Subspace Identification of a Glucose-Insulin model Using Meal Tracer Protocol Measurements

Subspace Identification of a Glucose-Insulin model Using Meal Tracer Protocol Measurements

Ali Al-Matouq[1]; Mohammed AlShahrani[2]; Carlo Novara[3]

Abstract— **In this study, the problem of identifying a low complexity state space model describing glucose and insulin dynamics from low sample meal tracer experiments is investigated. Triple tracer meal protocol measurements (sampled as low as 15 samples per meal) together with continuous glucose monitoring measurements, measured concurrently at a rate of 5 minutes per sample, are used. A new formulation to estimate the missing input and output measurements at such low sample rates is developed. Nuclear norm minimization is used to exploit low rankness of the stacked input and output matrix, while the ℓ_1 norm is used to exploit an available sparse basis for the glucose flux profiles. Simulation results, using the UVa Padova simulator, show that the technique outperforms previous methods and also demonstrate the possibility of identifying state space models from triple tracer measurements with good prediction performance under non-ideal conditions.**

I. INTRODUCTION

Glucose metabolism during meals is influenced by many factors, including meal composition, rate of gastric emptying, insulin/glucagon secretion, insulin sensitivity, health condition, activity level and other factors as discussed for example in [2] and [3]. Sophisticated clinical measurements are used to study the contribution of these different factors on glucose dynamics. Two commonly used techniques are the dual tracer meal protocol developed by Steele et. al. in [4] and the triple tracer meal protocol developed by Basu et.al in [5] for measuring glucose fluxes which use tracers consumed with a certain meal and/or administered to the patient.

In the dual tracer technique, for example, the patient consumes glucose that is labeled with one tracer while another tracer is infused intravenously at a constant rate. Three glucose fluxes are then reconstructed from concentration measurements of both the traced and untraced glucose assuming a certain model structure. The glucose fluxes are 1- Glucose rate of appearance from the intestine (GRA), its signal denoted by $u_{ra}(t)$, which measures the rate of glucose appearing from the intestine due to consumed meals; 2- Endogenous glucose production (EGP), denoted by $u_{egp}(t)$, which measures the rate of glucose produced by the liver and 3- Glucose disappearance due to insulin (U), denoted by $u_{ins}(t)$, which measures the rate of glucose absorption by muscles and adipose tissue due to insulin activation. A similar approach is used in the triple tracer meal protocol developed in [5] that uses two tracers infused intravenously and a the third tracer consumed orally with the meal. Both techniques require frequent blood sampling to measure the traced and untraced glucose and to sufficiently infer postprandial glucose dynamics. However, the frequency of blood sampling need to be limited to avoid taking large blood volumes from the patient. As a result, dual tracer and triple tracer meal protocol measurements contain valuable information about the patient that could be exploited in control and estimation but are characterized by low, and possibly irregular, sample rate measurements.

Methods to identify a dynamic model using meal tracer protocol measurements use a two stage approach that seeks to first estimate the missing measurements followed by parameter identification techniques on both the measured and estimated signals. In [6], the triple tracer data set obtained in [7] was used to identify a first principle transport model known as the UVa Padova model. Each patient data set contained 24 sample measurements for each flux type and additional measurements for plasma glucose and plasma insulin concentration measurements spanning 7 hours. The missing measurements for glucose fluxes were reconstructed using deconvolution as described in [8] while the forcing function strategy was used on a subsystem level to identify model parameters [6]. These approaches, however, are subject to both structural and parameter uncertainties since they presume a certain model structure and estimate the missing measurements and model parameters in two isolated steps. Not mentioning the non-convex formulations used for estimating model parameters which introduces non-unique and sub-optimal solutions. Other techniques for identifying the UVa Padova model use continuous plasma glucose and plasma insulin concentration measurements, as studied in [9], but also suffer from the same issues and require considerable number of measurements.

In the case of tracer meal protocol measurements, more than 80% of measurements are missing. To overcome this problem, additional a priori information relevant to the shape and pattern of glucose fluxes is introduced in this study. Particularly, sparse encoding of a large set of plausible glucose flux profiles, as explained in [10], is used in order to obtain sparse basis vectors that for representing the space of plausible glucose flux profiles during meals. The approach then uses a rank heuristic, as in [1], and an additional ℓ_1 norm penalty. The nuclear norm penalty is used to exploit low rankness of the stacked input and output matrix for estimating the missing measurements while the ℓ_1 norm penalty exploits the constructed sparse basis for reconstructing the glucose flux profiles that follow certain patterns. The

[1]Ali Al-Matouq is with the faculty of Engineering Management of Prince Sultan University; PO box 66863 Rafha Street, Riyadh 11586, Saudi Arabia almatouq@psu.edu.sa

[2]Mohammed Alshahrani is with the Department of Mathematics, King Fahd University of Petroleum and Minerals

[3]C. Novara is with the Politecnico di Torino, Turin 10129, Italy.

estimated measurements are then used within the numerical algorithm for subspace identification N4SID [11] to identify model parameters. The identified models are then validated in simulation using the FDA approved UVA Padova simulator.

