Optimization of transport-reaction processes using nonlinear model reduction

Eugene Bendersky, Panagiotis D. Christofides*

Department of Chemical Engineering, School of Engineering and Applied Sciences, University of California, 405 Hilgard Avenue, Box 951592, Los Angeles, CA 90095-1592, USA

Abstract

This article focuses on optimization problems arising in the context of transport-reaction processes which are governed by nonlinear elliptic partial differential equations and proposes a computationally efficient method for their solution. The central idea of the method is to discretize the infinite-dimensional optimization problem by utilizing the method of weighted residuals with empirical eigenfunctions obtained by applying Karhunen–Loéve expansion to an appropriately constructed ensemble of solutions of the PDE equality constraints for different values of the design variables. This model reduction procedure leads to low-dimensional nonlinear programs that represent accurate approximations of the original infinite-dimensional nonlinear program, and whose solution can be obtained with standard optimization algorithms. The key issues of construction of the ensemble used for the computation of the empirical eigenfunctions and validity of the optimal solutions computed from the finite-dimensional programs are addressed. The proposed method is applied to two representative transport-reaction processes and is shown to be more efficient compared to conventional optimization approaches based on spatial discretization with the finite-difference method. © 2000 Elsevier Science Ltd. All rights reserved.

1. Introduction

Optimization problems arising in the context of transport-reaction processes typically involve maximization/minimization of a nonlinear functional (e.g., integral of a nonlinear function over the domain of definition of the process) subject to a finite set of partial differential equation (PDE) equality constraints and a finite set of spatially dependent inequality constraints. As an example of such an optimization problem, consider the maximization of the integral of the reaction rate of an exothermic catalytic reaction over the surface of the catalyst by choosing properly the temperature of the cooling medium used to remove heat from the catalyst and satisfying mass and energy balances for reactant species and catalyst, respectively, and constraining the

catalyst and cooling medium temperatures to be lower than certain values.

The conventional approach to solve optimization problems with PDE constraints is to directly apply standard spatial discretization techniques (e.g., finite differences, finite elements) to transform the nonlinear functional into a nonlinear function and the PDE constraints into a large set of algebraic constraints. Then, the resulting finite-dimensional nonlinear programming problem is solved by using standard methods (see, for example, Vasantharajan, Viswanathan & Biegler, 1990; Manousiouthakis & Sourlas, 1992; Floudas & Panos, 1992; Biegler, Nocedal & Schmid, 1995). Unfortunately, even though this approach is conceptually straightforward, it may require using a very large number of discretization points/elements in order to compute the optimal solution with the desired accuracy, and thus, it may be computationally expensive. The reason for which this approach may be computationally inefficient is that a brute force discretization with finite differences/elements does not

---

*Corresponding author. Tel.: 1-310-794-1015; fax: 1-310-206-4107.
E-mail address: pdc@seas.ucla.edu (P.D. Christofides)

account for the inherent characteristics of the PDE equality constraints.

One approach to solve optimization problems with PDE constraints, while accounting for the inherent characteristics of the PDEs in the discretization process, is to use Galerkin's method with the eigenfunctions of the linear spatial differential operator as basis functions for the discretization (Yu & Seinfeld, 1973). This approach is motivated from the fact that the main feature of elliptic PDEs is that the dominant structure of their solutions is usually characterized by a finite (typically small) number of degrees of freedom (Teman, 1988) (for example, in the case of systems with linear spatial differential operators this follows from the fact that the eigenspectrum of the spatial differential operator can be partitioned into a finite-dimensional slow one and an infinite-dimensional stable fast complement). Even though, this approach may significantly reduce the dimension of the optimization problem which results from the spatial discretization for few specific types of differential operators for which the eigenfunction expansions converge quickly (note that for general Sturm–Liouville operators such eigenfunction expansions converge very slowly), it cannot be directly applied to problems that involve nonlinear spatial differential operators (e.g., nonlinear dependence of the diffusion coefficient and thermal conductivity on temperature). The reason is that the eigenvalue problems of nonlinear spatial differential operators cannot be, in general, solved analytically, and thus, it is difficult to a priori (i.e. without having any information about the solution of the system) choose an optimal (in the sense that will lead to an accurate approximation) basis to expand the solution of the PDE system. An approximate way to address this problem (Ray, 1981) is to linearize the nonlinear spatial differential operator around a steady state and address the optimization problem on the basis of the resulting quasi-linear system. However, this approach is only valid in a small neighborhood of the steady state where the linearization takes place.

An alternative approach which is not based on linearization is to utilize detailed finite difference (element) simulations of the PDE system to compute a set of *empirical eigenfunctions* (dominant spatial patterns) of the system through Karhunen–Loéve (K–L) expansion (also known as proper orthogonal decomposition and principal component analysis). The K–L expansion was introduced by Lumley (see, for example, Lumley, 1981) as a procedure for computation of coherent structures (empirical eigenfunctions) in turbulent flows from experimental data of fluid flow quantities. However, the original formulation of K–L expansion was computationally intractable, especially for large data sets, until Sirovich proposed the method of snapshots in 1987 (Sirovich, 1987a,b). The use of K–L expansion for computing empirical eigenfunctions which are employed as

basis functions in Galerkin's method has been shown to lead to the derivation of accurate nonlinear low-dimensional approximations of several dissipative PDE systems arising in the modeling of diffusion-reaction processes and fluid flows (e.g., Park & Cho, 1996; Bangia, Batcho, Kevrekidis & Karniadakis, 1997), as well as to the synthesis of linear and nonlinear low-order controllers that can be readily implemented in practice (e.g., Shvartsman & Kevrekidis, 1998; Baker & Christofides, 1999; Baker & Christofides, 2000).

In the context of optimization of transport-reaction processes, few papers have appeared in the literature that utilize Galerkin's method with empirical eigenfunctions obtained through K–L expansion as the means of solving infinite-dimensional nonlinear programs. A notable exception is the recent paper (Park & Lee, 1998) where an ill-posed inverse heat convection problem was efficiently discretized by using Galerkin's method with empirical eigenfunctions and solved through the conjugate gradient method. In another recent work (Esposito & Floudas, 1998), fundamental issues on global optimization of nonlinear programs that involve coupled differential (specifically, initial value ordinary differential equations) and algebraic constraints were studied.

