Reshaping CGM data into $Ax = y$

Assume the raw and glucose readings data has been trimmed and aligned already such that α and β are known

Let $R \in \mathbb{R}^{\beta \times n}$ denote the raw data[1] matrix where $R_{i,j}$ is the measure of the j^{th} variable at the i^{th} second for $i \leq \beta$ and $j \leq n$.

$$R = \begin{bmatrix} R_{1,1} & R_{1,2} & R_{1,3} & R_{1,4} & \dots & R_{1,n} \\ R_{2,1} & R_{2,2} & R_{2,3} & R_{2,4} & \dots & R_{2,n} \\ R_{3,1} & R_{3,2} & R_{3,3} & R_{3,4} & \dots & R_{3,n} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ R_{\beta,1} & R_{\beta,2} & R_{\beta,3} & R_{\beta,4} & \dots & R_{\beta,n} \end{bmatrix}$$

Now let $y \in \mathbb{R}^{\alpha}$ be the vector of glucose measurements[2] which are recorded every 5 minutes (300 seconds) where y_k is the measure at the k^{th} 5-minute interval for $k \leq \alpha << \beta$. To have a system of equations of the form $A_i x = y_i$, the a row A_i must contain 5 minutes worth of rows from R. If we use all 300 possible rows (1 second per row) and all n variables from R then $Ax = y$ will be over-determined since $\alpha << 300n$. Instead, we use ℓ features and take a sample of τ rows of R per minute such that $5\ell\tau < \alpha$ and the system is under-determined.

For example, assume $\tau = 1$ and we choose variables R_1 and R_2 ($\ell = 2$) to predict glucose measures. Then rows only have length 10 and $Ax = y$ takes the form

$$\begin{bmatrix} R_{1,1} & R_{1,2} & R_{61,1} & R_{61,2} & R_{121,1} & R_{121,2} & R_{181,1} & R_{181,2} & R_{241,1} & R_{241,2} \\ R_{301,1} & R_{301,2} & R_{361,1} & R_{361,2} & R_{421,1} & R_{421,2} & R_{481,1} & R_{481,2} & R_{541,1} & R_{541,2} \\ R_{601,1} & R_{601,2} & R_{661,1} & R_{661,2} & R_{721,1} & R_{721,2} & R_{781,1} & R_{781,2} & R_{841,1} & R_{841,2} \\ \vdots & \vdots & \vdots & \vdots & & & & \vdots & & \\ R_{\alpha_0,1} & R_{\alpha_0,2} & R_{\alpha_1,1} & R_{\alpha_1,2} & R_{\alpha_2,1} & R_{\alpha_2,2} & R_{\alpha_3,1} & R_{\alpha_3,2} & R_{\alpha_4,1} & R_{\alpha_4,2} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_\alpha \end{bmatrix}$$

where $\alpha_i = 300\alpha + 61i$.

Noise Masking

Let $N \in \mathbb{R}^{\beta}$ be the noise vector such that N_i is the recorded noise level at the i^{th} second and let θ be the noise level we consider 'too noisy'. If $N_i > \theta$ we wish to mask the tuple corresponding to the i^{th} second. The matrices returned from this masking algorithm will have the form

$$\begin{bmatrix} R_{1,1} & R_{1,2} & \blacksquare & \blacksquare & R_{121,1} & R_{121,2} & \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ R_{301,1} & R_{301,2} & R_{361,1} & R_{361,2} & \blacksquare & \blacksquare & R_{481,1} & R_{481,2} & R_{541,1} & R_{541,2} \\ \blacksquare & \blacksquare & R_{661,1} & R_{661,2} & R_{721,1} & R_{721,2} & R_{781,1} & R_{781,2} & \blacksquare & \blacksquare \\ \vdots & \vdots & \vdots & \vdots & & & & \vdots & & \\ R_{\alpha_0,1} & R_{\alpha_0,2} & R_{\alpha_1,1} & R_{\alpha_1,2} & R_{\alpha_2,1} & R_{\alpha_2,2} & \blacksquare & \blacksquare & R_{\alpha_4,1} & R_{\alpha_4,2} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_\alpha \end{bmatrix}$$

[1] R is the concatenation of the "Breathing", "ECG", and "Summary" excel files

[2] y is the "glucose" excel file