Comparing with previous techniques in the literature for identifying dynamic models from tracer experiments we mention the following: First, the method yields a computationally tractable solution since a convex formulation is used for estimating the missing input and output measurements followed by the N4SID algorithm that uses standard linear algebra operations. Second, the method does not assume a certain model structure rather a general linear state space model structure is used which helps to reduce structural uncertainties. Third, the missing triple tracer measurements are estimated with the objective of obtaining a low dimensional range space of the observability matrix as compared to the previous two stage techniques that estimate the missing measurements without any account for the final identification step. The available sparse basis also helps to improve the technique developed in [1] for estimating missing measurements using the nuclear norm by reducing the requirement of having at least 50% of the measurements available. Our simulation results show that the technique outperforms the method in [1] and also demonstrate the possibility of identifying state space models from triple tracer measurements with good prediction performance under non-ideal conditions.

The following is the outline of this study. Section 2 will give the problem formulation used to identify the glucose fluxes and subspace model along with the necessary assumptions. Section 3 will present the simulation results for the identification problem. The following notation is used in this study: $\mathbb{R}$ represents the set of real numbers; $A \in \mathbb{R}^{n\times m}$ is an $n\times m$ matrix with real values; $\|z\|_{l_i}$ is the ℓ_i norm of vector z while $|z|$ is the number of non-zero elements in z. For matrix A, $\|A\|_2$ represents the spectral norm of the matrix; $\|A\|_F$ represents the Frobenius norm and $\|A\|_* = \sum_{i=1}^{\min\{n,m\}} \sigma_i(A)$ the sum of singular values of A.

II. Problem Formulation

The objective of this study is to identify a linear state space model describing glucose/insulin dynamics using low sampled triple tracer meal protocol measurements; regularly sampled continuous glucose monitoring measurements; insulin infusion recordings and meal carbohydrates. The method also seeks to estimate the missing triple tracer measurements simultaneously. We first give the required assumptions.

Assumption 2.1: The availability of subcutaneous glucose concentration measurements, measured possibly using a continuous glucose monitor, denoted by $g_{sc}(k)$ (mg/dL) where $k = 0,\cdots,N-1$ is a time index and N is total number of samples taken by the measurement device for the entire experiment. Furthermore, we assume that $g_{sc}(k)$ measurements are perturbed by zero mean finite variance white noise and that the sample rate for these measurements is given by t_{samp}.

Assumption 2.2: The availability of glucose flux measurements measured possibly using the dual tracer meal protocol [4] or triple tracer meal protocol [5] in mg/kg min i.e. *mg* per *kg* of patient body weight per minute. The glucose flux measurements include samples of glucose rate of appearance from the intestine $u_{ra}(k)$, samples of endogenous glucose production $u_{egp}(k)$ and samples of insulin dependent glucose utilization $u_{ins}(k)$ for $k \in \mathscr{T} \subset \{0,1,\cdots,N\}$. Furthermore, we assume that these measurements are perturbed by zero mean iid white noise and can be either regularly or irregularly sampled.

Assumption 2.3: The availability of samples of plasma glucose and insulin concentration measurements denoted by $g(k)$ (mg/dL) and i (pmol/L), respectively for $k \in \mathscr{T}$. Furthermore, we assume that these measurements are perturbed by zero mean finite variance white noise.

Assumption 2.4: The glucose and insulin dynamics for the patient can be described by a discrete-time multi-input multi-output (MIMO) state space model given by:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k \\ y_k &= Cx_k + Du_k + e_k, \quad k = 0,\cdots,N-1 \end{aligned} \tag{1}$$

where, $x_k \in \mathbb{R}^{n_x}$ is the state sequence, $y_k = [y_k^p,\ y_k^{sc}]^T$ is the system output vector sequence that contains the output measurements $y_k^p = [g(k)\ \ i(k)]^T$ and $y_k^{sc} = g_{sc}(k)$. The sequence $u_k = [u_{ra}(kt_s)\ \ u_{egp}(kt_s)\ \ u_{ins}(kt_s)]^T$ is the system input sequence that contains the glucose flux sequences and $e_k \in \mathbb{R}^3$ is an iid zero mean white random noise sequence vector that is uncorrelated with u_k and x_k.

Assumption 2.5: The availability of (1) an estimate of the time and amount of meal carbohydrates consumed by the patient denoted by $c_{carbs}(k)$ (mg) and (2) recordings of the amount and time of bolus and basal insulin administrated to the patient denoted by $u_{iir}(k)$ (IU/hr); i.e. insulin units per hour for $k = 0,\cdots,N-1$.