In this work, we propose a computationally efficient method for the solution of optimization problems in transport-reaction processes described by nonlinear elliptic PDEs. We initially construct an ensemble of solutions of the PDE equality constraints for different values of the design variables, which is then used to compute a set of empirical eigenfunctions for the optimization problem by employing K–L expansion. The empirical eigenfunctions are subsequently used as basis functions in the method of weighted residuals employed to transform the original infinite-dimensional optimization problem to a set of low-dimensional nonlinear programs. The finite-dimensional programs are iteratively solved by using successive quadratic programming (SQP), until a solution that satisfies the optimality conditions for the infinite-dimensional program is obtained. The key issues of construction of the ensemble used for the computation of the empirical eigenfunctions and validity of the optimal solution computed from the finite-dimensional programs are addressed. The proposed method is applied to two representative transport-reaction processes: (a) it is used to maximize the integral of the reaction rate over the length of a catalytic rod, where an exothermic reaction takes place and diffusive phenomena are dominant, by selecting coolant temperature, and (b) it is used to maximize the outlet concentration of the product species in a packed-bed reactor, where an exothermic reaction takes place and both diffusive and convective phenomena are important, by selecting coolant temperature. Comparisons with conventional approaches involving the use of finite difference schemes for spatial

discretization illustrate the computational efficiency of the proposed method.

2. Preliminaries

2.1. Formulation of the optimization problem

We consider a system of nonlinear elliptic PDEs with the following description:

$$0 = \mathcal{A}(x) + f(x, d) \tag{1}$$

subject to the boundary conditions:

$$Cx + D\frac{dx}{d\eta} = R, \quad \text{on } \Gamma, \tag{2}$$

where $x(z)$ denotes the vector of state variables whose dimension is assumed to be n, $z = [z_1, z_2, z_3] \in \Omega \subset \mathbb{R}^3$ is the vector of spatial coordinates, Ω is the domain of definition of the process and Γ its boundary. $\mathcal{A}(x)$ is an elliptic, possibly nonlinear, spatial differential operator which includes first- and second-order spatial derivatives, $f(x, d)$ is a nonlinear vector function, $d \in \mathbb{R}^p$ is the vector of design variables which are assumed to be constant, C, D are constant matrices, $dx/d\eta|_\Gamma$ denotes the derivative in the direction perpendicular to the boundary and R is a constant vector. Throughout the paper, we will use the inner product and norm in $L_2[\Omega]$ (where $L_2[\Omega]$ is the space of square integrable functions which are defined in Ω), which are defined, respectively, as

$$(\phi_1, \phi_2) = \int_\Omega \phi_1(z)\phi_2(z)\,dz, \quad \|\phi_1\|_2 = (\phi_1, \phi_1)^{1/2}, \tag{3}$$

where ϕ_1, ϕ_2 are two elements of $L_2[\Omega]$.

A typical optimization problem for the system of Eqs. (1) and (2) can be formulated as follows:

$$\min \quad \int_\Omega G(x, d)\,dz$$

s.t

$$0 = \mathcal{A}(x) + f(x, d), \quad (P) \tag{4}$$

$$Cx + D\frac{dx}{d\eta} = R \quad \text{on } \Gamma,$$

$$g(x, d) \leqslant 0,$$

where $\int_\Omega G(x, d)\,dz$ is the objective function and $g(x, d)$ is the vector of inequality constraints. Both $G(x, d)$ and $g(x, d)$ are assumed to be continuous differentiable with respect to their arguments.

The n-PDE equality constraints and the p inequality constraints create a region (in an appropriate infinite-dimensional Hilbert space) for the variables x, termed the feasible region, and the presence of n state variables $x(z)$ in (P) implies that we have p degrees of freedom for optimization. In the remainder of this subsection, we state the necessary and sufficient optimality conditions of the infinite-dimensional program (P) which will be used to check the accuracy of the optimal solutions obtained by solving various finite-dimensional programs, and to terminate the proposed optimization algorithm. To this end, we form the following Lagrangian functional:

$$L = \int_\Omega G(x, d)\,dz + \lambda^T(\mathcal{A}(x) + f(x, d)) + \mu^T g(x, d), \tag{5}$$

where the coefficients λ^T, μ^T are the Kuhn–Tucker multipliers. Denoting the optimal solutions as $(x^*(z), \lambda^*, \mu^*, d^*)$, the necessary optimality conditions have the following form:

1. Linear dependence of the gradients of the objective and constraint functions:

$$\nabla L(x^*(z), \lambda^*, \mu^*, d^*)$$
$$= \int_\Omega \nabla G(x^*, d^*)\,dz + \lambda^{*T}(\nabla \mathcal{A}(x^*)$$
$$\quad + \nabla f(x^*, d^*)) + \mu^{*T}\nabla g(x^*, d^*)$$
$$= 0, \quad \forall z \in \Omega. \tag{6}$$

2. Feasibility of the nonlinear program (P):

$$\nabla \mathcal{A}(x^*) + \nabla f(x^*, d^*) = 0 \quad \text{and}$$
$$g(x^*, d^*) \leqslant 0, \quad \forall z \in \Omega, \tag{7}$$

3. Complementarity condition:

$$\mu^{*T}\nabla g(x^*, d^*) = 0, \quad \forall z \in \Omega. \tag{8}$$

4. Nonnegativity of inequality constraint multipliers

$$\mu^{*T} \geqslant 0, \quad \forall z \in \Omega. \tag{9}$$

5. Constraint qualification: Linear independence of active constraint gradients:

$$[\nabla g_A(x^*, d^*)|\nabla \mathcal{A}(x^*) + \nabla f(x^*, d^*)], \quad \forall z \in \Omega, \tag{10}$$

where $g_A(x^*, d^*)$ denotes the set of active constraints.

On the other hand, the following condition is sufficient for optimality:

$$p^T \nabla_{xx} L(x^*(z), \lambda^*, \mu^*, d^*)p > 0,$$
$$(\nabla \mathcal{A}(x^*) + \nabla f(x^*, d^*))^T p = 0, \tag{11}$$
$$\nabla g_A(x^*, d^*)^T p = 0,$$

where $\nabla_{xx} L(x^*(z), \lambda^*, \mu^*, d^*)$ is the Hessian matrix of the Lagrangian and p are the allowable (constrained)

directions for the optimization variables based on the active constraints.

