has appeared in:
ARO Report 79-3
(Proc. 1979 Army Numer.Anal.
and Computers Conf.), 299–302

How does Agee's smoothing method work?

Carl de Boor
Mathematics Research Center
University of Wisconsin-Madison
Madison, Wisconsin 53706

Bill Agee of White Sands Missile Range has been using a somewhat unusual smoothing method with good success. He told me about it last summer and eventually I began to see what might make it work. This is what I want to explain to you.

First, here is a rough description of Agee's smoothing method. This description is rough around the edges since I shall not give any indication as to what is happening near the ends fo the smoothing interval and also leave out other detail.

Starting with a data point sequence (x_i, y_i), $i = 1, \ldots, N$ with $x_i = x_0 + ih$, all i, Agee uses some local smoothing to obtain the smoothed sequence $(x_i, \widetilde{y}_i)$, $i = 1, \ldots, N$. He then constructs Schoenberg's cubic variation diminishing smoothing spline

$$S := V\widetilde{y} := \sum_i \widetilde{y}_i C_i$$

(see, e.g., [1, pp. 159ff]). Next, he calculates the difference between $\widetilde{y}_i$ and this spline at x_i,

$$e_i := \widetilde{y}_i - S(x_i),$$

and appplies V to this error sequence, giving him a new spline Ve which he adds to the earlier spline to get a new approximation

$$\widehat{S} := S + Ve.$$

He may go through this process once again, getting

$$\widehat{\widehat{S}} : \widehat{S} + V\widehat{e} \qquad \text{with} \;\; \widehat{e} := \widetilde{y}_i - \widehat{S}(x_i)$$

and again, getting

$$\widehat{\widehat{\widehat{S}}} : \widehat{\widehat{S}} + V\widehat{\widehat{e}} \qquad \text{with} \;\; \widehat{\widehat{e}} := \widetilde{y}_i - \widehat{\widehat{S}}(x_i)$$

But eventually he stops, with $\widehat{S}$ or $\widehat{\widehat{S}}$ or perhaps even $\widehat{\widehat{\widehat{S}}}$ as his smooth approximation to the original data.

As a first step toward understanding this procedure, consider what would happen if he were to carry out the iteration indefinitely. Think of V as a map applied to a function,

$$Vf := \sum_i f(x_i) C_i$$

and, for this purpose, let $\widetilde{Y}$ be any function with $\widetilde{Y}(x_i) = \widetilde{y}_i$, all i. Then the first spline approximation is

$$S^{[1]} = S = V\widetilde{Y},$$

and, in general,

$$\begin{aligned} S^{[n+1]} &= S^{[n]} + V(\widetilde{Y} - S^{[n]}) \\ &= (1-V)S^{[n]} + V\widetilde{Y}, \qquad n = 1, 2, 3, \ldots\ . \end{aligned}$$

In fact, we can start this iteration formula with $n = 0$, provided we set

$$S^{[0]} := 0.$$

Supposing now $S^{[\infty]} = \lim_{n\to\infty} S^{[n]}$ to exist, it would have to be a fixed point of this iteration, i.e.,

$$S^{[\infty]} = (1-V)S^{[\infty]} + V\widetilde{Y}$$

which then implies that $VS^{[\infty]} = V\widetilde{Y}$, or, since the C_i's are linearly independent,

$$S^{[\infty]}(x_i) = \widetilde{y}_i, \quad \text{all } i.$$

Conclusion: If the process converges, then it converges to the *cubic spline interpolant* to the data $(\widetilde{y}_i)$.

Since $(\widetilde{y}_i)$ is the result of some local smoothing, it would therefore not be a bad idea to use $S^{[\infty]}$. But this is not what Agee does.

In order to understand why Agee's choice of $S^{[3]}$ or $S^{[4]}$ might be better than $S^{[\infty]}$, you have to look now more closely at the iteration map $(1 - V)$. For this, write

$$S^{[n]} =: \sum_i s_i^{[n]} C_i.$$

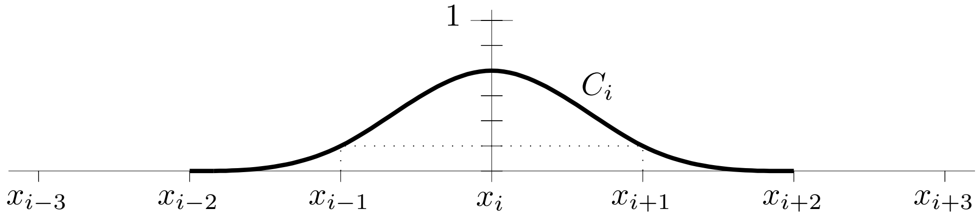
Then the coefficient vectors $s^{[n]}$ are generated by the iteration

$$s^{[n+1]} = (1-C)s^{[n]} + \widetilde{y}$$

with the iteration matrix

$$1 - C = (\delta_{ij} - C_j(x_i)),$$

and I now must tell you more about the functions C_i. These functions are cubic B-splines ($C_i = B_{i-2,4,x}$ in the language of [1]). In particular, C_i looks roughly like