Assumption 2.6: The matrix pair (A,C) is observable and the matrix pair (A,B) is reachable.

For subspace identification with missing inputs and outputs, the identification problem is posed as follows: given the subset of the noisy input and output sequences; i.e. u_k and y_k for $k \in \mathscr{T} \subset \{0,1,\cdots,N-1\}$ and assuming the system to be retrieved is a linear time invariant system of the form given in (1), we desire to find (a) an estimate of the state dimension of the system n_x; (b) a filtered estimate of the given input and output sequences $\hat{u}_k$ and $\hat{y}_k$ and state sequence $\hat{x}_k$ for all $k = 0,1,\cdots,N-1$ and (c) an estimate of the system matrices, A,B,C and D.

Towards that goal, we first write (1) as a matrix equation as in [12] as follows:

$$Y_{s,N} = \Theta_s X_{1,N} + T_{s,N} U_{s,N} + E_{s,N} \tag{2}$$

where, $Y_{s,N}$ and $X_{1,N}$ are constructed from the vector sequence y_k and x_k for $k = 0,\cdots,N-1$ as follows:

$$Y_{s,N} = \begin{bmatrix} y_0 & y_1 & \cdots & y_{N-s+1} \\ y_1 & y_2 & \cdots & y_{N-s+2} \\ \vdots & \vdots & \ddots & \vdots \\ y_{s-1} & y_s & \cdots & y_{N-1} \end{bmatrix}$$

$$X_{1,N} = \begin{bmatrix} x_0 & x_1 & \cdots & x_{N-s+1} \end{bmatrix}$$

where, s is the row-block size of the Hankel matrix $Y_{s,N}$. Similarly, Hankel matrices $U_{s,N}$ and $E_{s,N}$ are defined using the sequences u_k and e_k for $k = 0, \cdots, N-1$, respectively. On the other hand, matrices Θ_s and $T_{s,N}$ are defined as:

$$\Theta_s = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{s-1} \end{bmatrix}, T_{s,N} = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{s-2}B & CA^{s-3}B & \cdots & D \end{bmatrix}$$

To introduce the technique, we first rewrite (2) in terms of $\Theta_s X_{1,N}$ as follows:

$$\Theta_s X_{1,N} = \begin{bmatrix} -T_{s,N} & I_{sn_y} \end{bmatrix} \begin{bmatrix} U_{s,N} \\ Y_{s,N} - E_{s,N} \end{bmatrix} \tag{3}$$

where, I_{sn_y} is an identity matrix of dimension $sn_y \times sn_y$. From the above equation, we note the following:

1) If $s > n_x$, then rank of $\Theta_s X_{1,N}$ is n_x since Θ_s is full column by assumption 2.4 and $rank(X_{1,N}) = n_x$.
2) Since the matrix $\begin{bmatrix} -T_{s,N} & I_{sn_y} \end{bmatrix}$ is full row rank, then from Sylvester's inequality we can show that:

$$n_x \leq rank(\Phi) \leq n_x + sn_u \tag{4}$$

where,

$$\Phi(U_{s,N}, Y^*_{s,N}) = \begin{bmatrix} U_{s,N} \\ Y^*_{s,N} \end{bmatrix}$$

and $Y^*_{s,N} = Y_{s,N} - E_{s,N}$ is the noise-free output.

Hence, assuming that $rank(U_{s,N}) = sn_u$ is fixed and the state dimension n_x is small then the minimization of $rank(\Phi(U_{s,N}, Y^*_{s,N}))$ can be exploited for finding n_x. This was proposed in [1] through the solution of the following robust matrix completion problem:

$$\begin{aligned} &\min_{\tilde{U}_{s,N}, \tilde{Y}_{s,N} \in \mathscr{H}_s} \|\Phi(\tilde{U}_{s,N}, \tilde{Y}_{s,N})\|_* \\ &+ \lambda \mathscr{L}(\tilde{U}_{s,N}, \tilde{Y}_{s,N}, \{u_k\}_{k \in \mathscr{T}}, \{y_k\}_{k \in \mathscr{T}}) \end{aligned} \tag{5}$$

where the matrices $\tilde{U}_{s,N}$ and $\tilde{Y}_{s,N}$ are Hankel matrices formed using the decision variable vectors $\tilde{u}_0, \cdots, \tilde{u}_{N-1}$ and $\tilde{y}_0, \cdots, \tilde{y}_{N-1}$, respectively and $\mathscr{H}_s$ is the set of Hankel matrices of block size s. The first term is the nuclear norm defined as $\|\Phi\|_* = \sum_{i=1}^{s(n_u+n_y)} \sigma_i(\Phi)$ which is the sum of the singular values of the matrix Φ. This penalty is a convex relaxation of the rank function as discussed in [13] and its minimization can result into finding the minimum rank solution if certain conditions are satisfied [14]. The additional penalty $\mathscr{L}(\cdot)$ is a least squares objective that minimizes the sum of square residual errors between the available measurements and decision variables. The above convex program is commonly referred to as a robust matrix completion problem that can be used to estimate the missing elements of the matrix Φ from available noisy measurements of the matrix.

