CONSTRUCTING A FOOLING FUNCTION FOR THE TRAPEZOIDAL RULE

FRED J. HICKERNELL

1. Fooling Function

Choose a large $N = 2^n$ and define

$$x_i = \frac{i}{N}, \qquad i = 0, \dots, N,$$

These nodes are used to define a *periodic* piecewise linear function whose values at the nodes need to be solved. These function values, the derivative values, and the changes in the derivative values are defined as

$$\begin{aligned}
f_i &= f(x_i), \qquad i = 0, \dots, N, & f_N &= f_0 \\
d_i &= f_{i+1} - f_i, \qquad i = 0, \dots, N-1, & d_N &= d_0, \\
s_i &= d_{i+1} - d_i, \qquad i = 0, \dots, N-1.
\end{aligned}$$

Based on these definitions, it follows that

$$\begin{aligned}
d_i &= d_0 + (d_1 - d_0) + (d_2 - d_1) + \cdots + (d_i - d_{i-1}) \\
&= d_0 + s_0 + s_1 + \cdots + s_{i-1}, \qquad i = 0, \dots, N, \\
0 &= s_0 + s_1 + \cdots + s_{N-1}, \\
f_i &= f_0 + (f_1 - f_0) + (f_2 - f_1) + \cdots + (f_i - f_{i-1}) \\
&= f_0 + d_0 + d_1 + \cdots + d_{i-1} \\
&= f_0 + i d_0 + (i-1)s_0 + (i-2)s_1 + \cdots + s_{i-2} \\
&= f_0 + i d_0 + \sum_{j=0}^{i-1} (i-j-1) s_j, \qquad i = 1, \dots, N, \\
0 &= d_0 + d_1 + \cdots + d_{N-1} \\
&= N d_0 + (N-1)s_0 + (N-2)s_1 + \cdots + s_{N-2}.
\end{aligned}$$

Since f is piecewise linear, the integral of f is given by the trapezoidal rule:

$$\begin{aligned}
I(f) := \int_0^1 f(x)\,\mathrm{d}x &= \frac{1}{2N}(f_0 + 2f_1 + 2f_2 + \cdots + 2f_{N-1} + f_N) \\
&= \frac{1}{N}(f_0 + f_1 + f_2 + \cdots + f_{N-1}) \\
&= \frac{1}{N} \sum_{i=0}^{N-1} f_i \\
&= f_0 + d_0 \frac{1}{N} \sum_{i=0}^{N-1} i + \frac{1}{N} \sum_{i=1}^{N-1} \sum_{j=0}^{i-1} (i-j-1) s_j
\end{aligned}$$

$$
\begin{aligned}
&= f_0 + \frac{d_0(N-1)}{2} + \frac{1}{N}\sum_{j=0}^{N-2} s_j \sum_{i=j+1}^{N-1} (i-j-1) \\
&= f_0 + \frac{d_0(N-1)}{2} + \frac{1}{2N}\sum_{j=0}^{N-2} (N-1-j)(N-2-j)s_j
\end{aligned}
$$

2. Trapezoidal Rule

Suppose that there is a trapezoidal rule with $M = 2^m < N$ trapezoids. Then only every L^{th} function value is used, where $L = 2^l = N/M$, $l = n - m$:

$$
\begin{aligned}
T_M(f) &:= \frac{1}{M}(f_0 + f_L + f_{2L} + \cdots + f_{N-L}) = \frac{1}{M}\sum_{i=0}^{M-1} f_{iL} \\
&= f_0 + d_0 L \frac{1}{M}\sum_{i=0}^{M-1} i + \frac{1}{M}\sum_{i=1}^{M-1}\sum_{j=0}^{iL-1} (iL-j-1)s_j \\
&= f_0 + \frac{d_0 L(M-1)}{2} + \frac{1}{M}\sum_{j=0}^{N-L-1} s_j \sum_{i=\lceil (j+1)/L \rceil}^{M-1} (iL-j-1) \\
&= f_0 + \frac{d_0(N-L)}{2} + \\
&\qquad \frac{1}{2M}\sum_{j=0}^{N-L-1} \left(M - \left\lceil \frac{j+1}{L} \right\rceil\right)\left(N - L + L\left\lceil \frac{j+1}{L} \right\rceil - 2j - 2\right) s_j.
\end{aligned}
$$

