



Spectral filtering for improved performance of collocation discretization methods

Raymond A. Adomaitis

Department of Chemical Engineering and Institute for Systems Research, University of Maryland, College Park, MD 20742, USA

Received 28 January 2000; received in revised form 17 July 2001; accepted 17 July 2001

Abstract

Spectral filtering methods are investigated for reducing the Gibbs oscillations that result when discontinuous functions are projected onto globally defined trial function expansions. Several physical-space filters are studied to examine which is best suited to high-degree, mixed collocation methods used for time integration of non-linear boundary value problems with piece-wise continuous, time-dependent boundary conditions. Accuracy of the filtered global collocation methods relative to the non-filtered approaches is both in terms of point-wise and norm-wise solution convergence, making the filtered global collocation approach a potential alternative to spline formulations for some time integration applications. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Orthogonal collocation; Spectral filtering; Galerkin method; Numerical analysis; Gibbs phenomenon

1. Introduction

Solutions to boundary-value problems (BVPs) with spatially discontinuous forcing terms can exhibit poor convergence performance when the solutions are represented by a smooth, continuous, globally defined basis function expansion. Likewise, initial conditions that do not satisfy the problem boundary conditions, or discontinuities in the boundary conditions themselves, can result in solutions with non-diminishing oscillations characteristic of the Gibbs phenomenon. Even in cases where satisfactory solutions are found, accurate computation of secondary quantities, such as a flux at a specified spatial location, can be affected by the Gibbs oscillations both near to and away from the discontinuity.

To reduce the effects Gibbs oscillations have on solutions to BVPs, we will incorporate recent developments in spectral filtering methods in collocation discretization procedures based on globally defined trial function expansions. We take advantage of computational methods that make possible very high discretization degrees (Lin, Chang, & Adomaitis, 1999), and use

the filtering methods as a means of improving discretization method convergence performance, rather than the more traditional application as a post-processing step. We consider global polynomial collocation for time integration of two non-linear BVPs as representative test cases for the numerical methods presented in this paper.

Collocation-based time integrators have been used in numerous previous applications. Global polynomial collocation methods have been used for optimal control of batch processes (Biegler, 1984) and for solving initial-boundary value problems (Birnbaum & Lapidus, 1978; Villadsen & Sørensen, 1969). Spline collocation methods (collocation on finite elements) have been used extensively for optimal control (Cuthrell & Biegler, 1987, 1988; Logsdon & Biegler, 1993) and parameter estimation (Kim, Liebman, & Edgar, 1991) applications, and for the direct computation of limit-cycle solutions (Doedel, 1981). Typically, spline collocation methods are used in place of high-degree global discretization methods for applications in which discontinuities in time-dependent model parameters are expected. However, we should not always expect superior performance from spline collocation methods, as will be shown in the following comparison of spline and global collocation solutions to a representative BVP.

E-mail address: adomaiti@isr.umd.edu (R.A. Adomaitis).

1.1. A representative BVP

Consider the non-linear, time-dependent BVP describing reactant diffusion and consumption by a second-order reaction in a spherical catalyst pellet

$$\frac{\partial C}{\partial t} = \nabla^2 C - \phi^2 C^2.$$

This BVP is subject to boundary conditions

$$\frac{\partial C(0, t)}{\partial r} = 0, \quad C(1, t) = b(t),$$

initial condition $C(r, 0) = b(0)$ for $r < 1$, and the definition of $b(t)$:

$$b(t) = \begin{cases} 0.5 & 0 \leq t < t_{\max}/2 \\ 1 & t_{\max}/2 < t \leq t_{\max} \end{cases}$$

with $t_{\max} = 1$. A Galerkin projection procedure (described in Section 5) is used to produce a semi-discretized system consisting of a set of non-linear ordinary differential equations in time that can be integrated by a collocation method. In Fig. 1, time integration results produced using collocation on two discrete intervals (collocation on finite elements) are compared with solutions computed using global collocation.

When viewed in terms of the qualitative solution behavior we would expect from this model, we observe that the solutions produced by each collocation method are comparably inaccurate when the same number of collocation points is used in each procedure. While the observed oscillations are directly attributable to the jump discontinuity in the concentration boundary condition at $t = 1/2$, the mechanisms that give rise to this behavior are actually tied to the Galerkin projection as much as they are to the collocation-based time integration procedure: when using the global collocation procedure, time derivatives of the boundary concentration introduce a term that oscillates with amplitude peaks located at the time integration collocation points. While these time-derivatives are zero when using the spline formulation, projecting the solution at $t = 1/2 -$ onto the basis function expansion valid at $t = 1/2 +$ (used to define the initial condition of the second spline interval) introduces a spatially oscillating concentration profile, disrupting the solution in the second spline interval (see Fig. 2). We note that other formulations of the collocation on finite elements method (such as forcing the solution to be continuous at the knot point, or enforcing solution derivative continuity at the knot) can produce results which are even less accurate than solutions presented in Fig. 1.

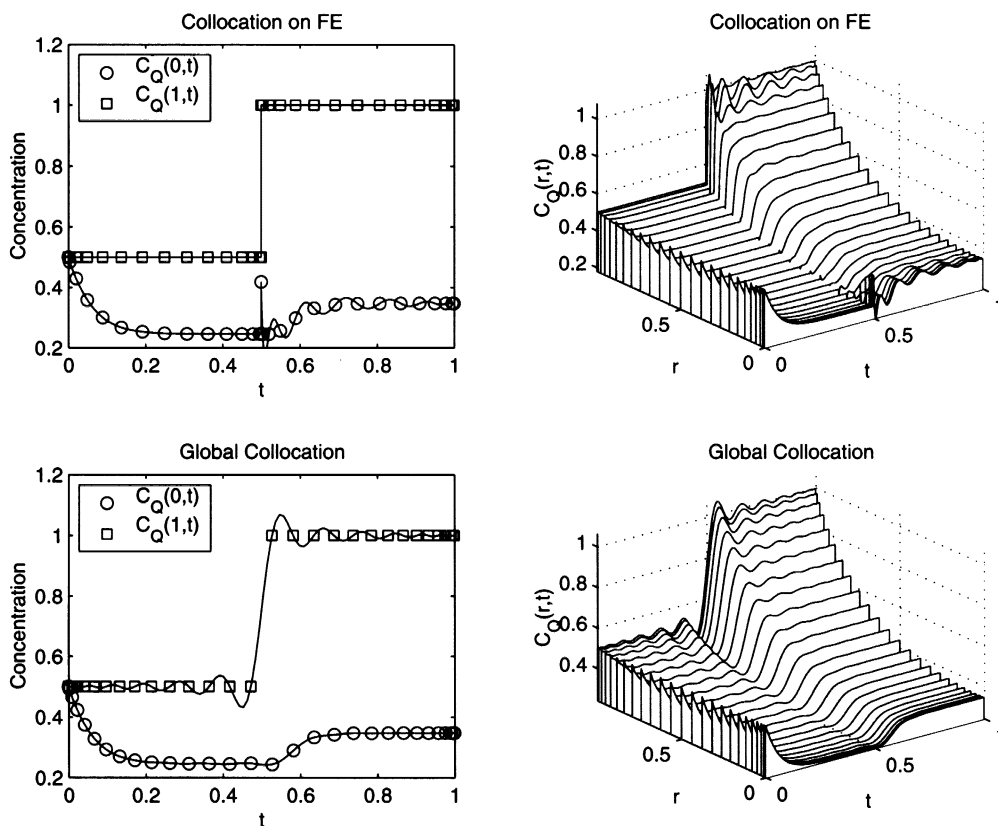


Fig. 1. Catalyst pellet dynamic simulation results produced by collocation on the intervals $0 \leq t \leq 1/2$, $1/2 \leq t \leq 1$ (top) and global collocation over the interval $0 \leq t \leq 1$ (bottom). At total of 30 collocation points are used for each simulation.

