Latent Variables Model Based MPC for People with Type 1 Diabetes

Xiaoyu Sun∗ Mudassir Rashid∗∗ Mohammad Reza Askari∗∗ Ali Cinar∗∗

∗ Departments of Biomedical Engineering, Illinois Institute of Technology, Chicago, IL 60616 USA (e-mail: xsun49@hawk.iit.edu).
∗∗ Departments of Chemical and Biological Engineering, Illinois Institute of Technology, Chicago, IL 60616 USA (e-mail: {mrashid3, cinar}@iit.edu, maskari@hawk.iit.edu)

Abstract: A model predictive control (MPC) system based on latent variables (LV) model generated by using partial least squares (PLS) method is developed. The difference in the performance of MPCs that use recursively updated LV models based on autoregressive time series modeling (with exogenous inputs - ARX) and PLS is studied. The effect of signal noise on MPC performance is also investigated for both types of models. MPC performance is evaluated by regulating the blood glucose concentration (BGC) of people with Type 1 diabetes mellitus (T1DM) in simulation studies. Signal noise in glucose concentration sensor data, delays caused by insulin absorption and action, and disturbances caused by consumption of meals make the regulation of BGC difficult. The proposed controller is evaluated with 10 in-silico adult subjects of the UVa/Padova simulator with different levels of signal noise. The results illustrate the effectiveness of the MPC based on LV model. The average time for BGC in the safe range (70-180 mg/dL) for the LV-based MPC is 83.23% compared to 79.68% for the MPC based on ARX model when intravenous BGC values are used. The average time in safe range decreases to 76.04% and 71.92%, respectively, when using the generic CGM sensor of the simulator. It is reduced further to 71.93% and 67.20% when additional noise is added to CGM readings.

Keywords: Model predictive control, latent variables model, measurement noise, type 1 diabetes, artificial pancreas system.

1. INTRODUCTION

Model predictive control (MPC) has become the preferred control system for most processes with nonlinearities and time-varying parameters. The models used in MPC to regulate such systems must provide reliable predictions of system variables in response to setpoint changes and disturbances. Often, the measured variables used for predictions would have signal noise, missing values and outliers. Hence models that are robust to such challenges in data must be developed. One approach for developing such models relies on latent variables computed by partial least squares (also called projection to latent structures) (PLS) techniques. PLS was first used in econometric modeling. Then, it revolutionized chemometrics (Geladi and Kowalski, 1986; Vinzi et al., 2010). The use of principal components analysis (PCA) and PLS in chemical process industries started by the development of multivariable process monitoring and fault diagnosis (Negiz and Cinar, 1997; MacGregor and Cinar, 2012). System identification based on latent variables progressed in the systems engineering community as well (Juricek et al., 2005; Negiz and Cinar, 1997).

Type 1 diabetes mellitus (T1DM) is a chronic disease caused by the autoimmune destruction of beta cells in pancreas, resulting in the inability of pancreas to produce insulin which regulates the transport and utilization of blood glucose (Hajizadeh et al., 2019b). Thus, exogenous insulin administrated by multiple daily injections or infusion by an insulin pump is necessary to maintain the blood glucose concentration (BGC) of people with T1DM within a safe range (70-180 mg/dL) (Yu et al., 2018). The closed-loop control system for automated insulin delivery, known as artificial pancreas (AP) system, is usually consists of a continuous glucose monitoring (CGM) sensor which measures the glucose to infer BGC, an insulin pump that delivers insulin to the subcutaneous tissue of the patient, and a feedback control algorithm that receives the CGM data to compute the optimal insulin amount to be infused by manipulating the infusion rate of the pump (Turksoy et al., 2013). However, delays caused by insulin absorption and action, and measurements of BGC, noise in CGM readings, and disturbances caused by meal consumption and exercise make the determination of the optimal infusion rate of insulin difficult, resulting in swings in BGC that may cause hypoglycemia or hyperglycemia.

