EIGENVECTOR RESEARCH INCORPORATED
Research, Training, and Software

196 Hyacinth Road
Manson, WA 98831
www.Eigenvector.com

Fitting Smooth Curves Part I: Fitting with Equality Constraints and Basis Functions

Neal B. Gallagher, Donal O'Sullivan

Key words: Data fitting, smoothing penalty, equality constraints, basis functions

Introduction: A flexible methodology for fitting smooth curves to data or spectra was provided by Eilers' Perfect Smoother[1]. Additional flexibility was incorporated into the `wsmooth` function in PLS_Toolbox and Solo[2] and is discussed in the Whittaker Smoother white paper from Eigenvector Research, Inc.[3] In this white paper, an objective function is introduced that extends the smooth fitting approach to include basis functions and equality constraints using the `datafit_engine` function.[2] Smoothness, basis functions and equality constraints are imposed using penalty functions. For additional information, please see [4-6].

Whittaker Smoother with Equality Constraints: The following objective function modifies the `wsmooth` objective to include equality constraints and are given in `datafit_engine` function

$$O(\mathbf{z}) = (\mathbf{y}-\mathbf{z})^{\mathrm{T}}\mathbf{W}_0(\mathbf{y}-\mathbf{z}) + \lambda_{\mathrm{s}}\mathbf{z}^{\mathrm{T}}\mathbf{D}_{\mathrm{s}}^{\mathrm{T}}\mathbf{W}_{\mathrm{s}}\mathbf{D}_{\mathrm{s}}\mathbf{z} + \lambda_{\mathrm{e}}(\mathbf{z}_{\mathrm{e}}-\mathbf{z})^{\mathrm{T}}\mathbf{W}_{\mathrm{e}}(\mathbf{z}_{\mathrm{e}}-\mathbf{z})$$

where $\mathbf{y}$ is a $N \times 1$ vector of measured data, $\mathbf{z}$ is smooth curve to be fit to the data, $\mathbf{W}_0$ is a diagonal matrix of weights (typically $0 \le w_{0,n} \le 1$ for $n = 1,\ldots,N$, $\mathbf{D}_{\mathrm{s}}$ is a second derivative operator (e.g., $\mathbf{D}_{\mathrm{s}}\mathbf{z}$ is the second derivative of $\mathbf{z}$) and λ_{s} is a scalar penalty on the smoothing term and $\mathbf{W}_{\mathrm{s}}$ is a diagonal matrix of weights with entries $0 \le w_{\mathrm{s},n} \le 1$ used to relax smoothing on selected regions of the signal. The equality constraints are given in the vector $\mathbf{z}_{\mathrm{e}}$, with corresponding diagonal matrix of weights, $\mathbf{W}_{\mathrm{e}}$ (with entries $0 \le w_{\mathrm{e},n} \le 1$), and scalar penalty, λ_{e}. When equality constraints are active $\lambda_{\mathrm{e}} > 0$. Elements of $\mathbf{z}_{\mathrm{e}}$ with real values and corresponding elements $w_{\mathrm{e},n} > 0$ are constrained. Elements of $\mathbf{z}_{\mathrm{e}}$ with NaN values have the corresponding elements $w_{\mathrm{e},n} = 0$ and are not constrained. For a given $\mathbf{W}_0$, λ_{s}, $\mathbf{W}_{\mathrm{e}}$, λ_{e} and $\mathbf{z}_{\mathrm{e}}$ the corresponding estimate of $\mathbf{z}$ is given by

$$\hat{\mathbf{z}} = \left(\mathbf{W}_0 + \lambda_{\mathrm{s}}\mathbf{D}_{\mathrm{s}}^{\mathrm{T}}\mathbf{D}_{\mathrm{s}} + \lambda_{\mathrm{e}}\mathbf{W}_{\mathrm{e}}\right)^{-1}\left(\mathbf{W}_0\mathbf{y} + \lambda_{\mathrm{e}}\mathbf{W}_{\mathrm{e}}\mathbf{z}_{\mathrm{e}}\right).$$