The study in [1], however, showed that the above minimization problem can estimate the missing elements of the matrix Φ for cases when the percentage of missing input and output measurements is less than 50%. In the case of tracer meal protocol measurements, more than 80% of measurements are missing and hence the above technique will not work unless additional a priori information is available. To overcome this problem, additional a priori information will be included in this study, which is the subject of the following assumption.

Assumption 2.7: The glucose flux profiles defined as:

$$\begin{aligned} U_{ra} &:= [u_{ra}(0), \cdots, u_{ra}((N-1)t_{samp})]^T \\ U_{egp} &:= [u_{egp}(0), \cdots, u_{egp}((N-1)t_{samp})]^T \\ U_{ins} &:= [u_{ins}(0), \cdots, u_{ins}((N-1)t_{samp})]^T \end{aligned}$$

live inside a low dimensional subspace of the space spanned by the positive linear combination of the column vectors of $D_{ra} \in \mathbb{R}^{N \times p_{ra}}$, $D_{egp} \in \mathbb{R}^{N \times p_{egp}}$ and $D_{ins} \in \mathbb{R}^{N \times p_{ins}}$ respectively; i.e:

$$U_{ra} = D_{ra}\alpha_{ra}, \quad U_{egp} = D_{egp}\alpha_{egp}, \quad U_{ins} = D_{ins}\alpha_{ins} \tag{6}$$

where, $\alpha_{ra} \in \mathbb{R}^{p_{ra}}_+$; $\alpha_{egp} \in \mathbb{R}^{p_{egp}}_+$ and $\alpha_{ins} \in \mathbb{R}^{p_{ins}}_+$ are positive sparse vectors. Here, p_{ra}, p_{egp} and p_{ins} are the column dimensions of the matrices D_{ra}, D_{egp} and D_{ins}, respectively.

Assumption 2.7 assumes the availability of a sparse basis for the glucose flux profiles. This assumption was made based on the observation that typical glucose flux profiles during meals follow certain patterns. In [10], sparse basis was constructed using extensive random simulations of the FDA approved UVa Padova simulator together with sparse encoding of these glucose fluxes. Consequently, we may express the Hankel matrix $U_{s,N}$ as follows:

$$U_{s,N} = D_{s,N} \otimes \alpha \tag{7}$$

where,

$$D_{s,N} = \begin{bmatrix} D_0 & D_1 & \cdots & D_{N-s+1} \\ D_1 & D_2 & \cdots & D_{N-s} \\ \vdots & \vdots & \ddots & \vdots \\ D_{s-1} & D_s & \cdots & D_{N-1} \end{bmatrix}$$

$$D_i := \begin{bmatrix} d^T_{ra,i} & 0 & 0 \\ 0 & d^T_{egp,i} & 0 \\ 0 & 0 & d^T_{ins,i} \end{bmatrix}, \text{ for } i = 0, 1, \cdots, N-1$$

$$\alpha := \begin{bmatrix} \alpha_{ra}, & \alpha_{egp}, & \alpha_{ins} \end{bmatrix}^T$$

where, $d^T_{ra,i}$ is the $i+1$ row of D_{ra}. Similarly, the row vectors $d^T_{egp,i}$ and $d^T_{ins,i}$ for $i = 0, \cdots, N-1$ correspond to the $i+1$ row of D_{egp} and D_{ins}, respectively.

As discussed in [10], the area under the curve of $u_{ra}(t)$ represents the amount of glucose absorbed from meal carbohydrates according to the following [15]:

$$f \sum_{k=1}^{N} c_{carbs}(k) \approx c_{bw} \sum_{k=0}^{N} u_{ra}(kT_s)T_s \tag{8}$$

where f is the fraction of meal carbohydrates absorbed as glucose in plasma (glucose bioavailability), c_{bw}(kg) the patient body weight.