Model predictive control (MPC) in AP systems aims to compute the optimal insulin infusion rate that minimize the deviation of glucose concentration values predicted by a predictive model and the desired glucose concentration by using as little insulin as possible. It is considered as a promising control method for managing BGC in both in-silico and clinical studies (Hovorka et al., 2004;
Soru et al., 2012). MPC based on physiological models (Hovorka et al., 2004; De Nicolao et al., 2011) where the glucose dynamics are described by a physiological model are proposed. However, it is hard to identify the model parameters and update the parameters online to capture the variations in BGC dynamics. MPC based on data-driven models use an empirical model where the model parameters can be identified recursively (Eren-Oruklu et al., 2009; Turksoy et al., 2014, 2013) to describe BGC dynamics accurately at any time. In particular, autoregressive (AR) models (Eren-Oruklu et al., 2009), and AR models with exogenous inputs (ARX) (Turksoy et al., 2014, 2013) are used to develop personalized BGC prediction model. The model can capture the glucose dynamics accurately by updating the model parameters recursively. However, the identified model parameters are influenced by the quality of CGM data. The prediction accuracy deteriorates with the presence of sensor noise.

An MPC based on latent variable (LV) model is proposed to handle measurement noise. Statistical methods based on LVs, such as PLS and PCA, have been proven to be powerful tools for data analysis, modeling, and prediction (MacGregor and Cinar, 2012; Zhao et al., 2012). In the empirical models for AP, the process variables, including CGM readings, insulin infusion rate and meal consumption, are treated as the inputs of the model, and the output of the model is the future BGC values. To compute the optimal insulin infusion rate, the identified model is used repeatedly over the prediction horizon, which increases the computational effort.

An alternative approach is to control the key variables (i.e., the BGC) while manipulating the inputs, once the key variables during the manipulation process is well-controlled, the quality after the manipulation can be guaranteed (Flores-Cerrillo and MacGregor, 2005; Golshan et al., 2010). In this work, a MPC aiming at controlling the BGC during the infusion of insulin is designed for people with T1DM. The LV-based MPC is consist of two key steps. First, a PLS model is developed to predict the glucose concentrations (the model output) from available historical CGM readings and variables including estimation of plasma insulin concentration (PIC) and gut absorption rate generated from CGM readings and insulin infusion rates (Hajizadeh et al., 2017) (the model inputs). In the second step, the MPC is formulated based on the PLS model to compute the optimal insulin infusion rate where constraints of PIC and insulin infusion rate are integrated.

In this paper, the LV-based MPC is evaluated with 10 in-silico adult subjects from the UVa/Padova T1DM simulator which is accepted by FDA (Man et al., 2014) with different levels of measurement noise and compared to the performance of a MPC based on ARX model. The average time in the safe range (70-180 mg/dL) and in hypoglycemia and hyperglycemia ranges for LV-based MPC and ARX based MPC are compared and the influence of measurement noise on these metrics are reported.

The rest of this paper is structured as follows. The LV-based MPC is described in Section 3. The rest of this paper is structured as follows. The LV-based MPC is described in Section 3. The second step, the MPC is formulated based on the PLS model to compute the optimal insulin infusion rate where constrains of PIC and insulin infusion rate are integrated. In the paper, the conclusion is reported and discussed in Section 4. Section 5 provides the conclusions.

2. LATENT VARIABLE MODEL

2.1 Partial Least Squares

The LV-based models are widely used in data analysis, monitoring, and prediction (Negiz and Cinar, 1997; MacGregor and Cinar, 2012). A variety of LV-based methods have been developed, with the main difference being how the LVs are generated. PLS (Geladi and Kowalski, 1986; Vinzzi et al., 2010) is a popular LV-based regression method. For the normalized input matrix \( X \in R^{n \times m} \) which is also normalized, where \( n \) is the number of samples, \( m \) and \( l \) are the number of input variables and output variables, respectively. The aim of the PLS method is to find the latent structure between the input matrix \( X \) and output matrix \( Y \) by maximizing the covariance of LV of the input matrix and the output matrix. Thus, the first LV \( t_1 \) can be expressed as

\[
t_1 = X w_1
\]

where \( w_1 \) is the first weight vector that maximizes the objective function

\[
\arg \min_{w_1} (w_1^T X^T Y Y^T X w_1)
\]

s.t. \( w_1^T w_1 = 1 \)