Remark 1. Referring to the nonlinear program (P) and the corresponding optimality conditions, the following remarks are in order: (a) we make no assumption of convexity of the functions and the feasibility region, and therefore, we focus on the computation of a local optimum, (b) the optimality conditions 3 and 4 mean that when an inequality constraint is inactive (i.e., $g_i(x^*, d^*) < 0$, $\forall z \in \Omega$) then the corresponding multiplier $\mu_i = 0$ (the constraint is ignored in the optimality conditions), while when an inequality constraint is active (i.e., $g_i(x^*, d^*) = 0$, for some $z \in \Omega$), then $\mu_i \geqslant 0$, (c) the second-order condition of Eq. (11), which guarantees optimality, requires positive curvature of the Lagrange function in the allowable (constrained) directions, p, and (d) the methodology that will be proposed in this work for computing the solution of (P) can be readily generalized to nonlinear programs, that include coupled nonlinear elliptic PDEs and nonlinear algebraic equality constraints, as well as explicit dependence of the functions G, f, g on the vector of spatial coordinates, z, and have the following general form:

$$\min \quad \int_\Omega G(x, z, d)\, \mathrm{d}z$$

s.t.

$$0 = \mathscr{A}(x) + f(x, z, d)$$

$$Cx + D\frac{\mathrm{d}x}{\mathrm{d}\eta} = R \quad \text{on } \Gamma, \tag{12}$$

$$R(x, z, d) = 0,$$

$$g(x, z, d) \leqslant 0,$$

where $R(x, z, d) = 0$ is a set of algebraic spatially dependent equality constraints.

In the next subsection, we consider a representative diffusion–reaction process example and formulate an optimization problem of the form of Eq. (4). This example will be used through the paper to illustrate the various theoretical concepts and evaluate the computational efficiency of the proposed optimization method.

2.2. Illustrative example

We consider a catalytic rod where an elementary exothermic reaction of the form A → B takes place. The temperature of the rod is adjusted by changing the temperature at five points located symmetrically along
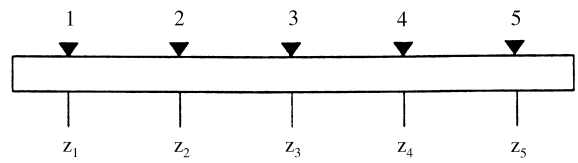

Fig. 1. Catalytic rod with five-point actuators. Actuator locations: $0.1l$, $0.3l$, $0.5l$, $0.7l$ and $0.9l$ where l is the length of the rod.

the length of the rod (see Fig. 1 for a schematic of the process). Under standard modeling assumptions, the steady-state profiles of the concentration of species A and temperature of the rod are given by the following two ordinary differential equations:

$$D\frac{\mathrm{d}^2 C_A}{\mathrm{d}\bar{z}^2} = k_1 C_A \mathrm{e}^{(-E/RT)} - h(C_A^\infty - C_A), \tag{13}$$

$$K\frac{\mathrm{d}^2 T}{\mathrm{d}\bar{z}^2} = \Delta H_R k_1 C_A \mathrm{e}^{(-E/RT)}$$
$$- U\left(\sum_{j=1}^5 \delta(\bar{z} - \bar{z}_j) T_{Jj} + b(z) T^\infty - T \right)$$

subject to the Dirichlet boundary conditions

$$@ \ \bar{z} = 0, l \quad C_A(z) = C_A^\infty,$$
$$@ \ \bar{z} = 0, l \quad T(z) = T^\infty, \tag{14}$$

where C_A is the concentration of species A, T is the rod temperature, D, K, l are the diffusivity, conductivity and length of the rod, respectively, $k_1, E, \Delta H_R$ are the reaction rate constant, activation energy and enthalpy of reaction, respectively, C_A^∞ is the concentration of A in the bulk, $\delta(\cdot)$ is the standard Dirac function, $\bar{z}_j$ and T_{Jj} are the location and temperature of the jth point actuator, respectively, $b(z)$ is a function which is unity for all $\bar{z} \in (0, l)$ except for the points $\bar{z}_j$ where $b(z) = 0$, T^∞ is the temperature of the bulk, h and U are the mass and heat transfer coefficients between the bulk and the rod, respectively, and R is the universal gas constant.

A typical optimization problem for this process is to maximize the rate of production of species B throughout the rod (i.e., maximize the rate of reaction of species A) by adjusting the temperature of the five-point actuators, while constraining the rod temperature at all positions and the temperatures of the point actuators to be lower than certain maximum values. The reader may refer to Bendersky and Christofides (1999) for the study of this problem with distributed actuation. Mathematically, this optimization problem is formulated

as follows:

$$\max \quad \int_0^l k_1 C_A e^{(-E/RT)} \, d\bar{z},$$

$$\text{s.t.} \quad D \frac{d^2 C_A}{d\bar{z}^2} = k_1 C_A e^{(-E/RT)} - h(C_A^\infty - C_A),$$

$$K \frac{d^2 T}{d\bar{z}^2} = \Delta H_R k_1 C_A e^{(-E/RT)} - U(T_e - T),$$

$$T_e = \sum_{j=1}^{5} \delta(\bar{z} - \bar{z}_j) T_{Jj} + b(z) T^\infty, \qquad (15)$$

$$\text{at } \bar{z} = 0, l; \quad C_A = C_A^\infty,$$

$$\text{at } \bar{z} = 0, l, \quad T = T^\infty,$$

$$T_{\min} \leqslant T(\bar{z}) \leqslant T_{\max}, \quad \forall \bar{z},$$

$$T_{J\min} \leqslant T_{Jj} \leqslant T_{J\max}, \quad (j = 1, \ldots, 5).$$

Using the following dimensionless variables,

$$x_1 = \frac{C_A - C_A^\infty}{C_A^\infty}, \quad x_2 = \frac{T - T^\infty}{T_{\max}}, \quad z = \frac{\bar{z}}{l}, \quad \alpha = \frac{k_1}{h},$$

$$\beta = \frac{E}{RT_{\max}}, \quad \lambda_1 = \frac{hl^2}{D}, \quad u = \frac{T_e}{T_{\max}}, \quad u_j = \frac{T_{Jj}}{T_{\max}},$$

$$\gamma = \frac{-\Delta H_R k_1 C_A^\infty}{T_{\max} U}, \quad \lambda_2 = \frac{Ul^2}{K}, \quad v = \frac{T^\infty}{T_{\max}}$$

the following optimization problem, which is included in the general class of problems of Eq. (4), is obtained:

$$\min \left(-\int_0^1 x_1 e^{(-\beta/x_2 + v)} \, dz \right)$$

s.t.

