18 July 2001

(Terminology as in the paper.)

$$\tilde{X} = \tilde{X}_m + \tilde{Y} q^T$$

$\tilde{X}$: calibration spectra; $\tilde{X}_m$, $\tilde{Y}$, q: unknown

$$\text{(I)} \quad \tilde{Y}_R^T \tilde{X} \left(\tilde{X}^T \tilde{X}\right)^{-\text{PLS}} \tilde{X}^T \tilde{Y}_R = (\tilde{Y}_R^T \tilde{Y}_R) \cdot \frac{SNR_x^2}{1+SNR_x^2} \cdot \frac{SNR_y^2}{1+SNR_y^2}$$

$$= (\tilde{Y}_R^T \tilde{Y}_R) \cdot \frac{SNR^2}{1+SNR^2}$$

$\tilde{Y}_R$: calibration lab references

$$\text{(II)} \quad \tilde{Y}_R^T \tilde{X} \left(\tilde{X}_m^T \tilde{X}_m\right)^{-\text{PLS}} \tilde{X}^T \tilde{Y}_R = (\tilde{Y}_R^T \tilde{Y}_R) \cdot SNR_x^2 \cdot \frac{SNR_y^2}{1+SNR_y^2}$$

$$\Rightarrow \text{ratio} \frac{(\text{II})}{(\text{I})} = 1 + SNR_x^2 \quad \Rightarrow \quad \underline{SNR_x = \sqrt{\frac{(\text{II})}{(\text{I})} - 1}}$$

Problem: $\tilde{X}_m^T \tilde{X}_m$ unknown so you can not compute (II)

Solution: Find spectra (other than the calibration spectra) for which:

(a) $\tilde{X}^T \tilde{X} \approx \tilde{X}_m^T \tilde{X}_m$ because their RMS glucose signal ≈ 0; and

(b) their $\tilde{X}_m^T \tilde{X}_m$ is similar to what was in the calibration spectra.

- E.g., take 1000 prediction spectra and according to their 1000 predictions divide them into three ranges: low, med, and hi glucose (with, say, 200, 500 and 300 spectra). Say the std is 25 $\frac{mg}{dl}$ rms in each range. That means that

$$\left(\tilde{X}_n^T X_n\right)_{\text{estimate}} = \frac{\frac{\tilde{X}_{lo}^T \tilde{X}_{lo}}{200} + \frac{\tilde{X}_{med}^T \tilde{X}_{med}}{500} + \frac{\tilde{X}_{hi}^T \tilde{X}_{hi}}{300}}{3}$$

has about 3x <u>less</u> glucose signal than the original (~75 $\frac{mg}{dl}$).

- The more ranges, the cleaner the noise estimate.
- The cleaner the noise estimate (the less glucose signal in the spectral covariance matrix) the higher the SNR_x-result. In other words, in practice, you end up with a statement like "my SNR_x is <u>at least</u> ...."
- Once you know SNR_x you also know SNR_y from the total-SNR (see paper).

18 July 2001

Technicality:

Because the number of spectra in the calibration data set $\tilde{X}, \tilde{y}_R$ and the data set for the estimation of $\tilde{X}_m^T \tilde{X}_m$ may be different, all variances must be scaled. E.g.:

$$\frac{\frac{\tilde{y}_R^T \tilde{X}}{250} \cdot \left\{ \frac{\frac{\tilde{X}_{lo}^T \tilde{X}_{lo}}{200} + \frac{\tilde{X}_{med}^T \tilde{X}_{med}}{500} + \frac{\tilde{X}_{hi}^T \tilde{X}_{hi}}{300}}{3} \right\}^{-PLS} \cdot \frac{\tilde{X}^T \tilde{y}_R}{250}}{\frac{\tilde{y}_R^T \tilde{X}}{250} \cdot \left\{ \frac{\tilde{X}^T \tilde{X}}{250} \right\}^{-PLS} \cdot \frac{\tilde{X}^T \tilde{y}_R}{250}} \cong 1 + SNR_x^2$$

assuming: 250 calibration spectra
200 spectra in the low-glucose range
500 " " " medium " "
300 " " " high " "