We describe the measurement equation as:

$$\mathbf{y} = S\mathbf{v}^* + \boldsymbol{\varepsilon} \tag{9}$$

where, vector $\mathbf{y} \in \mathbb{R}^{N_T}$, $N_T = 2N_s + N$, is a vector containing all available measurements which is defined as follows:

$$\mathbf{y} = \left[u_{k_1}, u_{k_2}, \cdots u_{k_{N_s}}, y^p_{k_1}, y^p_{k_2}, \cdots y^p_{k_{N_s}}, y^{sc}_0, y^{sc}_2, \cdots, y^{sc}_{N-1}\right]^T$$

Here we used subscripts $k_1, k_2, \cdots, k_{N_s} \in \mathscr{T}$ to represent the time indicies of the available tracer measurements. The vector $\mathbf{v}^* \in \mathbb{R}^{N(n_u+n_y)}$ represents the non-repeated true elements of the matrix true matrix Φ^* defined as follows:

$$\Phi^*_{i,j} = \mathbf{v}^*_{i+j-1} \text{ for } i = 1,2,\cdots,s,\ j = 1,2,\cdots,N-s+1,$$
$$\Phi^*_{i,j} = \mathbf{v}^*_{i+j+N-s-1} \text{ for } i = s+1,\cdots,2s,\ j = 1,2,\cdots N-s+1$$

Finally, S is a selection matrix for connecting the decision variables to the available measurements in $\mathbf{y}$.

Consequently, we propose the following minimization problem that incorporates the above additional a priori information for estimating the missing input and output:

$$\begin{aligned} &\min_{\alpha \succeq 0, \mathbf{v}} \quad \|\mathbf{y} - S\mathbf{v}\|_2^2 + \lambda\left(\|\Phi\|_* + \|\alpha\|_1\right) \\ &\text{subject to} \\ &[I_{sn_u} \quad \mathbf{0}]\Phi = D_{s,N} \otimes \alpha, \\ &c_{bw} \sum_{k=0}^{N} d^T_{ra,k} \alpha_{ra} T_s \leq f \sum_{k=0}^{N} c_{carbs}(k) \end{aligned} \tag{10}$$

The above two constraints incorporate the additional a priori knowledge about $U_{s,N}$ and meal carbohydrates, where $d_{ra,k}$ represents the kth row of D_{ra}.

The tuning parameter λ can be used to set the level of trade-off between fitness with measurements and both low rankness of Φ and sparsity level of α. Here, one tuning parameter λ is used since the two penalties can be described as one nuclear norm penalty.

After solving (10), the MOESP algorithm (Multivariable Output-Error State space) identification technique, described for example in [16], can be used to estimate the system order and system matrices $\hat{A}, \hat{B}, \hat{C}$ and $\hat{D}$ using RQ factorization of the estimated matrix $\hat{\Phi}$. The above is a convex semi-definite minimization problem that can be solved using standard convex programming algorithms [17].

The value of s, the block size of the Hankel matrices, should be selected such that $s > \tilde{n}_x$, where $\tilde{n}_x$ is an estimate of the order of the system. A suitable value for the tuning parameter λ can be found by repeated solution of (10) over a logarithmically spaced set of values for λ until an appropriate model is found in terms of feasibility of the estimated missing measurements and minimum validation fit, which is defined as:

$$fit = 100\left(1 - \frac{\|\mathbf{y}^{pred} - \mathbf{y}\|_2}{\|\mathbf{y} - mean(\mathbf{y})\|_2}\right) \tag{11}$$

where, $\mathbf{y}^{pred}$ is the predicted output of the model obtained from the two step method (solving (10) and then implementing MOESP algorithm on the estimated input/output signals) and $mean(\mathbf{y})$ is the average value of the measurement vector.

III. Simulation Study

In the following we will demonstrate the potential of the new technique through two simulation experiments using the FDA approved UVa Padova model developed in [6].

To generate the required measurements for testing, we simulated the average adolescent, average adult and average child in-silico patients using the UVa Padova simulator (version 3.2) with the meal scenario explained in [10] (not shown here for brevity). For the sake of testing the robustness of the technique we first subjected the simulated glucose flux profiles to random nonlinear distortions. The nonlinear distortion was made by raising each element $u_{ra}(k)$, $u_{egp}(k)$ and $u_{ins}(k)$ generated in simulation by a random exponent that is uniformally distributed between 0.5 and 1.0. We then repeated the simulations using the deformed flux profiles as inputs in order to generate the required measurements for testing. We then perturbed the generated signals by zero mean white noise with variance of 0.01. The noise in $g_{sc}(k)$ was modeled as an SU Johnson distributed signal with an autoregressive component as explained in [18] and [19].

A. Experiment I: Comparison with the method in [1]

In this experiment, the measurements included 97 samples for noisy continuous glucose monitoring measurements $g_{sc}(k)$, 15 linearly spaced samples for the distorted glucose flux signals $u_{ra}(k)$, $u_{egp}(k)$ and $u_{ins}(k)$ and 15 linearly spaced samples for plasma insulin $i(k)$ and plasma glucose $g(k)$ signals. We then used the method developed in [1] by solving (5) for the purpose of estimating the remaining 82 measurements for $u_{ra}(k)$, $u_{egp}(k)$, $u_{ins}(k)$, $i(k)$ and $g(k)$ and to filter the available measurements. The term $\mathscr{L}$ was set as the least squares objective (ℓ_2 norm error) and with $\lambda = 0.03$. This minimization was done using CVX in Matlab (version 2.0) [20]. The noisy measurements vs. the estimated measurements for the average child patient are shown in Figure 1 top view. The relative root mean square error values $RRMSE$, as defined in [10], for this experiment are shown in Table I. Clearly the results indicate poor estimation performance for the glucose fluxes.