$$\frac{d^2 x_1}{dz^2} = [\alpha(x_1 + 1)e^{(-\beta/x_2 + v)} + x_1]\lambda_1,$$

$$\frac{d^2 x_2}{dz^2} = [-\gamma(x_1 + 1)e^{(-\beta/x_2 + v)} + x_2 + v - u]\lambda_2,$$

$$u(z) = \sum_{j=1}^{5} \delta(z - z_j) u_j + b(z) v,$$

$$\text{at } z = 0, 1, \quad x_1 = 0,$$

$$\text{at } z = 0, 1, \quad x_2 = 0,$$

$$x_{1\min} \leqslant x_1(z) \leqslant x_{1\max}, \quad \forall z,$$

$$x_{2\min} \leqslant x_2(z) \leqslant x_{2\max}, \quad \forall z,$$

$$u_{\min} \leqslant u_j \leqslant u_{\max}, \quad (j = 1, \ldots, 5).$$

2.3. Methodological framework for solution of optimization problem

Owing to the presence of the nonlinear PDE equality constraint of Eq. (1), the optimization problem (P) cannot be solved directly, and a numerical spatial discretization scheme should be employed to reduce the PDE system of Eq. (1) into a set of algebraic equations. The standard approach to address this problem is to utilize finite differences or finite elements to perform the spatial discretization, and then solve the resulting finite-dimensional nonlinear program using standard techniques. The main disadvantage of this approach is that the number of nonlinear algebraic constraints resulting from the spatial discretization, which yields an acceptable approximation, may be very large, thereby leading to a computationally expensive optimization problem. To circumvent this problem, we employ the following methodology for solving (P).

1. Initially, we form an ensemble of solutions of the PDE system of Eq. (1) for different values of the design variables.
2. Then, we apply Karhunen–Loéve expansion to this ensemble to derive a set of empirical eigenfunctions (dominant spatial patterns that minimize the mean square error over all ensemble elements).
3. These empirical eigenfunctions are then used as basis functions within a model reduction framework based on the method of weighted residuals to transform the original infinite-dimensional nonlinear program into a low-dimensional nonlinear program.
4. Finally, various finite-dimensional nonlinear programs are iteratively solved by using successive quadratic programming (SQP) until a solution that satisfies the optimality conditions for the infinite dimensional program (Eqs. (6)–(11)) is computed.

In the next three sections, we describe the computation of empirical eigenfunctions via Karhunen–Loéve expansion, the reduction on the order of (P) through the method of weighted residuals with empirical eigenfunctions, and the computation of the optimum through the solution of the various finite-dimensional programs via successive quadratic programming.

3. Computation of empirical eigenfunctions via Karhunen–Loéve expansion

The Karhunen–Loéve (K–L) expansion (also known as proper orthogonal decomposition, method of empirical eigenfunctions, and principal component analysis) is a procedure used to compute an optimal (in a sense that will become clear below) basis for a modal decomposition of the PDE system of Eq. (1) from an appropriately constructed set of data of this system obtained by

simulations. In this work, the ensemble of data is constructed by computing the solutions of the PDE system

$$0 = \mathcal{A}(x) + f(x, d),$$

$$Cx + D\frac{\mathrm{d}x}{\mathrm{d}\eta} = R \quad \text{on } \Gamma \tag{16}$$

for different values of d. Specifically, we construct a representative ensemble by discretizing the interval in which each design variable d_m $(m = 1, \ldots, p)$ is defined into m_{d_m} (not necessarily equispaced) intervals and computing one set of data for every possible combinations of the discrete values of d_m (see also Remark 3 below). Application of K–L expansion to this ensemble of data provides an orthogonal set of basis functions (known as empirical eigenfunctions) for the representation of the ensemble, as well as a measure of the relative contribution of each basis function to the total energy (mean square fluctuation) of the ensemble (empirical eigenvalues). A truncated series representation of the ensemble data in terms of the dominant basis functions has a smaller mean square error than a representation by any other basis of the same dimension (see Remark 4 below for a rigorous statement of this point). Therefore, the K–L expansion yields the most efficient way for computing the basis functions (corresponding to the largest empirical eigenvalues) capturing the dominant patterns of the ensemble.

For simplicity of the presentation, we describe the K–L expansion in the context of the system of Eq. (1) with $n = 1$ and assume that there is available a sufficiently large set of solutions of this system for different values of d, $\{\bar{v}_\kappa\}$, consisting of K sampled states, $\bar{v}_\kappa(z)$, (which are typically called "snapshots"). The reader may refer to Fukunaga (1990) and Holmes, Lumley and Berkooz (1996) for a detailed presentation and analysis of the K–L expansion. We define the ensemble average of snapshots as $\langle \bar{v}_\kappa \rangle := (1/K)\sum_{\kappa=1}^{K} \bar{v}_\kappa(z)$ (we note that non-uniform sampling of the snapshots and weighted ensemble average can be also considered; see, for example, Graham and Kevrekidis (1996)). Furthermore, the ensemble average of snapshots $\langle \bar{v}_\kappa \rangle$ is subtracted out from the snapshots, i.e.

$$v_\kappa = \bar{v}_\kappa - \langle \bar{v}_\kappa \rangle \tag{17}$$

so that only fluctuations are analyzed. The issue is how to obtain the most typical or characteristic structure (in a sense that will become clear below) $\phi(z)$ among these snapshots $\{v_\kappa\}$. Mathematically, this problem can be posed as the one of obtaining a function $\phi(z)$ that maximizes the following objective function:

$$\text{Max} \quad \frac{\langle (\phi, v_\kappa)^2 \rangle}{(\phi, \phi)},$$

$$\text{s.t.} \quad (\phi, \phi) = 1, \ \phi \in L^2([\Omega]). \tag{18}$$

The constraint $(\phi, \phi) = 1$ is imposed to ensure that the function, $\phi(z)$, computed as a solution of the above maximization problem, is unique. The Lagrangian functional corresponding to this constrained optimization problem is

$$\bar{L} = \langle (\phi, v_\kappa)^2 \rangle - \lambda((\phi, \phi) - 1) \tag{19}$$

and necessary conditions for extrema is that the functional derivative vanishes for all variations $\phi + \delta\psi \in L^2[\Omega]$, where δ is a real number

$$\frac{\mathrm{d}\bar{L}(\phi + \delta\psi)}{\mathrm{d}\delta}(\delta = 0) = 0, \quad (\phi, \phi) = 1. \tag{20}$$