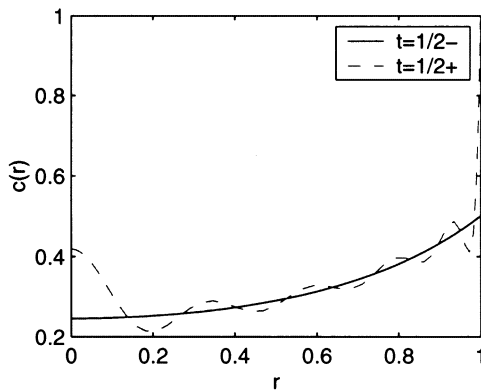


Fig. 2. Snapshot of the catalyst profile just prior to and after $t = 1/2$.

1.2. The filtered global collocation approach

We pose the filtered global collocation approach as an alternative time integration method applicable to some, but certainly not all, situations where a collocation on finite elements method typically would be used. Filtering methods have been an important tool in a variety of process control and signal processing applications (e.g. Morari & Zafiriou, 1989; Hamming, 1977), but the application of filtering to improving solution convergence of Galerkin/collocation methods for BVPs as presented in this paper appears to be new.

The primary motivation for developing the filtered collocation technique was the need for a MATLAB-based numerical integrator suitable for integrating the ODE/AE systems that result from semi-discretized time-dependent boundary value problems. While it is possible to use existing ODE/AE integrators for these applications, the simplicity of implementing global collocation was found valuable when this discretization method was used in conjunction with developing numerical methods for model reduction, parameter identification, numerical continuation of time-periodic solutions, and for developing object-oriented program-

ming concepts for implementing weighted residual methods using global trial functions (e.g. Huang, 2000). Furthermore, global methods allow accurate and sometimes exact (e.g. Adomaitis & Lin, 1998) discretization error assessments. This is necessary for comparing the relative performance of different filtering methods; the multiple-grid approach used as the basis for the MWR-tools methods (Lin et al., 1999) simplifies computing such global error estimates.

2. The Gibbs phenomenon

To illustrate basic issues in spectral filtering and the convergence of global polynomial approximations, consider projecting the piece-wise continuous function defined on the unit interval:

$$f(t) = \begin{cases} 0 & 0 \leq t < 0.55 \\ 1 & 0.55 < t \leq 1 \end{cases} \quad (1)$$

onto the sequence of trial functions $\psi_{n+1}(t) = P_n(t)/\sqrt{c_n}$ where the shifted Legendre polynomials $P_n(t)$ are defined by the recurrence formula

$$(n+1)P_{n+1}(t) = (2n+1)(2t-1)P_n(t) - nP_{n-1}(t) \quad (2)$$

with $P_0(t) = 1$ and $P_1(t) = 2t - 1$. The resulting sequence of functions is orthogonal with respect to inner product

$$\langle P_m, P_n \rangle = \int_0^1 P_m(t)P_n(t)dt = c_n\delta_{m,n}$$

with $c_n = 1/(2n+1)$. Examining the projection $f_N(t)$ of $f(t)$ onto the space spanned by the $\psi_n(t)$, $n = 1, \dots, N$

$$f_N(t) = \sum_{n=1}^N a_n\psi_n(t) \quad (3)$$

with

$$a_n = \langle f(t), \psi_n(t) \rangle,$$

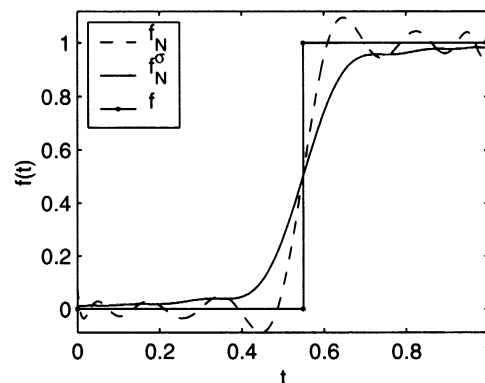
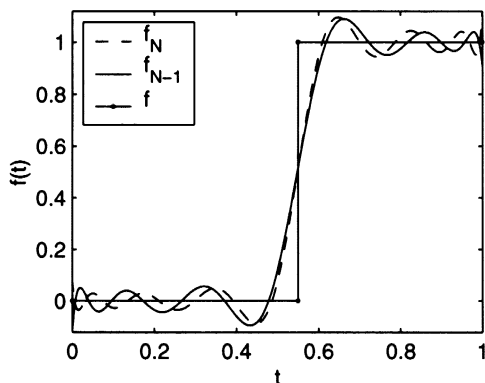


Fig. 3. Gibbs oscillations in two consecutive projections (left) for $N = 15$; first-order filtered results (right).

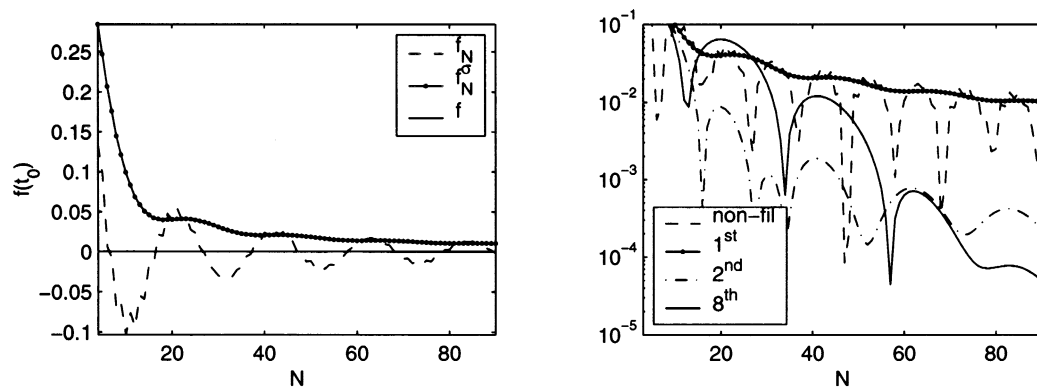


Fig. 4. Pointwise convergence of $f_N(t_0)$ and $f_N^\sigma(t_0)$ at $t_0=0.4$ for non-filtered and first-order filtered projections (left). Convergence of $|f(t_0) - f_N^\sigma(t_0)|$ for $t_0=0.4$ (right), comparing non-filtered results to those obtained using three different filter orders.