We then repeated the same experiment but using the technique developed in this study. The other parameters in (10) were set as follows: $s = 12$ and $\lambda = 0.03$. The sparse dictionaries D_{ra}, D_{egp}, D_{gu} were developed as explained in [10] and were used. The results are shown in Figure 1 lower row for the average child patient and the corresponding relative root mean square errors are shown in Table I. Clearly, the results demonstrate the potential of the technique in estimating the missing input and output measurements even for the measurements that do not have a sparse basis for them (i.e. g_{sc}, $g(k)$ and $i(k)$).

B. Experiment II: Subspace Identification using two meals

For the purpose of identifying a state space model from triple tracer experiments we first obtained the missing input and output measurements as explained in experiment 1 and then used the N4SID algorithm developed in [11] to identify a state space model. The measurements were first

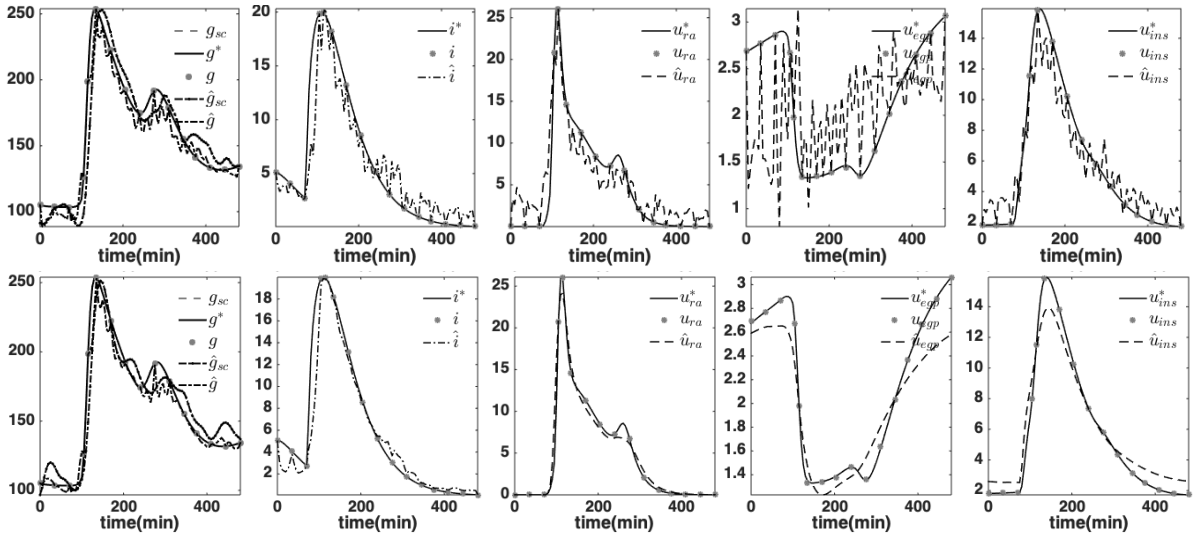

Fig. 1. Top row: (from left) the true signals g^*, i^* u^*_{ra}, u^*_{egp}, u^*_{ins} for the average adolescent patient vs. the noisy signals g, i u_{ra}, u_{egp}, u_{ins} and estimated signals $\hat{g}$, $\hat{i}$, $\hat{u}_{ra}$, $\hat{u}_{egp}$ and $\hat{u}_{ins}$ obtained using the method in [1] by solving (5). Bottom row, the same but with signals obtained by using the method in this study by solving (10).

Patient/ *RRMSE*	$\hat{u}_{ra}(old)$	$\hat{u}_{ra}(new)$	$\hat{u}_{egp}(old)$	$\hat{u}_{egp}(new)$	$\hat{u}_{ins}(old)$	$\hat{u}_{ins}(new)$
average adolescent	0.0907	0.0329	0.0973	0.0223	0.3851	0.2150
average adult	0.1130	0.0300	0.1049	0.0279	0.4281	0.2513
average child	0.0908	0.0386	0.0741	0.0685	0.3602	0.1800

TABLE I

RELATIVE ROOT MEAN SQUARE ERROR FOR ESTIMATED VALUES FOR GLUCOSE FLUXES USING THE METHOD IN [1] DENOTED BY $\hat{u}_{ra}(old)$, $\hat{u}_{egp}(old)$, $\hat{u}_{ins}(old)$ AND THE CORRESPONDING VALUES USING THE PROPOSED METHOD IN THIS STUDY DENOTED BY $\hat{u}_{ra}(new)$, $\hat{u}_{egp}(new)$, $\hat{u}_{ins}(new)$