The remaining LVs can be found by repeating the above process after deflating \( X \) and \( Y \) by \( t_1 \)

\[
p_1 = (t_1^T t_1)^{-1} t_1^T X
\]

\[
X = X - t_1 p_1^T
\]

\[
q_1 = (t_1^T t_1)^{-1} t_1^T Y
\]

\[
Y = Y - t_1 q_1^T
\]

where \( p_1 \) and \( q_1 \) are the eigenvector that corresponds to the largest eigenvalue of matrix \( (X^T Y Y^T X) \). The second LV \( t_2 \) can be found by repeating the above process after deflating \( X \) and \( Y \) by \( t_1 \)

\[
X = T P^T + E
\]

\[
Y = T Q^T + F
\]

where \( T = [t_1, t_2, \ldots, t_a] \) is the LV (score) matrix where \( t_i \) is the \( i \)th LV and \( t_j^T \) is the first \( a \) scores of the \( j \)th sample. \( P = [p_1, p_2, \ldots, p_a] \) is the loading matrix for the input matrix \( X \) and \( Q = [q_1, q_2, \ldots, q_a] \) is the loading matrix for the output matrix \( Y \). \( E \) and \( F \) are the residuals of the input matrix \( X \) and output matrix \( Y \), respectively.

2.2 Known Data Regression

In this work, we consider a situation where the variables in the output matrix \( Y \) is part of variables in the input matrix \( X \). Thus, for a new sample \( z \), only part of the variables can
be observed, defined as $z^*$, and the variables that are also presented in the output is not observable, defined as $z^\#$, it is natural to arrange the variables in the new sample as
\[
z^T = [z^T, z^\#T]
\]
(5)

Accordingly, the input data matrix $X$ and loading matrix $P$ can be partitioned into two parts as shown in Fig. 1
\[
X = [X^*, X^\#]
\]
\[
P^T = [P^T, P^\#T]
\]
(6)

If the known part of the new sample is used to estimate the score vector $\tau^T$ with a given PLS model (4), the following relation can be obtained (Nelson et al., 1996, 2006)
\[
\tau^T = z^T \Theta
\]
(7)

where $\Theta = (X^*^T X^*)^{-1} X^*^T T$. The corresponding output can be estimated as
\[
\hat{y}^T = \tau^T Q^T
\]
(8)

3. MPC BASED ON LV MODEL

3.1 Structure of data

In order to apply the LV-based modeling technique, the input and output data sets must be organized in an appropriate manner. In previous work in our group, plasma insulin concentration (PIC) and gut glucose absorption rate are estimated using unscned Kalman filter and have been proven helpful in glucose prediction and designing MPC for the AP (Hajizadeh et al., 2017, 2019b). Thus, the estimation of PIC and gut glucose absorption rate are used as exogenous inputs of the PLS model. The $n$th sample $x_n$ in the input matrix $X$ is arranged as
\[
x_n^T = [u_{O}^T(n), u_{I}^T(n), u_{M}^T(n)]
\]
where $u_{O}^T(n) = [g(n-L+1), \ldots, g(n)]$ contains $L$ measurements of BGC $g$, $u_{I}^T(n) = [I(n-d_I-L+1), \ldots, I(n-d_I)]$ and $u_{M}^T(n) = [M(n-L+1), \ldots, M(n)]$ are consist of estimation of PIC $I$ with the order of delay $d_I$ which is set to be 4 in this work accounting for physiological delay and gut glucose absorption rate $M$, respectively.

