Numerical differentiation of noisy near infrared spectra

Antti Autere,* Timo Mantere and Jarmo T. Alander
*University of Vaasa, Finland. *E-mail: aau@welho.com*

In this paper, the problem of numerical differentiation of noisy NIR spectra is formulated as two kinds of inverse problems: a deterministic and a statistical one. In this way, we are able to filter out noise usually present in the differentiation done by finite differences. The inverse problems are solved by optimisation. The two approaches result in similar derivatives.

Differentiation as an inverse problem

Direct numerical differentiation of noisy signals by, e.g. finite difference formula, amplifies the noise. Figure 1(a) shows a NIR spectrum measured from pork and cow meat and (b) its difference spectrum calculated by a formula $z(j)=y(j+1)-y(j)$, where the vector $y \in R^n$ denotes the NIR spectrum and $z \in R^n$ denotes its difference spectrum.

We differentiate noisy NIR spectra by formulating the task as an inverse problem.[3] In this way we are able to filter noise. Let $x \in R^n$ denote the derivative of the spectrum y and λ denote wavelength. Then

$$y(\lambda)=\int_0^{\lambda} x(\tau)d\tau+e(\lambda), Data=y(\lambda_j)=\int_0^{\lambda_j} x(\tau)dt+e(\lambda_j), \tag{1}$$

where $\lambda_j=j/n, j=1, \ldots, n$ and $e \in R^n$ is noise. The rightmost formula is a discretised version to the leftmost formula. Furthermore, let us approximate the integral (in the rightmost formula in (1)):

$$\int_0^{\lambda_j} x(\tau)dt \approx \frac{1}{n}\sum_{k=1}^{j} x(\lambda_k), \; y = Ax+e, \; \|Ax-y\|^2 + \delta^2 \|x\|^2. \tag{2}$$

In this way, we can formulate a forward problem [the middle formula in (2)], where the matrix $A \in R^{n \times n}$ is

$$A=\frac{1}{n}\begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 1 & 1 & & & \ldots \\ \ldots & & \ldots & & 0 \\ 1 & & & 1 & 0 \\ 1 & 1 & \ldots & 1 & 1 \end{pmatrix}. \tag{3}$$

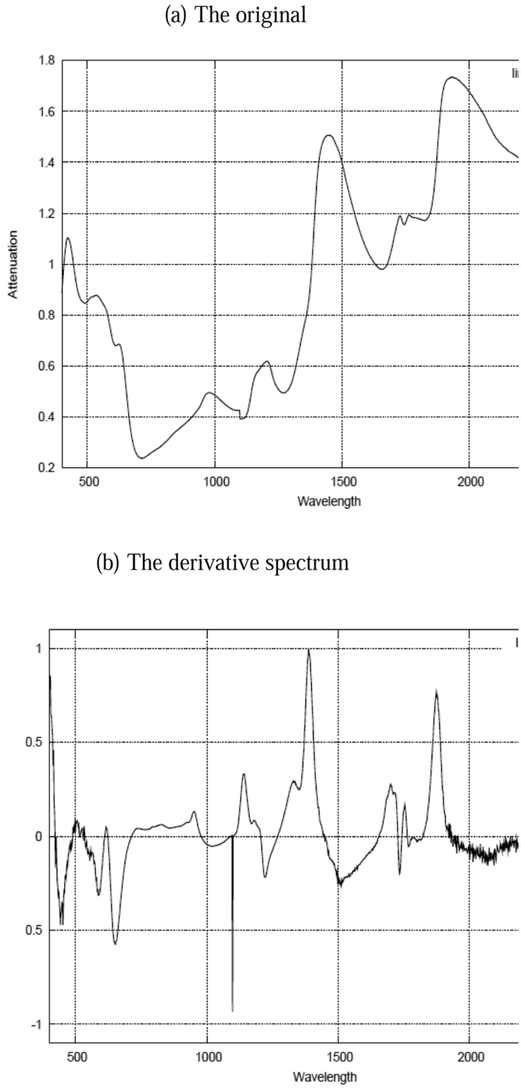

Figure 1. (a) A NIR spectrum and (b) its difference spectrum calculated by the formula $z(j) = y(j+1) - y(j)$.