we find the expected Gibbs phenomenon (Fig. 3). Three important features are observed in the projection results. First, oscillatory overshoot behavior near the discontinuity does not vanish with increasing truncation number N , but asymptotically approaches a constant value of approximately 8.949% of the jump magnitude (Gottlieb & Orszag, 1977) as N grows large. Second, the oscillatory effects are exhibited far from the discontinuity, for this problem, pointwise convergence away from the discontinuity and the interval endpoints is of order $1/N$, while at the endpoints convergence is proportional to $1/\sqrt{N}$ (cf. the exponential convergence when $f(t)$ is continuous). The slower convergence rate near the endpoints is directly attributable to the that property of the polynomial sequence to give

$$|P_n(t)| < 1, \quad 0 < t < 1$$

but

$$P_n(0) = (-1)^n, \quad P_n(1) = 1.$$

A third observation comes from comparing consecutive projections f_{N-1} and f_N (Fig. 3). When the ψ_n are taken from an orthogonal sequence, the oscillations tend to become out of phase away from the discontinuity, and the features near the discontinuity show little change. Therefore, we can average the two consecutive projections in the hope of obtaining better pointwise convergence performance. Continuing this averaging operation on the partial sums extends the smoothing behavior towards the discontinuity (Fig. 3, right). This averaging operation can be expressed by rewriting the partial sum (Eq. (3)) in terms of the spectrally filtered form

$$f_N^\sigma(t) = \sum_{n=1}^N a_n \sigma_n \psi_n(t)$$

with the σ_n derived to give the equivalent to the arithmetic mean of the partial sums f_n , $n = 1, \dots, N$ (the Cesaro sum):

$$\sigma_n = 1 - \frac{n-1}{N} = 1 - \eta$$

with $\eta = (n-1)/N$. The more uniform, but still slow, pointwise convergence rate is apparent in Fig. 4 (left). In fact, for the representative point chosen for this example, the average partial sum converges at essentially the same rate (but more uniformly) as the non-filtered projection.

3. Filters

The averaged partial sum filter belongs to the class of first-order filters; a filter of order p is a C^{p-1} function whose first $p-1$ derivatives vanish at $\eta=0$; specific definitions can be found in Gottlieb and Shu (1997). In the cited reference, it was shown that higher-order filters can provide better point-wise convergence when the filter order grows with truncation number N . One such filter is the Vandeven (1991) filter, derived to produce exponential pointwise convergence of the filtered partial sums away from the discontinuity for periodic functions:

$$\begin{aligned} \sigma(n) &= 1 - \frac{(2p-1)!}{(p-1)!^2} \int_0^\eta [s(1-s)]^{p-1} ds \\ &= 1 - \frac{(2p-1)!}{(p-1)!^2} \left[\frac{\eta^p}{p} + \sum_{n=1}^{p-1} \frac{(-1)^n (p-1)! \eta^{n+p}}{n!(p-1-n)! n+p} \right]. \end{aligned}$$

To study the potential improvements in pointwise convergence when using the higher-order filters, we examine $|f(t_0) - f_N^\sigma(t_0)|$ as a function of truncation number N and filter order p for the representative point $t_0=0.4$. Results are presented in Fig. 4. We observe that convergence for the second-order filter is, in general, better than the other filter orders for $N < 65$. For higher truncation number values, the eighth-order filter gives improved convergence at the point \$t_0=0.4\$, a result consistent with Gottlieb and Shu (1997).

4. Collocation projection

The tau method (Gottlieb & Orszag, 1977) can be used to force the truncated Legendre polynomial expansion $f_N(t)$ to satisfy a specified set of boundary conditions at $t = 0, 1$; the mixed collocation projection can be considered a discrete approximation to the tau method. If the collocation grid points are defined by t_n : $P_{N-2}(t_n) = 0$ plus the interval endpoints, and f and a are (column) vectors with elements $f(t_i)$ and a_i (see Eq. (3)), i, \dots, N , the discrete transformation array Q used to relate discretized function values f to mode amplitudes a by

$$f = Qa \tag{4}$$

is defined by the elements

$$Q_{i,j} = \psi_j(t_i).$$

Note that this definition of Q can be used to generate discrete transformation arrays that are numerically non-singular for $N > 1000$; such arrays computed using the traditional Vandermonde matrix transpose (Michelsen & Villadsen, 1972; Villadsen & Steward, 1967) typically become numerically singular for $10 < N < 20$. The physical-space filter (defined on the collocation grid) that gives the filtered, discrete ordinate values f^σ is derived using Eq. (4) to give

$$f_N^\sigma = Q\Sigma Q^{-1}f$$

$$f_N^\sigma = P_{sp} f \tag{5}$$

with $\Sigma_{i,i} = \sigma_i$ and $\Sigma_{i,j} = 0, i \neq j$. The σ_j depend on the filter order; consequently, we denote P_{sp}^p as a physical-space filter of order p .

With f representing (non-filtered) discrete points on Eq. (1), the filtered points are computed with Eq. (5) and interpolated values are computed using the mode amplitude coefficients obtained from Eq. (4). Representative results are presented in Fig. 5. Note that the

filtered profiles do not necessarily pass through the original collocation points, but rather through f_N^σ . Smoothing by the second-order filter is superior to smoothing by the eighth-order filter near the jump discontinuity due to the overshoot behavior of the higher-order filtered projection near the discontinuity. However, at the endpoints and away from the discontinuity, the higher-order filters produce smoother and more accurate results.

4.1. Numerical methods for quadrature-based projections

The filtered profiles are reconstructed on a grid that is finer than the collocation grid to assess the inter-collocation point behavior. While it is feasible to use Lagrange interpolation methods to reconstruct the inter-point values, a multiple-grid approach is introduced now because of the computational benefits it will produce when used in computational procedures developed later in this paper. The Legendre polynomial basis functions are computed using the recurrence relation (Eq. (2)) on a fine discretization grid $\hat{t}_m, m = 1, \dots, M$ with $M \gg N$. These quadrature points \hat{t} are defined as the combination of the unit interval endpoints and the roots of the Jacobi polynomial $J_{M-2}^{1,1}$ a polynomial sequence orthogonal with respect to inner product weight $t(1-t)$. The filtered functions are reconstructed on the (finer-scale) quadrature grid using

$$\hat{f}_N^\sigma = \hat{\Psi}Q^{-1}f^\sigma$$

where $\hat{\Psi}_{i,j} = \psi_j(\hat{t}_i)$.