For example, an optical emission (OES) spectrum is plotted in Figure 1 (blue) along with a smoothed version shown for $\mathbf{W}_0 = \mathbf{I}$ and $\lambda_{\mathrm{s}} = 1$ (red: moderate smoothing). [The OES are available at www.eigenvector.com.] Also shown are curves with equality constraints with $\mathbf{z}$ constrained to equal the value 44.5 in wavelength regions to $258<\lambda<280$ and $307<\lambda<311$. In addition to the moderate smoothing, the green and black curves have equality penalties of $\lambda_{\mathrm{e}} = 1$ and $\lambda_{\mathrm{e}} = 100$ respectively. As expected, the more strongly penalized fit (black curve) shows a much flatter line across the bottom of the peaks.

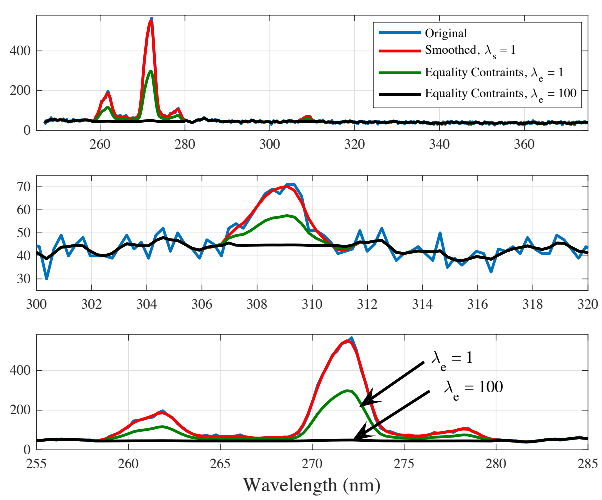

Figure 1: Uncorrected OES spectrum (blue), smoothed spectrum $\lambda_{\mathrm{s}} = 1$ (red). Equality constrained fits with penalties of $\lambda_{\mathrm{e}} = 1$ (green) and $\lambda_{\mathrm{e}} = 100$ (black). Zoom in on a small peak (middle) and significant sharp peaks (bottom).

Whittaker Smoother with Basis Function Constraints: At times, it is desirable to fit a curve to basis functions in addition to the equality constraints. This can be done using penalty functions that "guide" the solution rather than force a hard fit. The objective function is modified to (see the `datafit_engine` function [in PLS_Toolbox, run `datafit_engine demo` with options 1 and 2])

$$O(\mathbf{z},\mathbf{a}) = (\mathbf{y}-\mathbf{z})^{\mathrm{T}}\mathbf{W}_0(\mathbf{y}-\mathbf{z}) + \lambda_{\mathrm{s}}\mathbf{z}^{\mathrm{T}}\mathbf{D}_{\mathrm{s}}^{\mathrm{T}}\mathbf{W}_{\mathrm{s}}\mathbf{D}_{\mathrm{s}}\mathbf{z} +$$
$$\lambda_{\mathrm{e}}(\mathbf{z}_{\mathrm{e}}-\mathbf{z})^{\mathrm{T}}\mathbf{W}_{\mathrm{e}}(\mathbf{z}_{\mathrm{e}}-\mathbf{z}) + \lambda_{\mathrm{b}}(\mathbf{Pa}-\mathbf{z})^{\mathrm{T}}\mathbf{W}_{\mathrm{b}}(\mathbf{Pa}-\mathbf{z})$$

where $\mathbf{W}_{\mathrm{b}}$ is a diagonal matrix of weights with entries $0 \le w_{\mathrm{b},n} \le 1$, $\mathbf{P}$ is a $N \times K_{\mathrm{b}}$ set of basis vectors and $\mathbf{a}$ is a set of coefficients to be estimated. The corresponding estimator is

$$\begin{bmatrix} \hat{\mathbf{z}} \\ \hat{\mathbf{a}} \end{bmatrix} = \boldsymbol{\Gamma}^{-1}\left(\mathbf{W}_0\mathbf{y} + \lambda_{\mathrm{e}}\mathbf{W}_{\mathrm{e}}\mathbf{z}_{\mathrm{e}}\right) \text{ where}$$