We could solve the derivative x from the forward problem but then noise again would take over. Instead, x can be solved by *regularising* it. The classical way of doing this is the *Tikhonov regularisation* where we minimise the Tikhonov functional (the rightmost formula in (2)) by penalising the norm of x, see,[2] p. 16.

Another way to regularise x is to solve the forward problem approximately. We have chosen this way. We apply *conjugate gradient* (*CG*) optimiser to the forward problem ignoring the noise term e, see[2] p. 45 and[1] p. 420. It is known that in every iteration $k+1$ of CG $\|y - Ax^{k+1}\| = \|y - \mathbf{A}x^k\|$

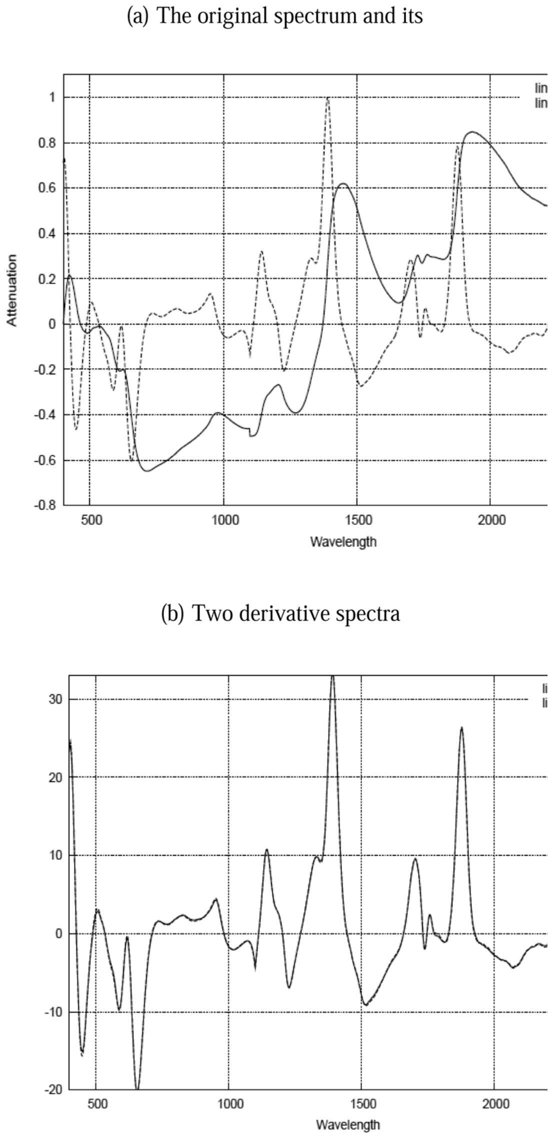

Figure 2. (a) the NIR spectrum and its derivative calculated by the premature convergence of CG (scaled in the same figure). (b) the derivatives calculated by the premature convergence of CG and by the sequential optimisation procedure.

and $\|x^{k+1}\| \geq \|x^k\|$. The latter formula corresponds to the regularisation term $\|x\|$ in the Tikhonov functional.

We iterate CG for some time and then stop. This is called premature convergence. Figure 2(a) shows the NIR spectrum (solid line) and its derivative (dashed line) calculated by 20 iterations of CG. The derivative in Figure 2(a) clearly has less noise than the one in Figure 1(b).

A statistical inverse problem

Consider a stochastic model, cf. the middle formula in (2):

$$\mathrm{Y} = \mathbf{A}X + E,\ X \sim N(\bar{x}, \gamma^2\mathbf{I}),\ E \sim N(0, \sigma^2 I), \tag{4}$$

where $\mathbf{I}$ is a $n \times n$ unit matrix and the Gaussian random variables X and E are mutually independent, see e.g.[2] p. 77. The parameters in the densities are x, the mean of the density of the derivative spectrum, γ and σ. We assume that the variances $\gamma \in$ [0.1, 30] and $\in$ [0.01, 1], cf. Figure (1) (a) and Figure (2) (b). Hence the prior probability density for the derivative spectrum $\bar{x}$ and the likelihood of the data y are, respectively

$$\pi_{prior}(x|\bar{x},\gamma) \propto \frac{1}{\gamma^n}\exp\left(-\frac{1}{2\gamma^2}\|x-\bar{x}\|^2\right),\ \pi(y|x,\sigma) \propto \frac{1}{\sigma^n}\exp\left(-\frac{1}{2\sigma^2}\|y-Ax\|^2\right) \tag{5}$$