Using the definitions of inner product and ensemble average, $\mathrm{d}\bar{L}(\phi + \delta\psi)/\mathrm{d}\delta$ at $(\delta = 0)$ can be computed as follows:

$$\frac{\mathrm{d}\bar{L}(\phi + \delta\psi)}{\mathrm{d}\delta}(\delta = 0)$$

$$= \frac{\mathrm{d}}{\mathrm{d}\delta}[\langle (v_\kappa, \phi + \delta\psi)(\phi + \delta\psi, v_\kappa) \rangle$$

$$\quad - \lambda(\phi + \delta\psi, \phi + \delta\psi)]_{\delta=0}$$

$$= 2\,Re[\langle (v_\kappa, \psi)(\phi, v_\kappa) \rangle - \lambda(\phi, \psi)]$$

$$= \left\langle \int_\Omega \psi(z)v_\kappa(z)\,\mathrm{d}z \int_\Omega \phi(\bar{z})v_\kappa(\bar{z})\,\mathrm{d}\bar{z} \right\rangle - \lambda \int_\Omega \phi(\bar{z})\psi(\bar{z})\,\mathrm{d}\bar{z}$$

$$= \int_\Omega \left(\left\{ \int_\Omega \langle v_\kappa(z)v_\kappa(\bar{z}) \rangle \phi(z)\,\mathrm{d}z \right\} - \lambda\phi(\bar{z}) \right) \psi(\bar{z})\,\mathrm{d}\bar{z}. \tag{21}$$

Since $\psi(\bar{z})$ is an arbitrary function, the necessary conditions for optimality take the form

$$\int_\Omega \langle v_\kappa(z)v_\kappa(\bar{z}) \rangle \phi(z)\,\mathrm{d}z = \lambda\phi(\bar{z}), \quad (\phi, \phi) = 1. \tag{22}$$

Introducing the two-point correlation function

$$K(z, \bar{z}) = \langle v_\kappa(z)v_\kappa(\bar{z}) \rangle = \frac{1}{K}\sum_{\kappa=1}^{K} v_\kappa(z)v_\kappa(\bar{z}) \tag{23}$$

and the linear operator

$$R := \int_\Omega K(z, \bar{z})\,\mathrm{d}\bar{z} \tag{24}$$

the optimality condition of Eq. (22) reduces to the following eigenvalue problem of the integral equation:

$$R\phi = \lambda\phi \Rightarrow \int_\Omega K(z, \bar{z})\phi(\bar{z})\,\mathrm{d}\bar{z} = \lambda\phi(z). \tag{25}$$

The computation of the solution of the above integral eigenvalue problem is, in general, a very expensive computational task. To circumvent this problem, Sirovich

introduced in 1987 (Sirovich, 1987a,b) the method of snapshots. The central idea of this technique is to assume that the requisite eigenfunction, $\phi(z)$, can be expressed as a linear combination of the snapshots i.e.

$$\phi(z) = \sum_k c_k v_k(z). \tag{26}$$

Substituting the above expression for $\phi(z)$ in Eq. (25), we obtain the following eigenvalue problem:

$$\int_\Omega \frac{1}{K} \sum_{\kappa=1}^{K} v_\kappa(z) v_\kappa(\bar z) \sum_{k=1}^{K} c_k v_k(\bar z) \, d\bar z = \lambda \sum_{k=1}^{K} c_k v_k(z). \tag{27}$$

Defining the κK element

$$B^{\kappa k} := \frac{1}{K} \int_\Omega v_\kappa(\bar z) v_k(\bar z) \, d\bar z \tag{28}$$

the eigenvalue problem of Eq. (27) can be equivalently written in the following matrix form

$$Bc = \lambda c. \tag{29}$$

The solution of the above eigenvalue problem (which can be obtained by utilizing standard methods from matrix theory) yields the eigenvectors $c = [c_1 \ \dots \ c_K]$ which can be used in Eq. (26) to construct the eigenfunction $\phi(z)$. From the structure of the matrix B, it follows that is symmetric and positive semi-definite, and thus, its eigenvalues, λ_κ, $\kappa = 1, \dots, K$, are real and non-negative. Furthermore,

$$\int_\Omega \phi_i(z) \phi_j(z) \, dz = 0, \quad i \neq j. \tag{30}$$

Remark 2. We note that the basis that we compute using KL decomposition is independent of the functional that we try to minimize. Therefore, the same basis can be used to perform computationally efficient optimizations with respect to different functionals associated with the same underlying set of partial differential equations.

Remark 3. We note that the value of m_{d_m} should be determined based on the effect of the design variable d_p on the solution of the system of Eq. (16) (if, for example, the effect of the variable d_1 is larger than the effect of the variable d_2, then m_{d_1} should be larger than m_{d_2}).

Remark 4. The optimality of the empirical eigenfunctions obtained via K–L expansion can be explained as follows. Consider a snapshot $v_\kappa(z)$ of the ensemble of snapshots, v_κ, and the set of empirical eigenfunctions obtained by applying K–L expansion to v_κ, and let

$$v_\kappa(z) = \sum_{l=1}^{L} \gamma_l \phi_l(z) \tag{31}$$

be the decomposition of $v_\kappa(z)$ with respect to this basis. Assume that the eigenfunctions have been ordered so the corresponding eigenvalues satisfy $\lambda_1 > \lambda_2 > \cdots > \lambda_{l+1}$. Then, it can be shown (Holmes et al., 1996) that if $\{\psi_1, \psi_2, \dots, \psi_\kappa\}$ is some arbitrary set of orthonormal basis functions in which we expand $v_\kappa(z)$, then the following result holds for any L:

$$\sum_{l=1}^{L} \langle (\phi_l, v_\kappa)^2 \rangle = \sum_{l=1}^{L} \lambda_l \geqslant \sum_{l=1}^{L} \langle (\psi_l, v_\kappa)^2 \rangle. \tag{32}$$

This implies that the projection on the subspace spanned by the empirical eigenfunctions will on average contain the most energy possible compared to all other linear decompositions, for any number of modes L.

