the widely-used Kalman filter. The proposed state estimator utilized the sequence of past measurements of the output to form the linear regression system, which has to be solved in the least squares sense.

Since it is a relatively common encounter for a state observer to perform suboptimally in the application of model predictive control-based artificial pancreas primarily due to the unknown and hence usually empirically tuned noise model parameters, it was essential to show that the output error of this suboptimal state observer forms a correlated sequence, which can be effectively predicted. To this end, we derived an analytical model of this output error and then proposed two reduced model structures to be used in practice. Namely the autoregressive and the moving average models were considered, while their estimation was possible using only data available from the operation of the state observer. Using these estimated models as predictors of the output error of the state observer to correct the predictions of glycemia has also led to an improved performance of the model predictive control of glycemia.

Process noise affecting the stochastic state-space empirical model of glycemia dynamics was introduced primarily to model input uncertainties, particularly the uncertainty of meal announcing and insulin kinetics. However, the covariance matrix in the process noise model had to be treated as a priori unknown, so we designed a novel method for estimating its diagonal entries from the available experimental data. The proposed estimation method was based on fitting the autocorrelation function of the output of the estimated stochastic state-space model to the sample autocorrelation function obtained from the experimental data.

Another partial subproblem of the complex topic of the advisory system was the identification of the empirical model of glycemia dynamics from virtual diabetic data. Therefore, we designed a novel identification method for the estimation of the two-input, single-output nonparametric impulse response model. The proposed identification method was based on correlation functions and the generalized least squares method with the estimate regularization. As an extension of this approach, we later designed its online version to allow for an effective real-time parameter reestimation and adaptation of impulse responses due to the presence of parametric time variability of a real diabetic subject. Moreover, a new optimal kernel-based regularization strategy was designed to improve the robustness properties of the parameter estimate.

Finally, we addressed the problem of optimizing the conventional insulin treatment by proposing an innovative model-based bolus calculator algorithm to generate real-time advice on the optimal time and quantity of insulin bolus administration to compensate for meal-related carbohydrate intake. Thus the resulting preprandial-postprandial glycemia response of the subject could be improved and mimic the physiological response of a healthy subject. The proposed optimal bolus calculator strategy was based on a sparse and impulse control action, which is more feasible to apply in clinical practice.

Bibliography

[1] S. Schmidt and K. Nørgaard, "Bolus calculators," *Journal of Diabetes Science and Technology*, vol. 8, 05 2014.

[2] C. Boughton, S. Hartnell, J. Allen, and R. Hovorka, "The importance of prandial insulin bolus timing with hybrid closed-loop systems," *Diabetic Medicine*, vol. 36, pp. 1716 – 1717, 2019.

[3] H. Kirchsteiger, J. Jørgensen, E. Renard, and L. del Re, eds., *Prediction Methods for Blood Glucose Concentration: Design, Use and Evaluation*. Lecture Notes in Bioengineering, Springer, 2016.

[4] D. Romeres, M. Schiavon, A. Basu, C. Cobelli, R. Basu, and C. Dalla Man, "Exercise effect on insulin-dependent and insulin-independent glucose utilization in healthy individuals and individuals with type 1 diabetes: a modeling study," *American Journal of Physiology-Endocrinology and Metabolism*, vol. 321, no. 1, pp. E122–E129, 2021.

[5] C. Toffanin, E. M. Aiello, S. D. Favero, C. Cobelli, and L. Magni, "Multiple models for artificial pancreas predictions identified from free-living condition data: A proof of concept study," *Journal of Process Control*, vol. 77, pp. 29–37, 2019.

[6] C. Dalla Man, R. Rizza, and C. Cobelli, "Mixed meal simulation model of glucose-insulin system," Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings, pp. 307–310, 2006.

[7] C. Dalla Man, R. A. Rizza, and C. Cobelli, "Meal simulation model of the glucose-insulin system," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 10, pp. 1740–1749, 2007.