Inner products used for the projection operations discussed in Section 2 and for computing the error norms corresponding to the collocation projections also are computed by quadrature on the fine grid. Finally, because our focus is on applications of collocation-based time integration methods, we compute the derivative of the filtered function (not the filtered derivative) as

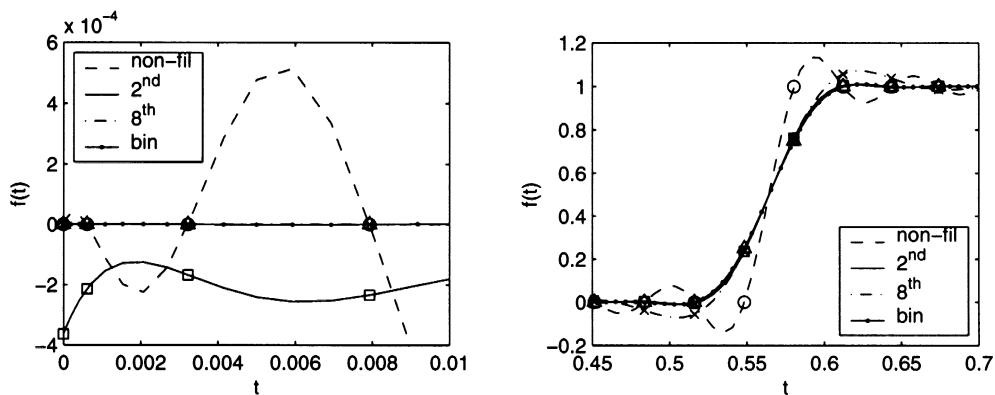


Fig. 5. Collocation projection results; non-filtered results compared with several physical-space filters away from (left) and near (right) the discontinuity. Note that the binomial and eighth-order filters produce nearly identical results near the interval ends. Collocation points for the non-filtered, second, eighth, and binomial filtered projections are marked by $\circ, \square, \times,$ and \triangle , respectively.

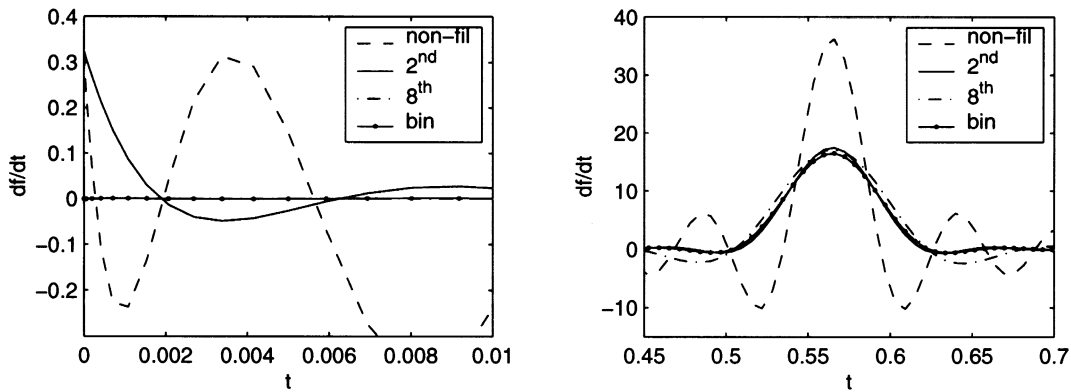


Fig. 6. Oscillations in the first derivative of non-filtered and filtered collocation projections. Again, the binomial and eighth-order filters produce nearly identical results near the interval ends.

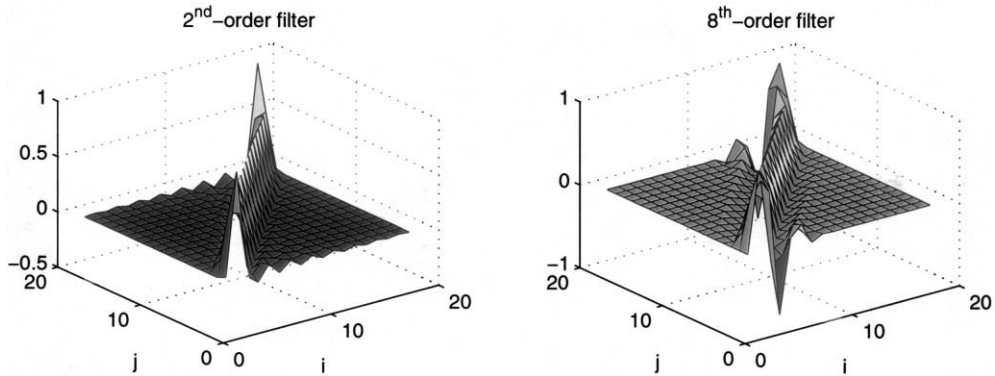


Fig. 7. The physical-space filters corresponding to the Vandeven filter with $N = 18$.

$$\frac{d}{dt} \hat{f}_N^\sigma = \hat{\Psi} Q^{-1} D f^\sigma$$

with

$$D_{i,j} = \sum_{k=1}^N \frac{d\psi_k(t_i)}{dt} Q_{k,j}^{-1}. \tag{6}$$

As in the case of projections of $f(t)$ itself, the derivatives of discontinuous functions filtered using higher order filters feature reduced inter-point oscillation near the interval ends, but continue to have an overshoot behavior near the discontinuity (Fig. 6).

4.2. Modified physical-space filters

The pointwise convergence behavior differences produced using different filter orders, both near to and away from the discontinuity, can be explained by examining the elements of each physical-space filter P_{sp}^p (Fig. 7) array. For example, plotting the elements of P_{sp}^2 reveals that the second-order filter has an oscillatory behavior that diminishes with increasing distance from the filter array diagonal elements. The oscillations reduce the localizing effect of the filter, allowing long-range transmission of information from the discontinuity to the interval endpoints and reducing the

filter performance near the interval endpoints (Fig. 7, left). The eighth-order filter is much more localized resulting in improved performance near the interval ends; however, the visible secondary ‘bumps’ lead to the previously observed overshoot behavior in the neighborhood of the discontinuity, reducing the filter performance in the neighborhood of the discontinuity.

Because of the drawbacks associated with each of the filters studied, modified physical-space filters were investigated, including those that did not have direct transform-space counterparts. One such filter is generated as a weighted sum of filters of different order. In this composite physical-space filtering approach, two arrays are defined: P_{sp}^2 denoting a second-order physical-space filter and P_{sp}^8 denoting an eighth-order physical-space filter. The composite filter then is computed using

$$P_{sp_i}^c = 4t_i(1-t_i)P_{sp_i}^2 + (1-4t_i(1-t_i))P_{sp_i}^8$$

where $P_{sp_i}^c$ denotes the i th row of P_{sp}^c . The resulting filter is shown in Fig. 8; this filter combines the localizing nature of the high-order filters with the reduced overshoot behavior near the diagonal, characteristic of the lower-order filters.

We also consider a simple binomial filter, commonly used for image processing applications (Jahne, 1993). It is defined by $P_{sp}^b(i, i) = 1$ for $i = 1, N$ and $P_{sp}^b(i, i - 1) = P_{sp}^b(i, i + 1) = 1/4$, $P_{sp}^b(i, i) = 1/2$ for $i = 2, \dots, N - 1$; the filter is shown in Fig. 8. Both the composite and binomial filters produced similar results, as measured by the improved convergence near the interval endpoints and the reduced overshoot near the discontinuity. Results for the binomial filter are presented in Figs. 5 and 6. We note that higher-order binomial filters, necessary for cases where stronger filtering is required, can be obtained by repeated applications of this filter, i.e. $(P_{sp}^b)^2$, $(P_{sp}^b)^3$, and so on. We will refer to $(P_{sp}^b)^n$ as a binomial filter of order n .