Patient/*RRMSE*	$\hat{u}_{ra}(new)$	$\hat{u}_{egp}(new)$	$\hat{u}_{ins}(new)$	$\hat{g}(new)$
average adolescent	0.0465	0.0501	0.2393	0.0546
average adult	0.0262	0.0169	0.3017	0.0573
average child	0.0417	0.0540	0.1848	0.0597

TABLE II

RELATIVE ROOT MEAN SQUARE ERROR PERFORMANCE FOR ESTIMATED VALUES FOR EXPERIMENT II

Pat.	FPE	MSE	$\hat{n}_x$	fit_{val}
avg. adol.	13.52	13.91	8	(82.3 86.1,66.9)
avg. adult	3.894	7.3	8	(83.0,85.6,47.5)
avg. child	26.9	20.1	8	(67.6, 58.5, 43.0)

TABLE III

IDENTIFICATION PERFORMANCE

obtained using simulation similar to the approach discussed in experiment 1 above but extended for two consecutive meals with random carbohydrate content.

Shifted versions of the sparse dictionaries D_{ra}, D_{egp}, D_{gu} were used (one for each meal) using the technique discussed in [10]. Figure 2 shows the measurements used for this experiment and the estimated input and output measurements from the solution of (10) for the average adolescent patient. Table II shows the corresponding relative root mean square error values for the estimated fluxes and plasma glucose signals for the average patients.

The Matlab function ('N4SID') was then used on the estimated input and output measurements with the option of finding a state space model with input disturbance and no feed through. Table III gives the Akaike finite prediction error and the mean square error for the identified models for the average adolescent, average adult and average child patient, respectively.

Validation was conducted by testing the identified models using new simulation data generated using the UVa Padova model for 3 consecutive different meal scenarios for the same 3 in-silico average patients. The matlab function "compare" was used to generate the plots in Fig. 3 for the average adolescent patient. The fitness parameter in (11) was calculated using the true simulated values for the three outputs. The results are tabulated in Table III also under fit_{val} for the three output signals $\hat{g}_{sc}(k)$, $\hat{g}$, $\hat{i}$, respectively. The results indicate overall good performance and fitness with estimated data except for the prediction of plasma insulin for the child and adult patient. This could be improved by introducing more input excitation or using more experimental data.

IV. DISCUSSION AND CONCLUSION

A new technique for the reconstruction of missing input and output measurements in triple tracer experiments for subsequent subspace identification is proposed. The approach uses a machine learning prior in the form of a sparse basis for the glucose fluxes constructed using dictionary

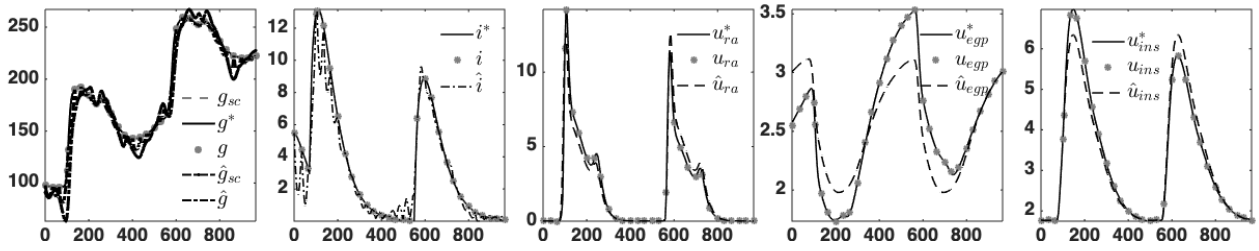

Fig. 2. The true signals g^*, i^* u^*_{ra}, u^*_{egp}, u^*_{ins} for the average adolescent patient vs. the noisy signals g, i u_{ra}, u_{egp}, u_{ins} and estimated signals $\hat{g}$, $\hat{i}$, $\hat{u}_{ra}$, $\hat{u}_{egp}$ and $\hat{u}_{ins}$ obtained from solving (10)

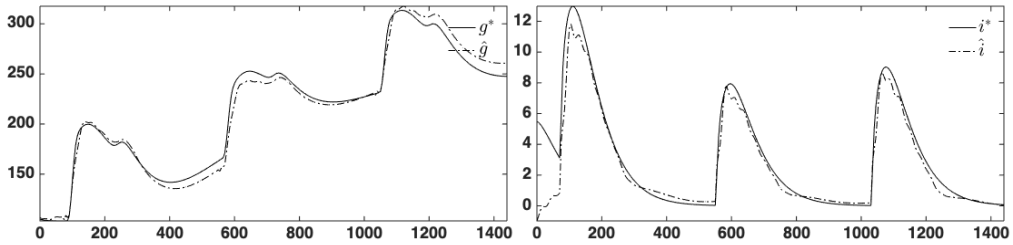

Fig. 3. Results obtained from Experiment II for the average adolescent patient using N4SID algorithm [11]

learning techniques on a large database of simulated fluxes. The nuclear norm is then used to exploit low rankness of the stacked input and output matrix, while the ℓ_1 norm is used to exploit the available sparse basis for typical glucose flux profiles during meals. Simulation results show that the technique outperforms the method in [1] and also demonstrate the possibility of identifying state space models from triple tracer measurements with good performance.