[8] L. Magni, D. Raimondo, C. D. Man, G. De Nicolao, B. Kovatchev, and C. Cobelli, "Model predictive control of glucose concentration in subjects with type 1 diabetes: an in silico trial," *IFAC Proceedings Volumes*, vol. 41, no. 2, pp. 4246 – 4251, 2008. 17th IFAC World Congress.

[9] L. Magni, D. M. Raimondo, L. Bossi, C. D. Man, G. D. Nicolao, B. Kovatchev, and C. Cobelli, "Model predictive control of type 1 diabetes: An in silico trial," *Journal of Diabetes Science and Technology*, vol. 1, no. 6, pp. 804–812, 2007.

[10] P. Tatjewski, "Effectiveness of dynamic matrix control algorithm with laguerre functions," *Archives of Control Sciences*, vol. vol. 31, no. No 4, pp. 795–814, 2021.

[11] R. Nebeluk and P. Marusak, "Efficient mpc algorithms with variable trajectories of parameters weighting predicted control errors," *Archives of Control Sciences*, vol. vol. 30, no. No 2, pp. 325–363, 2020.

[12] R. Haber, R. Bars, and U. Schmitz, *Generalized Predictive Control of Linear SISO Processes*, ch. 5, pp. 135–220. John Wiley & Sons, Ltd, 2011.

[13] R. Hovorka, V. Canonico, L. J. Chassin, U. Haueter, M. Massi-Benedetti, M. O. Federici, T. R. Pieber, H. C. Schaller, L. Schaupp, T. Vering, and M. E. Wilinska, "Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes," *Physiological Measurement*, vol. 25, pp. 905–920, jul 2004.

[14] R. S. Parker, F. J. Doyle, and N. A. Peppas, "A model-based algorithm for blood glucose control in type i diabetic patients," *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 2, pp. 148–157, 1999.

[15] J. Matousek and B. Gärtner, *Understanding and Using Linear Programming*. Universitext, Springer Berlin Heidelberg, 2006.

[16] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, vol. 82, pp. 35–45, 03 1960.

[17] B. Anderson and J. Moore, *Optimal Filtering*. Dover Books on Electrical Engineering, Dover Publications, 2012.

[18] T. Amemiya, *Advanced Econometrics*. Cambridge, MA: Harvard University Press, 1st ed., 1985.

[19] R. Mehra, "Approaches to adaptive filtering," *IEEE Transactions on Automatic Control*, vol. 17, no. 5, pp. 693–698, 1972.

[20] R. Mehra, "On the identification of variances and adaptive kalman filtering," *IEEE Transactions on Automatic Control*, vol. 15, no. 2, pp. 175–184, 1970.

[21] G. Golub and C. Van Loan, *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, 2013.

[22] L. Ljung, *System Identification: Theory for the User*. Prentice Hall information and system sciences series, Prentice Hall PTR, 1999.

[23] N. Sandgren, P. Stoica, and P. Babu, "On moving average parameter estimation," in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pp. 2348–2351, 2012.

[24] J. Durbin, "Efficient estimation of parameters in moving-average models," *Biometrika*, vol. 46, no. 3/4, pp. 306–316, 1959.

[25] G. Jenkins and D. Watts, *Spectral Analysis and Its Applications*. Holden-Day series in time series analysis and digital signal processing, Holden-Day, 1969.

[26] J. A. Gubner, *Probability and Random Processes for Electrical and Computer Engineers*. Cambridge University Press, 2006.

[27] G. Box, G. Jenkins, G. Reinsel, and G. Ljung, *Time Series Analysis: Forecasting and Control*. Wiley Series in Probability and Statistics, Wiley, 2015.

[28] R. Davidson, C. Davidson, J. MacKinnon, and E. MacKinnon, *Econometric Theory and Methods*. Oxford University Press, 2004.

[29] J. McLaughlin and J. Raviv, "Nth-order autocorrelations in pattern recognition," *Information and Control*, vol. 12, no. 2, pp. 121–142, 1968.