5. Convergence analysis

To examine the effects filtering has on the convergence of collocation solution procedures, we consider the problem of computing the solution to the catalyst pellet model described in the Section 1. A semi-discrete Galerkin discretization technique is used as the first step of the solution procedure; first a Q -term trial function expansion is used to represent the concentration profile in r and t

$$C_Q(r, t) = b(t) + \sum_{q=1}^Q a_q(t)\eta_q(r).$$

The trial functions $\eta_q(r)$ are generated from the sequence

$$\left\{ r^{2q-2} \cos\left(\frac{\pi r}{2}\right) \right\}_{q=1}^Q$$

and functions in this sequence are made orthonormal with respect to an inner product weighted to correspond to a spherical geometry:

$$\langle \eta_m, \eta_n \rangle_{r^2} = \int_0^1 \eta_m(r)\eta_n(r)r^2 dr$$

using the Gram–Schmidt orthogonalization procedure.

The Galerkin projection produces the Q non-linear ordinary differential equations in time

$$\dot{a}_q(t) = \langle -\dot{b}(t) + \nabla^2 C_Q - \phi^2 C_Q^2, \eta_q \rangle_{r^2}$$

where the inner product operations are carried out by quadrature on a finer, physical-space discretization grid, using the numerical methods discussed in Section 4.1. Writing the mode coefficients \dot{a}_n in (column) vector form gives

$$\dot{a}(t) = \langle -\dot{b}(t) + \nabla^2 C_Q - \phi^2 C_Q^2, \eta^T \rangle_{r^2}.$$

We note that written in this way, the projection of the function $C_Q(r, t)$ onto the (column) vector η^T produces a column vector. This is important, because if we now expand the vector of equations into the discrete-time form at the t_n , $n = 1, \dots, N$, we obtain the $Q \times N$ array of equations

$$\dot{A} = \langle -\dot{b}^T + \nabla^2 C_Q - \phi^2 C_Q^2, \eta^T \rangle_{r^2}$$

where

$$C_{Q_n}(r) = C_Q(r, t_n) = b(t_n) + \sum_{q=1}^Q a_q(t_n)\eta_q(r).$$

The collocation discretization in time gives the large set of non-linear algebraic equations

$$D[A]_2 - \langle -(Db)^T + \nabla^2 C_Q - \phi^2 C_Q^2, \eta^T \rangle_{r^2} = 0 \quad (7)$$

where the notation $D[A]_2$ denotes that the discrete differentiation operation (scaled for $t_{max} \neq 1$) takes place along the second dimension of the array, i.e.

$$D[A]_2 = E$$

with

$$E_{p,q} = \sum_{k=1}^N \frac{D_{q,k} A_{p,k}}{t_{max}}$$

and with D computed using Eq. (6).

The fully discretized modeling Eq. (7) can be solved using the Newton–Raphson method, subject to the concentration initial conditions $a_q(0) = 0$, $q = 1, \dots, Q$. Representative solution and discretization error results

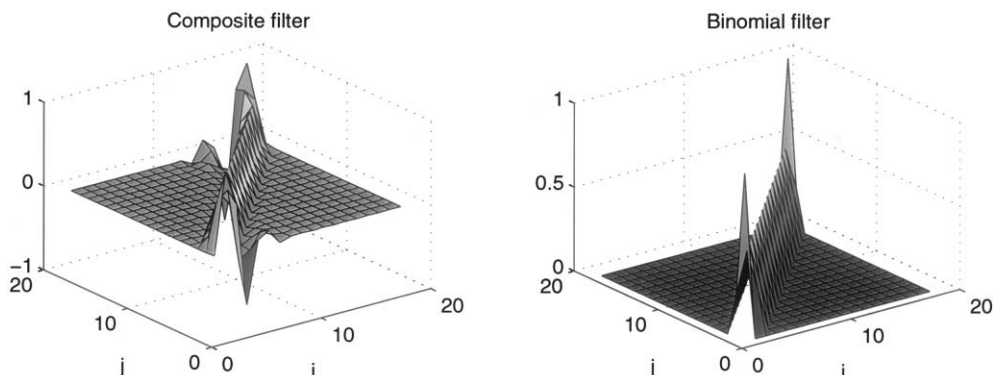


Fig. 8. Composite and binomial physical-space filters for $N = 18$.

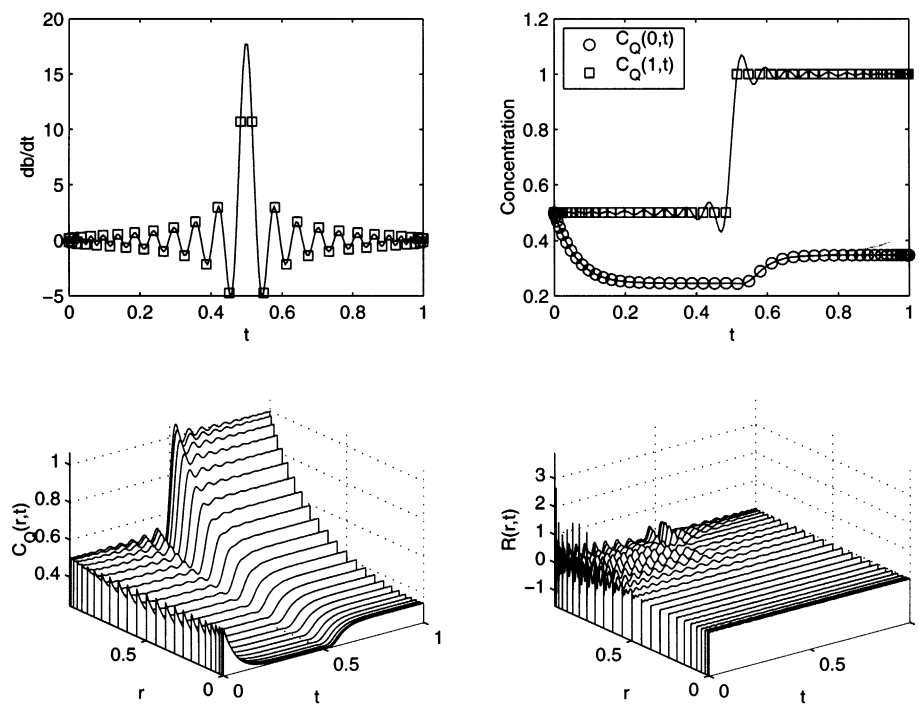


Fig. 9. Simulation results produced using $N = 50$, $Q = 10$, and $\phi = 4$, based on using a non-filtered $b(t)$. Shown are the forcing function time derivative (top left), reactant concentration at $r = 0, 1$ (top right), the concentration profile (bottom left), and the residual function (bottom right).

are presented in Fig. 9. The residual function $R(r, t)$ is defined on the quadrature grid as

$$R(\hat{r}_i, \hat{t}_j) = \left[\frac{\partial}{\partial t} - \nabla^2 \right] C_Q(\hat{r}_i, \hat{t}_j) + \phi^2 [C_Q(\hat{r}_i, \hat{t}_j)]^2$$

where the \hat{r}_i and \hat{t}_j are defined by the quadrature grid points (not the coarser collocation grid). Standard Lagrange interpolation polynomial collocation is used to define discretized equivalents to both the time and spatial derivatives on the quadrature grid (see Lin et al., 1999). This multiple grid approach is critical to accurate numerical analysis of solutions computed for the fully discretized system. In some cases, the multiple grid approach actually allows the exact (within round-off error) evaluation of the residual function (Adomaitis & Lin, 1998), and so coupled with accurate quadrature algorithms, this technique can be considered a relatively rigorous method of solution convergence analysis.