ACKNOWLEDGMENT

The first author would like to acknowledge the monetary support of Prince Sultan University for presenting this paper.

REFERENCES

[1] Z. Liu, A. Hansson, and L. Vandenberghe, "Nuclear norm system identification with missing inputs and outputs," *Systems & Control Letters*, vol. 62, no. 8, pp. 605–612, 2013.

[2] Y. C. Kudva, R. E. Carter, C. Cobelli, R. Basu, and A. Basu, "Closed-loop artificial pancreas systems: physiological input to enhance next-generation devices," *Diabetes care*, vol. 37, no. 5, pp. 1184–1190, 2014.

[3] F. J. Doyle, L. M. Huyett, J. B. Lee, H. C. Zisser, and E. Dassau, "Closed-loop artificial pancreas systems: engineering the algorithms," *Diabetes care*, vol. 37, no. 5, pp. 1191–1197, 2014.

[4] R. Steele, C. Bjerknes, I. Rathgeb, and N. Altszuler, "Glucose uptake and production during the oral glucose tolerance test," *Diabetes*, vol. 17, no. 7, pp. 415–421, 1968.

[5] R. Basu, B. Di Camillo, G. Toffolo, A. Basu, P. Shah, A. Vella, R. Rizza, and C. Cobelli, "Use of a novel triple-tracer approach to assess postprandial glucose metabolism," *American Journal of Physiology-Endocrinology And Metabolism*, vol. 284, no. 1, pp. E55–E69, 2003.

[6] C. D. 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.

[7] R. Basu, C. Dalla Man, M. Campioni, A. Basu, G. Klee, G. Jenkins, G. Toffolo, C. Cobelli, and R. Rizza, "Mechanisms of postprandial hyperglycemia in elderly men and women: Gender specific differences in insulin secretion and action," *Diabetes*, vol. 55, pp. 2001–2014, 2006.

[8] G. Pillonetto, G. Sparacino, and C. Cobelli, "Reconstructing insulin secretion rate after a glucose stimulus by an improved stochastic deconvolution method," *IEEE transactions on biomedical engineering*, vol. 48, no. 11, pp. 1352–1354, 2001.

[9] R. Visentin, C. Dalla Man, and C. Cobelli, "One-day bayesian cloning of type 1 diabetes subjects: toward a single-day uva/padova type 1 diabetes simulator," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 11, pp. 2416–2424, 2016.

[10] A. A. Al-Matouq, T.-M. Laleg-Kirati, C. Novara, I. Rabbone, and T. Vincent, "Sparse reconstruction of glucose fluxes using continuous glucose monitors," *IEEE/ACM transactions on computational biology and bioinformatics*, 2019.

[11] P. van Overschee and B. de Moor, *Subspace identification for linear systems: theory, implementation, applications*. Kluwer Academic Publishers, 1996, no. v. 1. [Online]. Available: http://books.google.com/books?id=c0vvAAAAMAAJ

[12] M. Verhaegen and A. Hansson, "N2sid: Nuclear norm subspace identification of innovation models," *Automatica*, vol. 72, pp. 57–63, 2016.

[13] M. Fazel, T. K. Pong, D. Sun, and P. Tseng, "Hankel matrix rank minimization with applications to system identification and realization," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 3, pp. 946–977, 2013.

[14] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM review*, vol. 52, no. 3, pp. 471–501, 2010.

[15] C. Della Man, A. Caumo, and C. Cobelli, "The oral glucose minimal model: estimation of insulin sensitivity from a meal test," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 5, pp. 419–429, 2002.

[16] M. Verhaegen and V. Verdult, *Filtering and system identification: a least squares approach*. Cambridge university press, 2007.

[17] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge university press, 2004.

[18] M. Breton and B. Kovatchev, "Analysis, modeling, and simulation of the accuracy of continuous glucose sensors," *Journal of Diabetes Science and Technology*, vol. 2, no. 5, pp. 853–862, 2008.

[19] B. P. Kovatchev, M. Breton, C. Dalla Man, and C. Cobelli, "In silico preclinical trials: a proof of concept in closed-loop control of type 1 diabetes," *Journal of diabetes science and technology*, vol. 3, no. 1, pp. 44–55, 2009.

[20] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.0 beta," http://cvxr.com/cvx, Sep. 2012.