We will also consider a trapezoidal rule with $M/2$ trapezoids, which then uses every $2L^{\text{th}}$ function value:

$$
\begin{aligned}
T_{M/2}(f) &:= \frac{1}{M}(f_0 + f_{2L} + f_{4L} + \cdots + f_{N-2L}) = \frac{2}{M}\sum_{i=0}^{M/2-1} f_{2iL} \\
&= f_0 + \frac{d_0(N-2L)}{2} + \\
&\qquad \frac{1}{M}\sum_{j=0}^{N-2L-1} \left(\frac{M}{2} - \left\lceil \frac{j+1}{2L} \right\rceil\right)\left(N - 2L + 2L\left\lceil \frac{j+1}{2L} \right\rceil - 2j - 2\right) s_j.
\end{aligned}
$$

The error of this trapezoidal rule is

$$
I(f) - T_M(f) = \frac{1}{N}[(1-L)f_0 + f_1 + \cdots + f_{L-1} + (1-L)f_L + \cdots + f_{N-1}]
$$

3. Constrained Optimization

Given $m, n, p \in \mathbb{N}$ with $p \le m \le n$, and given

A $n \times n$ symmetric, positive definite

B $n \times m$

C $p \times m$

$\boldsymbol{d}$ $m \times 1$,

we want to find $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{y} \in \mathbb{R}^p$ such that

$$\begin{aligned} (\boldsymbol{x}, \boldsymbol{y}) \text{ minimizes} \quad & \boldsymbol{x}^T \mathsf{A} \boldsymbol{x} \\ \text{subject to} \quad & \mathsf{B}^T \boldsymbol{x} + \mathsf{C}^T \boldsymbol{y} = \boldsymbol{d}. \end{aligned}$$

First we write $\boldsymbol{x}$ in terms of B as follows:

$$\boldsymbol{x} = \mathsf{A}^{-1} \mathsf{B} \boldsymbol{\xi} + \mathsf{B}_\perp \boldsymbol{\xi}_\perp,$$

where $\mathsf{B}_\perp$ is $n \times n - m$ such that $(\mathsf{B} \,|\, \mathsf{B}_\perp)$ has full rank, and the columns of $\mathsf{B}_\perp$ are perpendicular to the columns of B. The m-vector $\boldsymbol{\xi}$ and the $n - m$-vector $\xi_\perp$ are the new unknowns replacing $\boldsymbol{x}$. This implies that

$$\begin{aligned} \boldsymbol{x}^T \mathsf{A} \boldsymbol{x} &= \boldsymbol{\xi}^T \mathsf{B}^T \mathsf{A}^{-1} \mathsf{B} \boldsymbol{\xi} + \boldsymbol{\xi}_\perp^T \mathsf{B}_\perp^T \mathsf{A} \mathsf{B}_\perp \boldsymbol{\xi}_\perp \\ \boldsymbol{d} &= \mathsf{B}^T \mathsf{A}^{-1} \mathsf{B} \boldsymbol{\xi} + \mathsf{C}^T \boldsymbol{y} \end{aligned}$$

Thus, one should choose $\boldsymbol{x}_\perp = \boldsymbol{0}$. One may now solve for $\boldsymbol{\xi}$ in terms of $\boldsymbol{y}$:

$$\boldsymbol{\xi} = (\mathsf{B}^T \mathsf{A}^{-1} \mathsf{B})^{-1} (\boldsymbol{d} - \mathsf{C}^T \boldsymbol{y}).$$