5.1. Comparison of filtered and non-filtered methods

When using the non-filtered collocation projection method, the reconstructed $b_N(t)$ shows significant inter-point oscillations, which translate into severe oscillations in $\dot{b}_N(t)$ (Fig. 9). The oscillations affect both the solution and the residual function in the time dimension. When the collocation projection of $b(t)$ is filtered, however, $\dot{b}_N(t)$ is significantly smoother, and the solution is more accurate; the increased accuracy can be

seen in the smoother and smaller residual function (Fig. 10). The collocation discretization procedure incorporating the physical-space filter is nearly identical to the non-filtered technique; however, the problem to be solved now consists of

$$D[A]_2 - \langle -(Db^\sigma)^T + \nabla^2 C_Q - \phi^2 C_Q^2, \eta^T \rangle_{r,2} = 0 \quad (8)$$

with

$$C_Q(r, t_n) = b_n^\sigma + \sum_{q=1}^Q a_q(t_n) \eta_q(r)$$

and

$$b^\sigma = P_{sp}^n b$$

with $n = 2, 8, c$, or b , corresponding to the second, eighth, composite, and (first-order) binomial filters, respectively.

Computing the norm of the fully discretized solution, we find that the filtered solution procedure appears to converge with exponential accuracy, i.e. $\|R(N)\| \approx \alpha \exp(-\beta N)$ for large values of N ; furthermore, all of the filtered solutions converge at a comparable rate, a rate substantially better than the non-filtered approach (Fig. 11). A better method for assessing the accuracy of the filtered collocation method relative to the non-filtered method is to plot the solution residual function norms as a function of the accuracy of the projection of $b(t)$. The accuracy of the filtered projection of $b(t)$ is computed with

$$\|r_b\|^2 = \int_0^{t_{\max}} [b(t) - b_N^{\sigma}(t)]^2 dt$$

and the non-filtered residual is computed analogously. Results presented in Fig. 11 demonstrate that all filtered methods result in substantially increased solution accuracy relative to the non-filtered method when compared on an equal $b(t)$ discretization error basis.

6. Application to a reverse-flow reactor

Consider computing time-periodic solutions to the reverse-flow reactor model studied by Khinast and Luss (1997, 2000) as a representative application of the filtered collocation approach to time integration of BVPs. In this system, the time scales of the flow-rever-

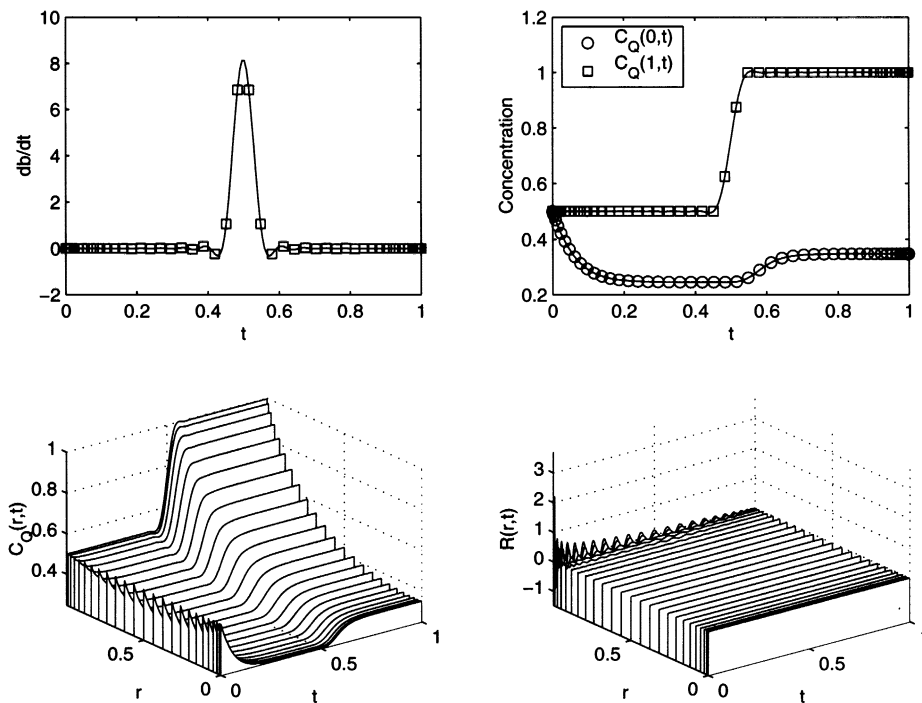


Fig. 10. Simulation results produced using $N = 50$, $Q = 10$, and $\phi = 4$, based on using a (first-order binomial) filtered $b(t)$. Figure format is the same as Fig. 9.

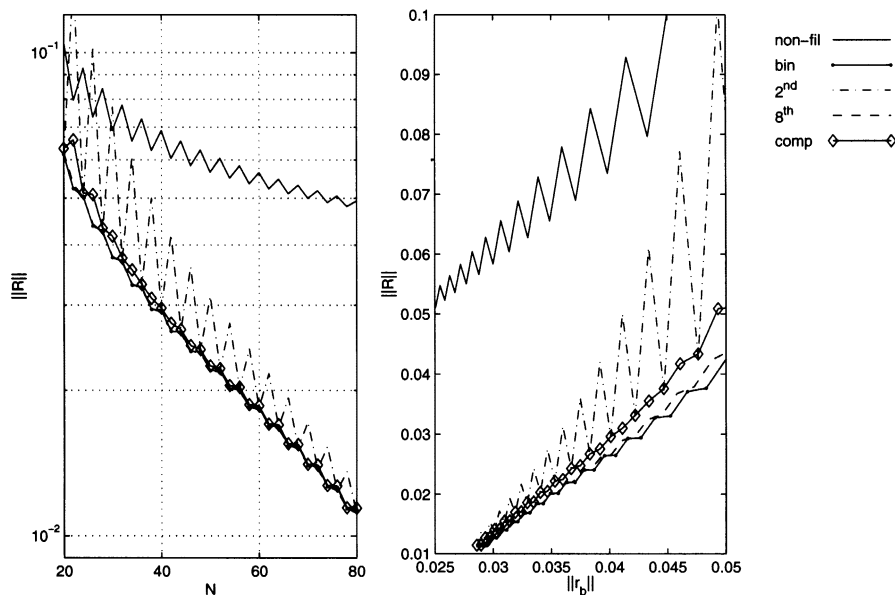


Fig. 11. Residual norm of the fully discretized solution plotted as a function of collocation number (left) and as a function of $b(t)$ projection residual norm (right). Note that the curves representing the eighth order, binomial, and composite filters fall nearly on top of each other.

sal relative to the reactor dynamics (hot-spot movement) generate abrupt changes in the reactor boundary conditions. The benefit of the filtered collocation approach in this application is that while the direct collocation method will converge when the number of points is sufficient to resolve the abrupt change, it is possible to achieve better (more uniform) convergence—and potentially a sufficiently accurate solution with fewer collocation points—if the filtered method is employed.