[30] D. Chazan and B. Weiss, "Higher order autocorrelation functions as translation invariants," *Information and Control*, vol. 16, no. 4, pp. 378–383, 1970.

[31] C. Fabris and B. Kovatchev, *Glucose Monitoring Devices: Measuring Blood Glucose to Manage and Control Diabetes*. Elsevier Science, 2020.

[32] A. Marconato, M. Schoukens, and J. Schoukens, "Filter-based regularisation for impulse response modelling," *IET Control Theory & Applications*, vol. 11, 10 2016.

[33] T. Chen, H. Ohlsson, and L. Ljung, "On the estimation of transfer functions, regularizations and Gaussian processes—Revisited," *Automatica*, vol. 48, no. 8, pp. 1525–1535, 2012.

[34] J. Sherman and W. J. Morrison, "Adjustment of an Inverse Matrix Corresponding to a Change in One Element of a Given Matrix," *The Annals of Mathematical Statistics*, vol. 21, no. 1, pp. 124 – 127, 1950.

[35] S. Mehmood, I. Ahmad, H. Arif, U. E. Ammara, and A. Majeed, "Artificial pancreas control strategies used for type 1 diabetes control and treatment: A comprehensive analysis," *Applied System Innovation*, vol. 3, no. 3, 2020.

[36] J. Tasic, M. Takacs, and L. Kovacs, "Control engineering methods for blood glucose levels regulation," *ACTA POLYTECHNICA HUNGARICA*, vol. 19, no. 7, pp. 127–152, 2022.

[37] S. J. Moon, I. Jung, and C.-Y. Park, "Current advances of artificial pancreas systems: A comprehensive review of the clinical evidence," *DIABETES & METABOLISM JOURNAL*, vol. 45, pp. 813–839, NOV 2021.

Author's Publications

[Auth1] Martin Dodek, Eva Miklovičová, and Marián Tárník. Implementation of selected control theory algorithms for embedded real-time systems. In R. Paulen and M. Fikar, editors, *2021 23rd International Conference on Process Control (PC)*, pages 19–24, June 2021.

[Auth2] Martin Dodek and Eva Miklovičová. Improvements for BLDC motor control. In Alena Kozáková, editor, *ELITECH´21 : 23th Conference of Doctoral Students*. Spektrum STU, May 2021.

[Auth3] Martin Dodek and Eva Miklovičová. Physiology-compliant empirical model for glycemia prediction. *International Review of Automatic Control (IREACO)*, 14(6), 2021.

[Auth4] Martin Dodek and Eva Miklovičová. Maximizing performance of linear model predictive control of glycemia for T1DM subjects. *Archives of Control Sciences*, vol. 32(No 2):305–333, 2022.

[Auth5] Martin Dodek and Eva Miklovičová. Optimal state estimation for the artificial pancreas. In *2022 23rd International Carpathian Control Conference (ICCC)*, pages 88–93, 2022.

[Auth6] Martin Dodek, Eva Miklovičová, and Marián Tárník. Correlation method for identification of a nonparametric model of type 1 diabetes. *IEEE Access*, 10:106369–106385, 2022.

[Auth7] Martin Dodek and Eva Miklovičová. Estimation of process noise variances from the measured output sequence with application to the empirical model of type 1 diabetes. *Biomedical Signal Processing and Control*, 84:104773, 2023.

[Auth8] Zuzana Vitková, Martin Dodek, Jarmila Pavlovičová, and Anton Vitko. Using a state-bounding observer to predict the guaranteed limits of drug amounts in rats after oral administration based on an uncertain pharmacokinetic model. *Pharmaceutics*, 14(4), 2022.

[Auth9] Marián Tárník, Martin Ernek, Martin Dodek, Eva Miklovičová, Adrián Ilka, and Tomáš Murgaš. Identification of generator active power oscillations stability measure. In *2022 29th International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 1–4, 2022.

[Auth10] Martin Dodek and Eva Miklovičová. Predicting the output error of the suboptimal state estimator to improve the performance of the MPC-based artificial pancreas. *Control Theory and Technology*, May 2023.