From Bayes' formula, the posterior probability density for x is

$$\pi_{post}(x|y,\bar{x},\gamma,\sigma) \propto \pi_{prior}(x|\bar{x},\gamma)\pi(y|x,\sigma) \tag{6}$$

$$\propto \frac{1}{\gamma^n\sigma^n}\exp\left(-\frac{1}{2\gamma^2}\|x-\bar{x}\|^2-\frac{1}{2\sigma^2}\|y-Ax\|^2\right) \tag{7}$$

The posterior density is a Bayesian solution of the statistical inverse problem. In practice, we calculate the maximum of π_{post} to get the *maximum* a *posteriori estimate* (*MAP*) for x. For the maximisation, we take the negative logarithm of π_{post} to obtain a minimisation problem: $x_{\mathrm{MAP}} = argmin_x V(x|\ y, \bar{x}, \gamma, \sigma)$ where

$$V(x|y,\bar{x},\gamma,\sigma) = \frac{2}{2\sigma^2}\left(\|y-Ax\|^2+\left(\frac{\sigma}{\gamma}\right)^2\|x-\bar{x}\|^2+2\sigma^2 n\ln\gamma+2\sigma^2 n\ln\sigma\right) \tag{8}$$

For the minimisation of (8), we use the following sequential optimisation procedure. By taking the partial derivatives for variables x, γ and σ in (8) and equaling them to zero, we obtain:

1. Initialise $\bar{x} = \bar{x}^0 = 0$ -vector, $\gamma = \gamma^0$ and $\sigma = \sigma^0$, set k = 1.
2. Update x:

$$\begin{aligned} x^k &= \arg\max_{x\in R^n}\{\pi_{post}(x|y,\bar{x}^{k-1},\gamma^{k-1},\sigma^{k-1})\} \\ &= \arg\min_{x\in R^n}\left\{\|y-Ax\|^2+\left(\frac{\sigma^{k-1}}{\gamma^{k-1}}\right)^2\|x-\bar{x}^{k-1}\|^2\right\} \end{aligned}$$

3. Update γ and σ and $\bar{x}$:

$$\sigma^k = \frac{1}{n^{1/2}} \| y - Ax^k \|, \;\; y^k = \frac{1}{n^{1/2}} \| x^k - \bar{x}^{k-1} \|, \;\; \bar{x}^k = x^k$$

4. Increase k and repeat from step 2 until convergence.

The second derivatives of γ and σ at γ^k and σ^k were positive in all the iterations. We used again the conjugate gradient method (CG) with 20 iterations to complete step 2. In step 2, we actually have a Tikhonov functional with $\delta=\sigma/\gamma$, cf. the rightmost formula in (2). The regularising coefficient δ is thus a function of the variances of the densities of the derivative spectrum and the noise.

We ran the above optimisation procedure with different starting values σ^0/γ^0. Figure 2(b) shows the derivative spectra corresponding to the lowest obtained minimum of V and that resulted by the premature convergence of CG, cf. Figure 2(a). Both the spectra are almost identical, the *RMS* error between them is 0.5 percent. Hence the Bayesian treatment gives support to the hypothesis that the derivative obtained by the CG in the previous section is very close to the true one.

Conclusions

Numerical differentiation of noisy NIR spectra was done by solving inverse problems. The solvers use conjugate gradient method and a sequential optimisation scheme. The derivatives obtained by solving both the ordinary and the statistical inverse problem were similar. In this way, the noise level was decreased compared to the one present in the differentiation done by finite differences.

Acknowledgements

The European Union Interreg Botnia-Atlantica project Field-NIRCe is acknowledged for funding.

References

1. M.S. Bazaraa, H.D. Sherali and C.M. Shetty, *Nonlinear Programming, Theory and Algorithms*. John Wiley & Sons Ltd, Chichester, UK (2006).
2. J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Springer, Heidelberg, Germany (2005).
3. E. Somersalo, "A numerical differentiation example", in *Lecture Notes on Inverse Problems* (not published).