4. Computation of finite-dimensional nonlinear programs

In this section, the empirical eigenfunctions, $\phi_k(z)$, will be used to derive well-defined low-order approximations of the infinite-dimensional nonlinear program (P) by using the method of weighted residuals. The central idea of the method of weighted residuals (see Finlayson (1980) for a comprehensive review of this method) is to approximate the exact solution of $x(z)$ by an infinite series of orthogonal basis functions (that form a complete set) defined on Ω with constant coefficients, substitute the series expansion into Eq. (1) to form the residual, and then force the residual to be orthogonal to a complete set of weighted functions (i.e., the inner product of the residual with a complete set of weighting functions in $L_2[\Omega]$ is set equal to zero) to compute a set of algebraic equations whose unknowns are the constant coefficients of the series expansion of the solution.

In order to describe the application of the method of weighted residuals to (P) and show that the resulting finite-dimensional nonlinear programs are well-posed (in a sense made precise in Proposition 1 below), we assume, for simplicity in the notation, that $n = 1$ and expand the solution $x(z)$ in an infinite series in terms of a complete set of basis functions $\phi_k(z)$ as follows:

$$x(z) = \sum_{k=1}^{\infty} a_k \phi_k(z), \tag{33}$$

where a_k are constant coefficients. We point out that even though our approach uses basis functions, $\phi_k(z)$, computed by K–L expansion from an appropriately constructed ensemble of solutions of the elliptic PDE, the results of this section are also applicable when any other complete set of basis functions, defined in $L_2[\Omega]$ and satisfy the boundary conditions of Eq. (2), is employed in Eq. (33). Substituting the expansion of Eq. (33)

into (P), we get

$$\min \quad \int_\Omega G\left(\sum_{k=1}^\infty a_k \phi_k(z), d\right) dz$$

$$\text{s.t} \quad 0 = \mathscr{A}\left(\sum_{k=1}^\infty a_k \phi_k(z)\right) + f\left(\sum_{k=1}^\infty a_k \phi_k(z), d\right), \qquad (34)$$

$$g\left(\sum_{k=1}^\infty a_k \phi_k(z), d\right) \leqslant 0.$$

Multiplying the PDE and the inequality constraints with the weighting functions, $\psi_v(z)$, and integrating over the entire spatial domain (i.e., taking inner product in $L_2[\Omega]$ with the weighting functions), the following infinite-dimensional nonlinear program is obtained:

$$\min \quad \int_\Omega G\left(\sum_{k=1}^\infty a_k \phi_k(z), d\right) dz$$

$$\text{s.t} \quad 0 = \int_\Omega \psi_v(z) \mathscr{A}\left(\sum_{k=1}^\infty a_k \phi_k(z)\right) dz$$

$$\qquad\qquad + \int_\Omega \psi_v(z) f\left(\sum_{k=1}^\infty a_k \phi_k(z), d\right) dz, \qquad (35)$$

$$\int_\Omega \psi_v(z) g\left(\sum_{k=1}^\infty a_k \phi_k(z), d\right) dz \leqslant 0, \ v = 1, \dots, \infty.$$

Truncating the series expansion of $x(z)$ up to order N and keeping the first N equations (i.e. $v = 1, \dots, N$), the infinite-dimensional program of Eq. (35) reduces to the following finite-dimensional one:

$$\min \quad \int_\Omega G\left(\sum_{k=1}^N a_{kN} \phi_k(z), d\right) dz$$

$$\text{s.t} \quad 0 = \int_\Omega \psi_v(z) \mathscr{A}\left(\sum_{k=1}^N a_{kN} \phi_k(z)\right) dz$$

$$\qquad\qquad + \int_\Omega \psi_v(z) f\left(\sum_{k=1}^N a_{kN} \phi_k(z), d\right) dz, \qquad (36)$$

$$\int_\Omega \psi_v(z) g\left(\sum_{k=1}^N a_{kN} \phi_k(z), d\right) dz \leqslant 0, \ v = 1, \dots, N,$$

where a_{kN} is the approximation of a_k obtained by an Nth-order truncation. From Eq. (36), it is clear that the form of the algebraic equality and inequality depends on the choice of the weighting functions, as well as on N. The weighting functions determine the type of weighted residual method being used (see Remark 5 below).

Theorem 1 that follows establishes that the finite-dimensional nonlinear program of Eq. (36) is a well-defined approximation of the infinite-dimensional program (P) in the sense that its optimal solution converges to the optimal solution of (P) as $N \to \infty$. This results follows from the smoothness of the functions $G(x, d), \mathscr{A}(x), f(x, d), g(x, d)$ and the completeness of the set of basis functions, $\phi_k(z)$, and it is essential for computing an accurate solution for (P) based on a low-dimensional nonlinear program (see algorithm in the next section). The reader may also refer to the recent works (Dennis,

ElAlem & Maciel, 1997; Alexandrov, Dennis, Lewis & Torczon, 1998; Dennis, ElAlem & Williamson, 1999) which present new approaches, based on the trust region idea from nonlinear programming, for studying global convergence of low-order approximate programs of varying accuracy to high-order optimization problems. The proof of the theorem is given in the appendix.

Theorem 1. *Let J_{opt}^N be the optimal solution of the finite-dimensional nonlinear program of Eq.* (36) *and let J_{opt} be the optimal solution of* (P). *Then, under the assumption that the functions $G(x, d), \mathscr{A}(x), f(x, d), g(x, d)$ are continuously differentiable and the set of basis functions, $\phi_k(z)$ is complete, the following result holds:*

$$\lim_{N \to \infty} J_{\text{opt}}^N = J_{\text{opt}}. \qquad (37)$$

Remark 5. When the number of basis functions, $\phi_k(z)$, required to obtain a good approximation (measured in a desired norm) of the solution of the system of partial differential equations (equality constraints), is small, then the weighting functions are usually chosen to be identical to the basis functions, in which case the method of weighted residuals reduces to Galerkin's method.