Then the quantity to minimize becomes

$$\begin{aligned} \boldsymbol{x}^T \mathsf{A} \boldsymbol{x} &= \boldsymbol{\xi}^T \mathsf{B}^T \mathsf{A}^{-1} \mathsf{B} \boldsymbol{\xi} \\ &= (\boldsymbol{d} - \mathsf{C}^T \boldsymbol{y})^T (\mathsf{B}^T \mathsf{A}^{-1} \mathsf{B})^{-1} (\boldsymbol{d} - \mathsf{C}^T \boldsymbol{y}) \\ &= \boldsymbol{y}^T \mathsf{C} (\mathsf{B}^T \mathsf{A}^{-1} \mathsf{B})^{-1} \mathsf{C}^T \boldsymbol{y} - 2 \boldsymbol{d}^T (\mathsf{B}^T \mathsf{A}^{-1} \mathsf{B})^{-1} \mathsf{C}^T \boldsymbol{y} + \boldsymbol{d}^T (\mathsf{B}^T \mathsf{A}^{-1} \mathsf{B})^{-1} \boldsymbol{d} \end{aligned}$$

The value of $\boldsymbol{y}$ that minimizes this quantity is

$$\boldsymbol{y} = [\mathsf{C} (\mathsf{B}^T \mathsf{A}^{-1} \mathsf{B})^{-1} \mathsf{C}^T]^{-1} \mathsf{C} (\mathsf{B}^T \mathsf{A}^{-1} \mathsf{B})^{-1} \boldsymbol{d}$$

To solve this numerically in a stable way, perhaps we should use singular value decompositions. First we form $(\mathsf{B}^T \mathsf{A}^{-1} \mathsf{B})^{-1}$:

$$\begin{aligned} \mathsf{A} &= \mathsf{V}_1 \Lambda_1^2 \mathsf{V}_1^T, \quad \Lambda_1, \mathsf{V}_1 \ n \times n, \quad \Lambda_1 \text{ diagonal}, \quad \mathsf{V}_1^T \mathsf{V}_1 = \mathsf{I}, \\ \Lambda_1^{-1} \mathsf{V}_1^T \mathsf{B} &= \mathsf{U}_2 \Lambda_2 \mathsf{V}_2^T, \quad \mathsf{U}_2 \ n \times m, \quad \Lambda_2, \mathsf{V}_2 \ n \times n, \\ & \qquad \mathsf{U}_2^T \mathsf{U}_2 = \mathsf{V}_2^T \mathsf{V}_2 = \mathsf{I}, \quad \Lambda_2 \text{ diagonal}, \\ \mathsf{B}^T \mathsf{A}^{-1} \mathsf{B} &= \mathsf{B}^T \mathsf{V}_1 \Lambda_1^{-2} \mathsf{V}_1^T \mathsf{B} = \mathsf{V}_2 \Lambda_2^2 \mathsf{V}_2^T, \\ (\mathsf{B}^T \mathsf{A}^{-1} \mathsf{B})^{-1} &= \mathsf{V}_2 \Lambda_2^{-2} \mathsf{V}_2^T. \end{aligned}$$

Next we form $[\mathsf{C} (\mathsf{B}^T \mathsf{A}^{-1} \mathsf{B})^{-1} \mathsf{C}^T]^{-1}$:

$$\begin{aligned} \Lambda_2^{-1} \mathsf{V}_2^T \mathsf{C}^T &= \mathsf{U}_3 \Lambda_3 \mathsf{V}_3^T, \quad \mathsf{U}_3 \ m \times p, \quad \Lambda_3, \mathsf{V}_3 \ p \times p, \\ & \qquad \mathsf{U}_3^T \mathsf{U}_3 = \mathsf{V}_3^T \mathsf{V}_3 = \mathsf{I}, \quad \Lambda_3 \text{ diagonal}, \\ \mathsf{C} (\mathsf{B}^T \mathsf{A}^{-1} \mathsf{B})^{-1} &= \mathsf{V}_3 \Lambda_3 \mathsf{U}_3^T \Lambda_2^{-1} \mathsf{V}_2^T, \\ \mathsf{C} (\mathsf{B}^T \mathsf{A}^{-1} \mathsf{B})^{-1} \mathsf{C}^T &= \mathsf{V}_3 \Lambda_3^2 \mathsf{V}_3^T, \\ [\mathsf{C} (\mathsf{B}^T \mathsf{A}^{-1} \mathsf{B})^{-1} \mathsf{C}^T]^{-1} &= \mathsf{V}_3 \Lambda_3^{-2} \mathsf{V}_3^T. \end{aligned}$$