Future BGC values are usually used as the model output, and a MPC based on this kind of models computes the optimal operations by minimizing the deviation of future BGCs predicted by the model from the desired value after the insulin is delivered. However, in this work, we focus on BGC along with the infusion of insulin because the BGC will be in range if the BGC is well controlled while the insulin is infused. Thus, the $n$th output $y_n$ in the output matrix $Y$ is
\[
y_n^T = [g(n-l+1), \ldots, g(n)]
\]
where $l(l < L)$ is the number of BGC values predicted, known as prediction horizon.

3.2 Integrating insulin absorption model with LV model

In the AP system, insulin infusion rate is the manipulated variable. Hence, it is necessary to incorporate the insulin absorption model in the LV-based model because the estimates of PIC are used as inputs in the LV-based model. In this work, the model that describes the absorption of subcutaneously administered insulin proposed by Hovorka et al. (2004) is chosen for computing future PIC. The discrete insulin absorption model is
\[
S_{1,k+1} = S_{1,k} + T_s \left( u_{I,k} - \frac{S_{1,k}}{t_{max}} \right)
\]
\[
S_{2,k+2} = S_{2,k+1} + T_s \left( \frac{S_{1,k+1}}{t_{max}} - \frac{S_{2,k+1}}{t_{max}} \right)
\]
\[
I_{k+3} = I_{k+2} + T_s \left( \frac{S_{2,k+2}}{t_{max}V_I} - k_e I_{k+2} \right)
\]
(9)

where $S_1$ and $S_2$ are the two compartments representing absorption of subcutaneously administered insulin and $u_I$ represents the insulin infusion rate (including basal and bolus insulin). $I$ is the PIC with distribution volume $V_I$. $T_s$ and $t_{max}$ are the sampling time and time to maximum of absorption of administered insulin, respectively. Note that $I_{k+3}$ can be calculated if $u_{I,k}$ is known where the delay between insulin infusion and PIC can be observed. Furthermore, there is a delay $d_I$ between CGM and PIC. Thus to predict $l$ future insulin infusion and PIC can be observed. The estimation of future $(l - (d_I + 2))$ insulin infusion rates and PIC values are needed. The estimation of future $(l - (d_I + 2))$ insulin concentrations $\hat{I}$ can be expressed as
\[
\hat{I} = A_I u + C_I
\]
(10)

where $A_I$ is the coefficient matrix, $C_I$ is the constant vector and $u$ represents the $(l - (d_I + 2))$ future insulin infusion rates. By scaling the coefficient matrix and constant vector with means and standard deviations of PIC, the normalized PIC $\bar{I}_n$ which is part of the model input can be calculated:
\[
\bar{I}_n = A_I u + C_I
\]
(11)

The score vector $\hat{\tau}$ of the PLS model estimated from the incomplete input can be calculated by
\[
\hat{\tau} = \Theta_1^T \hat{z}_1 + \Theta_2^T \bar{I}_n
\]
(12)

where $z_1$ contains all the known variables in the new input sample, $\Theta_1$ and $\Theta_2$ are the coefficient matrices that correspond to the known variables and the normalized
future PIC, respectively. And the normalized output \( \hat{y}_n \) given in (8) can be reformulated as
\[
\hat{y}_n = Q\hat{\tau} = Q\Theta_2^T A_1 u + ... \text{the same variables are used as the input where } L = 24 \text{ and the next CGM in the future is used as the model output,}
\]
The CGMs \( \hat{y} \) while insulin is infused can be estimated as following after integrating the means and standard deviations with the coefficient matrix and constant vector
\[
\hat{y} = A_2 u + C_2 \tag{14}
\]
3.3 MPC based on LV model

MPC is widely adopted in designing AP systems due to its ability to incorporate constraints while making a control decision. The purpose of a MPC in the AP system is to compute the optimal insulin infusion rate by minimizing differences between the predicted BGC and the target BGCs under some constraints. From previous works in our group (Hajizadeh et al., 2019b,a), the PIC has been proven important for BGC management. Thus, in this work, the same objective function and constraints for PIC are adopted (Hajizadeh et al., 2019b).