Remark 6. As a practical implementation note, we point out that even though it is theoretically expected that the use of more basis functions in the series expansion of Eq. (33) would improve the accuracy of the computed approximate model of Eq. (36), the use of empirical eigenfunctions corresponding to very small eigenvalues should be avoided because such eigenfunctions are contaminated with significant round-off errors.

5. Computation of optimal solution

The objective of this section is to provide a computationally efficient procedure for the computation of an accurate optimal solution of the infinite dimensional nonlinear program of Eq. (1). The central idea is to use standard successive quadratic programming (SQP) algorithms to solve various finite-dimensional approximations of (P) obtained through application of the method of weighted residuals with empirical eigenfunctions until the optimal solution is computed with the desired accuracy. SQP is used for its simplicity and efficiency in the context of low-dimensional programs (it requires fewer function evaluations than other methods, and it does not require feasible points at intermediate iterations and converges to optimal solution from an infeasible path). A brief description of the standard SQP algorithm is provided for completeness in Remark 7 below (see also Vasantharajan et al. (1990), Floudas and Panos (1992) and Biegler et al. (1995) for details and analysis of the algorithm). The validity of the optimal solution computed by SQP is checked by verifying the optimality conditions of Eqs. (6)–(11). We

formulate the procedure used for the computation of the optimal solution of the infinite-dimensional program (P) in the form of the following algorithm:

Step 1: Compute an initial guess for N based on the importance of the eigenvalues corresponding to the empirical eigenfunctions.

Step 2: Use the model reduction procedure of Section 4 to derive the corresponding finite-dimensional program of Eq. (36).

Step 3: Solve the resulting finite-dimensional program through SQP to compute an optimal solution.

Step 4: Use Eq. (33) with $K = 1, \ldots, N$ to compute the optimal solution corresponding to the infinite-dimensional program and examine if this solution satisfies the optimality conditions of Eqs. (6)–(11).

Step 5: If the optimality conditions are satisfied, then stop. If not, then go back to Step 2 and perform model reduction with $N = N + 1$.

The structure of the above algorithm is motivated by the fact that the discrepancy between the infinite-dimensional program (P) and its finite-dimensional approximation of Eq. (36) decreases, as the number of basis functions, N, used in the expansion of Eq. (33) increases (at least, up to the point where round-off errors are not important). This is a consequence of the hierarchy of the empirical eigenfunctions obtained from K–L expansion; that is the first eigenfunction, ϕ_1, corresponds to the largest eigenvalue, λ_1, and is the most typical structure (in a sense made precise in Eq. (18)) of the members of the snapshots, $\{\bar{v}_k\}$, while the second eigenfunction, ϕ_2, corresponds to the next largest eigenvalue, λ_1, and is the next most typical structure, and so on. On the other hand, the convergence of the above algorithm is a direct consequence of the result of Theorem 1.

Remark 7. Utilizing a vector notation, the finite-dimensional program of Eq. (36) can be written in the following compact form:

$$\min \quad F(x)$$
$$\text{s.t} \quad h(x) = 0, \tag{38}$$
$$g(x) \leqslant 0.$$

The computation of an optimal solution of the above nonlinear program via SQP involves the iterative solution of the following quadratic problem which includes linear equality constraints (Biegler, Grossman & Westerberg, 1997):

$$\min \quad (\nabla F(x^j)^{\mathrm{T}} v + \tfrac{1}{2} v^{\mathrm{T}} B^j v),$$
$$\text{s.t.} \quad h(x^j) + \nabla h(x^j)^{\mathrm{T}} v = 0, \tag{39}$$
$$g(x^j) + \nabla g(x^j)^{\mathrm{T}} v \leqslant 0,$$

where x^j is the vector of variables for the jth iteration of SQP, B^j is Hessian approximation with BFGS, and $v = \Delta x$. The solution of the quadratic program of Eq. (39)

can be readily done by using standard solvers (see Biegler et al., (1997) for details).

Remark 8. We note that even though, for simplicity, we chose to use SQP for solving the finite-dimensional nonlinear programs obtained from the spatial discretization, other local (e.g., reduced gradient, etc.) as well as global optimization algorithms (e.g., Manousiouthakis & Sourlas, 1992; Floudas & Panos, 1992) can be used to solve the finite-dimensional programs.

Remark 9. We note that an alternative approach to perform reduction of the order of the infinite-dimensional program of Eq. (4) is to initially apply to this program finite differences/finite elements to derive a very high-order finite-dimensional nonlinear program, and then employ reduced-basis methods (Rheinboldt, 1993; Rabier & Rheinboldt, 1995) or dual variable methods (Hall, Porsching & Mesina, 1992; Chou & Porsching, 1998) to efficiently solve the resulting approximate program. Furthermore, modified finite-difference schemes (based, for example, on Pade' approximants) have been also used to efficiently solve third- and fourth-order boundary value problems (Al-Said, Noor & Rassias, 1998; Al-Said & Noor, 1998).

5.1. *Illustrative example*: *catalytic rod*

We now apply the proposed optimization method to the catalytic rod example introduced in Section 2.2. The values of the dimensionless parameters of the process used in the calculations are $\alpha = 2.0$, $\beta = 0.3$, $\gamma = 0.5$, $\lambda_1 = 1.0$, $\lambda_2 = 100.0$ and $v = 0.8$; the values of the constraints are $x_{1\min} = -1.0$, $x_{1\max} = 0$, $x_{2\min} = -0.5$, $x_{2\max} = 0.2$, $u_{\min} = 0.5$ and $u_{\max} = 2.0$; and the locations of the point actuators are: $z_j = 0.1, 0.3, 0.5, 0.7, 0.9$.