Because of the large Lewis number value for the reactor system in the cited work, we use the quasi-steady state assumption for the gas phase concentration, giving the differential equation model describing the reactant conversion $x(z, \tau)$

$$0 = \frac{1}{\phi_m^2} \frac{d^2x}{dz^2} - \frac{f}{Da} \frac{dx}{dz} + B(T)(1-x)$$

subject to boundary conditions

$$x(0, \tau) = 0, \quad \frac{dx(1, \tau)}{dz} = 0$$

for feed gas flows corresponding to $f \geq 0$ and

$$x(1, \tau) = 0, \quad \frac{dx(0, \tau)}{dz} = 0$$

for $f < 0$. This choice of boundary conditions can be justified by the short reactor residence time and complete conversion of reactants under the chosen operating conditions. The reactor temperature $T(z, \tau)$ is computed as the solution to

$$\frac{\partial T}{\partial \tau} = \frac{\sigma}{Le\phi_h^2} \frac{\partial^2 T}{\partial z^2} - \frac{\sigma f}{DaLe} \frac{\partial T}{\partial z} + \frac{\beta\sigma}{Le} B(T)(1-x) - \frac{\sigma\Delta}{Le}(T-1)$$

with boundary conditions

$$\frac{\partial T(0, \tau)}{\partial z} - \Phi_0(t) \frac{\phi_h^2}{Da} (T(0, \tau) - 1) = 0,$$

$$\frac{\partial T(1, \tau)}{\partial z} - \Phi_1(t) \frac{\phi_h^2}{Da} (T(1, \tau) - 1) = 0.$$

The boundary condition functions Φ are defined by

$$\Phi_0(t) = \begin{cases} 0 & \text{for } 0.5 < \tau < 1.5 \\ 1 & \text{otherwise} \end{cases}$$

$$\Phi_1(t) = \begin{cases} -1 & \text{for } 0.5 < \tau < 1.5 \\ 0 & \text{otherwise} \end{cases}$$

The dimensionless reactant gas velocity f is defined by $f = \Phi_0 + \Phi_1$. The parameter values are based on Khinast, Gurumoorthy, and Luss (1998) and are given as $Le = 686.6$, $Da = 5.745 \times 10^{-5}$, $\phi_h^2 = 1.4482 \times 10^{-2}$, $\phi_m^2 = 4.16$, $\beta = 0.155$, $\gamma = 25.785$, $\varepsilon = 0.69$, $\alpha_v = 2628.8$

$\text{m}^2 \text{ surf per m}^3 \text{ react}$, $k_c = 0.115 \text{ m/s}$, and $k_{\text{inf}} = 1.815 \times 10^7 \text{ l/s}$. Symmetric period-1, two-peak limit-cycle solutions with substantial hot-spot movement (Fig. 5a of Khinast et al., 1998) result when $\sigma = 3(6.9 \times 10^{-3})$ (corresponding to $t_f = 180 \text{ s}$) and $\Delta = 3500$ are used.

The semi-discretized system is generated using the Galerkin projection method and the globally defined trial function expansions

$$x(z, \tau) = \sum_{q=1}^Q a_q(\tau) \eta_q(z) \quad f \geq 0$$

$$x(z, \tau) = \sum_{q=1}^Q a_q(\tau) \gamma_q(z) \quad f < 0$$

$$T(z, \tau) = 1 + \sum_{q=1}^Q c_q(\tau) \psi_q(z) + d_0(\tau)(1-z)^2 + d_1(\tau)z^2$$

where the basis functions are computed as solutions to the following eigenvalue problems

$$\lambda \eta = \eta'' \quad \eta(0) = \eta'(1) = 0$$

$$\lambda \gamma = \gamma'' \quad \gamma'(0) = \gamma(1) = 0$$

$$\lambda \psi = \psi'' \quad \psi(0) = \psi(1) = 0$$

A semi-discretized set of ODEs in time is produced by substituting the basis function expansions into the modeling equations, and projecting the resulting residual functions onto each basis function. The discretization procedure applied to the reactant conversion balance is straightforward and produces a set of nonlinear algebraic equations at each point in time. Because the time-dependent terms $d_0(\tau)$ and $d_1(\tau)$, computed using the energy balance boundary conditions, are discontinuous the Galerkin solution procedure produces oscillating time-derivative terms \dot{d}_0 and \dot{d}_1 in the time-collocation solution if no filtering is used; in fact, these oscillations will increase as the number of collocation points is increased. However, accurate periodic solutions are found when a third-order binomial filter is applied to $\Phi_0(\tau)$ and $\Phi_1(\tau)$: simulation results corresponding to a time-periodic solution are presented in Fig. 12. The time-periodic solutions are computed directly using a Newton–Raphson, limit-cycle fixed point algorithm; the algorithm typically converged to the limit cycle solution within six iterations. We note that computing the sensitivities of the states at each collocation point in time with respect to the initial conditions comes at essentially no additional computational cost over what is required to compute the Jacobian array elements necessary for the collocation-based time integration procedure.

A representative converged (with respect to time-collocation point and trial function truncation numbers) limit-cycle solution is shown in Fig. 12. These results are consistent with those reported in Fig. 5a of Khinast et al. (1998). We note that the non-filtered global collocation results corresponding to the same time-dis-