In particular, C_i vanishes off the interval $[x_{i-2}, x_{i+2}]$, and the matrix C, as well as the matrix $1 - C$, is therefore tridiagonal. Explicitly, $1 - C$ has the general row

$$-\frac{1}{6} \quad \frac{1}{3} \quad -\frac{1}{6}$$

and has therefore $\|1 - C\|_\infty = 2/3 < 1$. This insures that the iteration process converges. But we need more information.

If you have ever looked at the iterative solution of linear systems, then you will remember that the manner of convergence depends on the *spectrum* of the iteration matrix. This requires me to bring in one more property of the matrix C, namely the fact that C *is totally positive* (see, e.g., [1; p. 201]). This means that all minors of C are nonnegative. For our purposes, this has the following consequence.

C has a complete set of eigenvectors $v^{[1]}, \ldots v^{[N]}$, corresponding to a sequence

$$\lambda_1 > \lambda_2 > \cdots > \lambda_N \geq 0$$

of eigenvalues, all simple and nonnegative. For the particular matrix C, $1 = \lambda_1$ and $\lambda_N > 0$.

With this, expand the solution $x = s^{[\infty]}$ of the linear system $Cx = \widetilde{y}$ in terms of the $v^{[i]}$,

$$s^{[\infty]} =: \sum_i a_i v^{[i]}.$$

Then, with $s^{[0]} = 0$, we get

$$\begin{aligned} s^{[\infty]} - s^{[k]} &= (1 - C)(s^{[\infty]} - s^{[k-1]}) = \cdots \\ &= (1 - C)^k (s^{[\infty]} - s^{[0]}) \\ &= (1 - C)^k s^{[\infty]} \ = \ \sum_i a_i \ (1 - \lambda_i)^k v^{[i]} \end{aligned}$$

since $(1 - C)v^{[i]} = (1 - \lambda_i)v^{[i]}$, hence

$$s^{[k]} = \sum_i [1 - (1 - \lambda_i)^k] \, a_i \, v^{[i]}.$$

Correspondingly, we get

$$S^{[k]} = \sum_i [1 - (1 - \lambda_i)^k] \, a_i \, V^{[i]},$$

with the cubic spline $V^{[i]}$ given by

$$V^{[i]} := \sum_j v_j^{[i]} \, C_j.$$

This shows in familiar fashion that $S^{[k]}$ approximates the different components $a_iV^{[i]}$ of the interpolant $S^{[\infty]}$ at different rates. For small i, $a_iV^{[i]}$ is already present in $S^{[k]}$ even for small k, since then λ_i is close to 1, hence $(1-\lambda_i)^k$ goes to zero quite fast as k increases. By contrast, λ_i is close to zero for large i, and this means that $(1-\lambda_i)^k$ goes to zero only slowly as k increases. Consequently, $S^{[k]}$ has those well approximated only for large k.

This is a good thing if your are trying to smooth because of the following. The total positivity of C also implies that $V^{[i]}$ has exactly $i-1$ sign changes. So, $V^{[1]}$ is constantly equal to 1, $V^{[2]}$ looks more or less like a straight line, etc. At the other extreme, $V^{[N]}$ changes sign in each interval (x_i, x_{i+1}). But this means that, in the iterative process, the slow moving components of the interpolant $S^{[\infty]}$ are approximated quite quickly, while it takes many iterations to approximate also the fast moving components. This is the reason why it is a good move when smoothing data to stop the iteration after the first few steps, as Agee does.

In conclusion, it is possible to carry out the above analysis without any reference to total positivity since it is easy to construct the vectors $v^{[i]}$ explicitly. But, the above analysis still serves when the data points are not equally spaced in which case there are no explicit or simple formulæ for the $v^{[i]}$.

Reference

1. C. de Boor, A practical guide to splines, Springer-Verlag, New York, 1978.