\[
\begin{align*}
\arg \min_u & \quad (\hat{y} - y_p)^T Q_y (\hat{y} - y_p) + I_d^T Q_{pic} I_d + u^T Q_u u \\
\text{s.t.} & \quad \hat{y} = A_1 u + C_1 \\
& \quad I_d = I_p - I \\
& \quad I_{min} \leq I \leq I_{max} \\
& \quad u_{min} \leq u \leq u_{max}
\end{align*} \tag{15}
\]

where \( Q_y \), \( Q_{pic} \), and \( Q_u \) are the weight matrices. Specifically, \( Q_y \), generated from the glycemic risk matrix, is dynamically tuned according to CGM for regulate glucose concentration effectively and \( Q_u \), calculated from the plasma insulin risk and according to the estimation of PIC, aiming at make the controller more aggressive when the PIC is low and less aggressive if there are enough insulin present in the bloodstream. \( y_p \) is the setpoint for CGM and is set to be 110 mg/dL. \( I_p \), \( I_{min} \), and \( I_{max} \) are the desired and constraint boundaries for PIC which are specified for each individual. The glycemic risk index, plasma insulin risk index, and desired and boundaries for PIC are shown in Fig. 2 (Hajizadeh et al., 2019b). \( u_{min} \) and \( u_{max} \) are the constraints on future insulin infusion rates and are zero and 6 U/min, respectively.

The proposed MPC based on LV model updated online is shown in Fig. 3. At each sampling time, the optimal insulin infusion rate is calculated by solving the constrained optimization problem that minimizes the objective function (15) and the first insulin infusion rate of the solution is sent to the insulin pump. At the next sampling time, the PIC and gut glucose absorption rate are estimated, the LV-based model is updated with the new observation and estimations. The optimal insulin infusion rate is recomputed by solving the optimization problem with updated constraints and models. The process is repeated at each sampling time.

For comparison, the historical BGC information, estimation of PIC and gut glucose absorption rate are used as the input of ARX model and the output of the model is the future BGC values. The ARX model where the model parameters are updated using recursive least square (RLS) algorithm is then integrated in the objective function (15) to give predictions of future BGC values.

4. RESULTS AND DISCUSSION

The performance of the proposed MPC based on LV model is tested on the in-silico adult cohort of the UVa/Padova metabolic simulator (Man et al., 2014) and the scenario is listed in Table 1 where the first day is an open-loop experiment for initialization and the following five days are closed-loop operation for testing the proposed MPC algorithm. The results are compared with the MPC based on ARX model. In the simulation study, the number of variables in the input samples \( L \) is 36 for LV model to achieve a good prediction, while for the ARX model, the same variables are used as the input where \( L = 24 \) and the next CGM in the future is used as the model output,
and the insulin absorption model is incorporated as well. The prediction horizon \( l \) in both controllers are set to be 12 (1 hour) for short term glucose regularization. The influence of three different levels of measurement noise is studied. For the noise free case (noise level 0), CGM measurements from IV sensor of the simulator are used in the modeling process. For the second case (noise level 1), the data from generic CGM of the UVa/Padova simulator are used in the modeling process. In the third case (noise level 2), Gaussian noise with zero mean and standard deviation of 1 mg/dL is added to the generic CGM readings. The performance of the controllers is evaluated by the percentage of time of the BGC in different ranges.

<table>
<thead>
<tr>
<th>Meal</th>
<th>Day 1 and 4</th>
<th>Day 2 and 5</th>
<th>Day 3 and 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breakfast</td>
<td>48 g 09:45</td>
<td>50 g 09:10</td>
<td>40 g 09:00</td>
</tr>
<tr>
<td>Lunch</td>
<td>47 g 13:30</td>
<td>70 g 13:45</td>
<td>68 g 14:00</td>
</tr>
<tr>
<td>Dinner</td>
<td>75 g 17:45</td>
<td>65 g 18:00</td>
<td>75 g 18:20</td>
</tr>
<tr>
<td>Snack</td>
<td>31 g 21:30</td>
<td>20 g 22:00</td>
<td>25 g 22:30</td>
</tr>
</tbody>
</table>