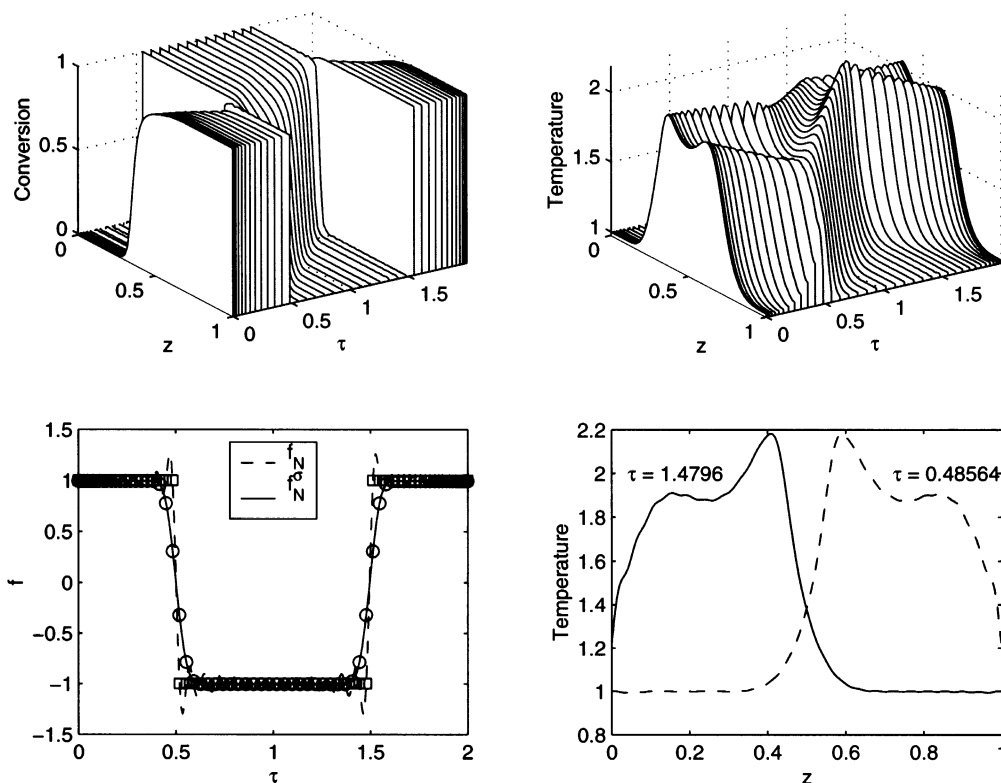


Fig. 12. Simulation snapshots of the conversion and temperature profiles of a reverse-flow reactor. $N = 80$ collocation points in time and $Q = 45$ trial functions were used to produce these simulation results.

cretization and basis function truncation numbers produced completely incorrect results—for example, for the conditions corresponding to the simulation results shown in Fig. 12, the non-filtered approach resulted in a single-peak hot-spot which is qualitatively incorrect.

7. Concluding remarks

This paper focused on the application of filtering methods to reduce the Gibbs oscillations that result when discontinuous functions are projected onto globally defined trial function expansions. Various physical-space implementations of the filters were studied and modified to suit the mixed collocation projection method. It was found that a binomial physical-space filter possessed many of the desirable structural attributes of the range of filters examined; the simplicity with which this filter can be generated makes a strong argument for choosing it over the other physical-space filters for collocation applications. With physical-space filter in hand, the convergence of a representative collocation based discretization application was studied to examine the effect discontinuities had on both normed and point-wise solution convergence. Several methods for quantifying solution global accuracy were presented and used to compare the filtered and non-filtered ap-

proaches. All these methods for assessing solution accuracy demonstrated the substantial improvements produced when filtering was incorporated as part of the collocation solution technique. It was found that the filtered collocation discretization techniques allow high-degree discretizations of problems with time-discontinuous terms that would otherwise make global collocation methods inappropriate.

The original motivation for studying polynomial collocation-based time integration was to expand on the weighted residual methods computational library MWRtools (Adomaitis, Lin, & Chang, 2000; Lin et al., 1999). The goal was to simplify time integration of semi-discretized BVPs, especially those that result in ODE/AE systems. Current work focuses on implementing the polynomial collocation methods in conjunction with developing data structures that define the modeling equations and solutions to the fully discretized system. Development of these simulation tools was motivated by parameter estimation and batch process optimization applications in chemical vapor deposition reactors, solid-state gas sensor system model development, and to study alternative chemical reactor designs and modes of operation, such as the reverse-flow system discussed in this paper. Additional details and demonstration scripts can be obtained at the MWRtools project website www.ench.umd.edu/software/MWRtools.

Acknowledgements

The author acknowledges the support of the National Science Foundation through grants ECS-0082381 and CTS-0085633.

References

- Adomaitis, R. A., & Lin, Y.-h. (1998). A technique for accurate collocation residual calculations. *Chemical Engineering Journal*, 71, 127–134.
- Adomaitis, R. A., Lin, Y.-h., & Chang, H.-Y. (2000). A computational framework for boundary-value problem based simulations. *Simulation*, 74, 30–40.
- Biegler, L. T. (1984). Solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. *Computers and Chemical Engineering*, 8, 243–248.
- Birnbaum, I., & Lapidus, L. (1978). Studies in approximation methods IV: Double orthogonal collocation with an infinite domain. *Chemical Engineering and Science*, 33, 455–462.
- Doedel, E. (1981). AUTO: a program for the automatic bifurcation analysis of autonomous systems. *Congressus Numerantium*, 30, 265–284.
- Cuthrell, J. E., & Biegler, L. T. (1987). On the optimization of differential–algebraic process systems. *American Institute of Chemical Engineering Journal*, 33, 1257–1270.
- Cuthrell, J. E., & Biegler, L. T. (1988). Simultaneous optimization and solution methods for batch reactor control profiles. *Computers and Chemical Engineering*, 13, 49–62.
- Gottlieb, D., & Orszag, S. A. (1977). *Numerical analysis of spectral methods*. SIAM CBMS-NSF regional conference series in Applied Mathematics, vol. 26.
- Gottlieb, D., & Shu, C.-W. (1997). On the Gibbs phenomenon and its resolution. *SIAM Review*, 39, 644–668.
- Hamming, R. W. (1977). *Digital filters* Prentice-Hall signal processing series. Englewood Cliffs, NJ: Prentice-Hall.
- Huang, J. (2000). An object-oriented programming approach to implement global spectral methods: application to dynamic simulation of a chemical infiltration process. M.S. thesis, University of Maryland.
- Jahne, B. (1993). *Digital image processing. Concepts, algorithms, and scientific applications* (2nd ed.). Berlin: Springer.
- Khinast, J., & Luss, D. (1997). Mapping regions with different bifurcation diagrams of a reverse-flow reactor. *American Institute of Chemical Engineering Journal*, 43, 2034–2047.
- Khinast, J., & Luss, D. (2000). Efficient bifurcation analysis of periodically-forced distributed parameter systems. *Computers and Chemical Engineering*, 24, 139–152.
- Khinast, J., Gurumoorthy, A., & Luss, D. (1998). Complex dynamic features of a cooled reverse-flow reactor. *American Institute of Chemical Engineering Journal*, 44, 1128–1140.
- Kim, I.-W., Liebman, M. J., & Edgar, T. F. (1991). A sequential error-in-variables method for nonlinear dynamic systems. *Computers and Chemical Engineering*, 15, 663–670.
- Lin, Y.-h., Chang, H.-Y., & Adomaitis, R. A. (1999). MWR tools: a library for weighted residual methods. *Computers and Chemical Engineering*, 23, 1041–1061.
- Logsdon, J. S., & Biegler, L. T. (1993). A relaxed reduced space SQP strategy for dynamic optimization problems. *Computers and Chemical Engineering*, 17, 367–373.
- Michelsen, M. L., & Villadsen, J. (1972). A convenient computational procedure for collocation constants. *Chemical Engineering Journal*, 4, 64–68.
- Morari, M., & Zafriou, E. (1989). *Robust process control*. Englewood Cliffs, NJ: Prentice-Hall.
- Vandeven, H. (1991). Family of spectral filters for discontinuous problems. *Journal of Scientific Computing*, 6, 159–192.
- Villadsen, J., & Sørensen, J. P. (1969). Solution of partial differential equations by a double collocation method. *Chemical Engineering Science*, 24, 1337–1349.
- Villadsen, J. V., & Steward, W. E. (1967). Solution of boundary-value problems by orthogonal collocation. *Chemical Engineering Science*, 22, 1483–1501.