$$\boldsymbol{\Gamma} = \begin{bmatrix} \left(\mathbf{W}_0 + \lambda_{\mathrm{s}}\mathbf{D}_{\mathrm{s}}^{\mathrm{T}}\mathbf{D}_{\mathrm{s}} + \lambda_{\mathrm{e}}\mathbf{W}_{\mathrm{e}}\right) & -\lambda_{\mathrm{b}}\mathbf{W}_{\mathrm{b}}\mathbf{P} \\ -\lambda_{\mathrm{b}}\mathbf{P}^{\mathrm{T}}\mathbf{W}_{\mathrm{b}} & \lambda_{\mathrm{b}}\mathbf{P}^{\mathrm{T}}\mathbf{W}_{\mathrm{b}}\mathbf{P} \end{bmatrix}.$$

Figure 2 again shows the original OES signal (blue) and the moderately smoothed signal (red). However, this time the green curve includes a third order polynomial fit with $\lambda_{\mathrm{b}} = 1e^5$ and all diagonal elements of $\mathbf{W}_{\mathrm{b}}$ were equal to one. In contrast the black curve shows results of a second fit performed where the elements of $\mathbf{W}_{\mathrm{b}}$ in the regions 258<λ<280, 307<λ<311 and 393<λ<397 were set to zero – thus, the fit to the basis functions (third order polynomial) is not influence by the major peaks. The green curve fit is clearly influenced by trying to fit the peaks while the black curve shows a smooth curve throughout the majority of the OES spectrum including the major peaks.

Conclusions: Smoothing is a useful tool for providing interpretable trends and processing of spectra. Additional flexibility is provided when using equality constraints and basis functions.

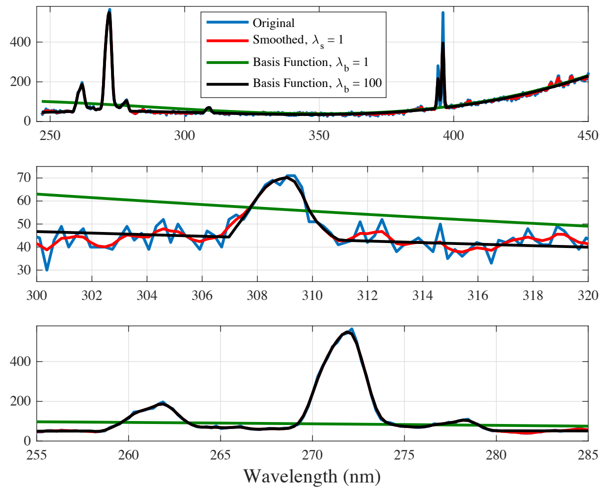

Figure 2: Uncorrected OES spectrum (blue), smoothed spectrum $\lambda_{\mathrm{s}} = 1$ (red). Fit to a third order polynomial through all points in the OES spectrum (green) and fit while neglecting the peaks (black). Zoom in on a small peak (middle) and significant sharp peaks (bottom).

References:

[1] Eilers, PHC, "A Perfect Smoother," *Anal. Chem.* 2003, **75**, 3631-3636.

[2] PLS_Toolbox and Solo. Eigenvector Research, Inc., Manson, WA USA 98831; software available at www.eigenvector.com.

[3] Gallagher, NB, "Whittaker Smoother," white paper Eigenvector Research, Inc., www.eigenvector.com.

[4] Gallagher, NB, "Fitting Smooth Curves Part II: Fitting with a Robust Algorithm," white paper Eigenvector Research, Inc., www.eigenvector.com.

[5] Gallagher, NB, "Fitting Smooth Curves Part III: Baselining with an Asymmetric Least-Squares Algorithm," white paper Eigenvector Research, Inc., www.eigenvector.com.

[6] Gallagher, NB, "Fitting Smooth Curves Part IV: Baselining with Asymmetric Least-Squares and Basis Functions" white paper Eigenvector Research, Inc., www.eigenvector.com.