Then we solve for $\boldsymbol{y}$:

$$\begin{aligned} \boldsymbol{y} &= [\mathsf{V}_3 \Lambda_3^{-2} \mathsf{V}_3^T][\mathsf{V}_3 \Lambda_3 \mathsf{U}_3^T \Lambda_2^{-1} \mathsf{V}_2^T] \boldsymbol{d} \\ &= \mathsf{V}_3 \Lambda_3^{-1} \mathsf{U}_3^T \Lambda_2^{-1} \mathsf{V}_2^T \boldsymbol{d}. \end{aligned}$$

Finally we solve for $\boldsymbol{x}$ via $\boldsymbol{\xi}$:

$$
\begin{aligned}
\mathsf{C}^T\boldsymbol{y} &= [\mathsf{V}_2\mathsf{\Lambda}_2\mathsf{U}_3\mathsf{\Lambda}_3\mathsf{V}_3^T]\mathsf{V}_3\mathsf{\Lambda}_3^{-1}\mathsf{U}_3^T\mathsf{\Lambda}_2^{-1}\mathsf{V}_2^T\boldsymbol{d} \\
&= \mathsf{V}_2\mathsf{\Lambda}_2\mathsf{U}_3\mathsf{U}_3^T\mathsf{\Lambda}_2^{-1}\mathsf{V}_2^T\boldsymbol{d} \\
\boldsymbol{d} - \mathsf{C}^T\boldsymbol{y} &= \mathsf{V}_2\mathsf{\Lambda}_2(\mathsf{I} - \mathsf{U}_3\mathsf{U}_3^T)\mathsf{\Lambda}_2^{-1}\mathsf{V}_2^T\boldsymbol{d} \\
\boldsymbol{\xi} &= (\mathsf{B}^T\mathsf{A}^{-1}\mathsf{B})^{-1}(\boldsymbol{d} - \mathsf{C}^T\boldsymbol{y}) \\
&= [\mathsf{V}_2\mathsf{\Lambda}_2^{-2}\mathsf{V}_2^T]\mathsf{V}_2\mathsf{\Lambda}_2(\mathsf{I} - \mathsf{U}_3\mathsf{U}_3^T)\mathsf{\Lambda}_2^{-1}\mathsf{V}_2^T\boldsymbol{d} \\
&= \mathsf{V}_2\mathsf{\Lambda}_2^{-1}(\mathsf{I} - \mathsf{U}_3\mathsf{U}_3^T)\mathsf{\Lambda}_2^{-1}\mathsf{V}_2^T\boldsymbol{d} \\
\boldsymbol{x} &= \mathsf{A}^{-1}\mathsf{B}\boldsymbol{\xi} \\
&= [\mathsf{U}_2\mathsf{\Lambda}_2\mathsf{V}_2^T]\mathsf{V}_2\mathsf{\Lambda}_2^{-1}(\mathsf{I} - \mathsf{U}_3\mathsf{U}_3^T)\mathsf{\Lambda}_2^{-1}\mathsf{V}_2^T\boldsymbol{d} \\
&= \mathsf{U}_2(\mathsf{I} - \mathsf{U}_3\mathsf{U}_3^T)\mathsf{\Lambda}_2^{-1}\mathsf{V}_2^T\boldsymbol{d}
\end{aligned}
$$