Table 2 summarizes the performance of the MPC based on LV-based model and ARX model. The average time for BGC in the safe range (70-180 mg/dL) for the LV-based MPC is 83.23% compared to 79.68% for the MPC based on ARX model where BGC values are measured by the noise free IV sensor. The average time in safe range decreases by 7.19% to 76.04% for LV-based MPC, while average time in safe range decreases by 7.77% to 71.92% for MPC based on ARX model when using the generic CGM sensor of the simulator. When additional noise is added to the generic CGM readings, the average time in the safe range decreases further for both controllers, but the decrease for LV-based MPC controller is smaller. The results indicate that the proposed MPC method based on LV model is more robust when there is noise in CGM readings. Since there is no rescue carbohydrates given throughout the simulation study, few subjects experience hypoglycemia, however, none of them experience severe hypoglycemia (BGC<55 mg/dL). As the level of measurement noise increases, the control performance of both controllers deteriorates as expected, however, the BGC is more tightly controlled for LV-based MPC. Specifically, the minimal value of the BGC (Table 2) is larger for LV-based MPC compared to ARX based MPC, indicating that the proposed MPC method has the ability to provide more reliable control of the BGC to prevent hypoglycemia without much change in maximum BGC when sensor noise is present.

The comparison of BGC variation under LV-based MPC and ARX-based MPC (Fig. 4) indicates that the peaks of BGC after meal is smaller and the BGC is more stable for the LV-based MPC.

Fig. 5 shows the mean and standard deviation of BGC for different levels of noise. The controller performance is more reliable as the CGM noise levels increase illustrating the effectiveness of the proposed LV-based MPC in controlling BGC when sensor noise is present.

5. CONCLUSIONS

The LV-based MPC is able to provide reliable management of the blood glucose levels for people with T1DM when realistic signal noise exists. According to the simulation results, the controller can avoid the occurrence of hypoglycemia events and manage the BGC after consumption of meals more effectively. In future work this LV-based controller will be extended to multivariable MPC.

ACKNOWLEDGEMENTS

The work of X. Sun was supported by the China Scholarship Council under grant 20190608136. Funds provided by the Hyosung S. R. Cho Endowed Chair at Illinois Institute of Technology to A. Cinar is gratefully acknowledged.

REFERENCES


Table 2. Percentage of time in different glycemic ranges across all subjects for closed-loop glucose control with LV-based MPC and ARX based MPC (Mean/STD, STD: standard deviation)

<table>
<thead>
<tr>
<th>Performance Index</th>
<th>noise level 0</th>
<th>noise level 1</th>
<th>noise level 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;70 mg/dL (%)</td>
<td>ARX</td>
<td>LV</td>
<td>ARX</td>
</tr>
<tr>
<td>70-180 mg/dL (%)</td>
<td>0/0</td>
<td>0.06/0.15</td>
<td>0.12/0.27</td>
</tr>
<tr>
<td>180-250 mg/dL (%)</td>
<td>79.68/11.28</td>
<td>83.23/11.68</td>
<td>71.92/10.93</td>
</tr>
<tr>
<td>&gt;250 mg/dL (%)</td>
<td>20.32/11.28</td>
<td>15.73/10.73</td>
<td>26.84/9.63</td>
</tr>
<tr>
<td>Min (mg/dL)</td>
<td>84.60/5.62</td>
<td>80.30/10.81</td>
<td>71.10/7.36</td>
</tr>
<tr>
<td>Max (mg/dL)</td>
<td>225.60/18.76</td>
<td>232.30/40.49</td>
<td>255.80/24.79</td>
</tr>
</tbody>
</table>

Fig. 4. Performance comparison of LV based MPC and ARX based MPC (adult subject 9)

Fig. 5. Performance comparison of LV-based MPC and ARX-based MPC