Since our objective is to test the computational efficiency and accuracy of the proposed optimization method, we initially solved the optimization problem (PE) by directly applying a second-order accurate finite difference scheme to reduce it to a finite-dimensional problem. Specifically, we used 206 discretization points to derive a finite-dimensional program which includes 412 state and 5 design variables, 412 equality constraints and 834 inequality constraints. We note that even though the number of discretization points, 206, used to discretize the nonlinear program of Eq. (4) is very large (owing to the poor convergence properties of the finite-difference scheme), the computation of an accurate (i.e., independent of the discretization) solution is critical for the precise evaluation of the performance of the proposed method. The adequacy of 206 discretization points to yield an accurate solution is established in Fig. 2, where the profiles of the rod temperature and concentration of species *A* for 206 and 412 discretization points are compared and are shown to be almost identical. The
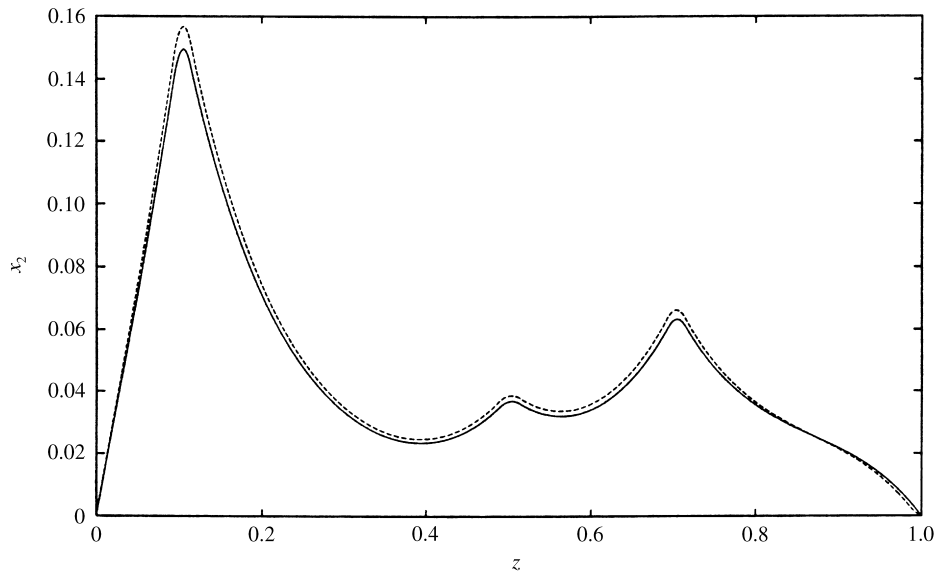
Fig. 2. Comparison of rod temperature profiles obtained from finite difference method with 206 (dashed line) and 412 (solid line) discretization points. Point actuator values: $u_1 = 2.0$, $u_2 = 0.5$, $u_3 = 0.7$, $u_4 = 1.0$ and $u_5 = 0.5$.

finite-dimensional nonlinear program obtained from the discretization was solved by using SQP. The quadratic problems in the SQP method were solved by using MINOS (Biegler et al., 1997). The computation of the optimal solution required 22 min and led to $u_1 = 2.0$, $u_2 = 1.654$, $u_3 = 1.533$, $u_4 = 1.654$, $u_5 = 2.0$, and $J_{\text{opt}} = 0.6573$ (note that, as expected, owing to the symmetry of the process with respect to the center of the rod $u_1 = u_5$ and $u_2 = u_4$).

We now use the proposed method to solve (PE). We initially construct an ensemble of solutions (snapshots) of the process model (equality constraints of (PE)) by taking a snapshot for each design variable at 5 equally spaced values between its maximum and minimum value; this leads to $5^3 = 125$ snapshots which give us 125 sets of values for x_1 and x_2. We apply Karhunen–Loéve expansion to this set of snapshots to compute 4 empirical eigenfunctions for concentration ($\phi_{1k}(z)$) and 4 for temperature ($\phi_{2k}(z)$); they capture more than 99% of the energy of the corresponding ensemble of snapshots, for both cases. The first three concentration and temperature empirical eigenfunctions are shown in Figs. 3 and 4, respectively; both sets of eigenfunctions are orthogonal. Expanding $x_1(z)$ and $x_2(z)$ as linear combinations of the empirical eigenfunctions

$$x_1(z) = \sum_{k=1}^{4} a_{1k} \phi_{1k}(z), \quad x_2(z) = \sum_{k=1}^{4} a_{2k} \phi_{2k}(z), \tag{40}$$

where a_{1k} and a_{2k} are constant coefficients, substituting into the optimization problem and taking the inner product of the PDE equality constraints with the empirical eigenfunctions, we obtain

$$\min \left(- \int_0^1 e^\tau \sum_{k=1}^{4} a_{1k} \phi_{1k}(z) \, \mathrm{d}z \right),$$

s.t.

$$0 = \int_0^1 A_1(z) \phi_{1k}(z) \, \mathrm{d}z - a_{1k} \lambda_1 \int_0^1 \phi_{1k}^2(z) \, \mathrm{d}z$$
$$- \lambda_1 \alpha \int_0^1 Q(z) \phi_{1k}(z) \, \mathrm{d}z \quad (k = 1, \ldots, 4),$$

$$0 = \int_0^1 A_2(z) \phi_{2k}(z) \, \mathrm{d}z - a_{2k} \lambda_2 \int_0^1 \phi_{2k}^2(z) \, \mathrm{d}z$$
$$+ \lambda_2 \gamma \int_0^1 Q(z) \phi_{2k}(z) \, \mathrm{d}z$$
$$- 0.5 \lambda_2 \int_0^1 \phi_{2k}(z) \, \mathrm{d}z + \lambda_2 \int_0^1 u(z) \phi_{2k}(z) \, \mathrm{d}z$$
$$(k = 1, \ldots, 4),$$

$$- 1.0 \leqslant \sum_{k=1}^{4} \alpha_{1k} \phi_{1k}(z) \leqslant 0, \quad \forall z$$

$$- 0.5 \leqslant \sum_{k=1}^{4} \alpha_{2k} \phi_{2k}(z) \leqslant 0.2, \quad \forall z$$

$$0.5 \leqslant u_j \leqslant 2.0 \quad (j = 1, \ldots, 5), \tag{41}$$

where

$$A_1(z) = \sum_{k=1}^{4} \left(\alpha_{1k} \frac{\mathrm{d}^2 \phi_{1k}(z)}{\mathrm{d}z^2} \right),$$

$$A_2(z) = \sum_{k=1}^{4} \left(\alpha_{2k} \frac{\mathrm{d}^2 \phi_{2k}(z)}{\mathrm{d}z^2} \right),$$

$$Q(z) = \left(\sum_{k=1}^{4} \alpha_{1k} \phi_{1k}(z) + 1 \right) e^\tau,$$

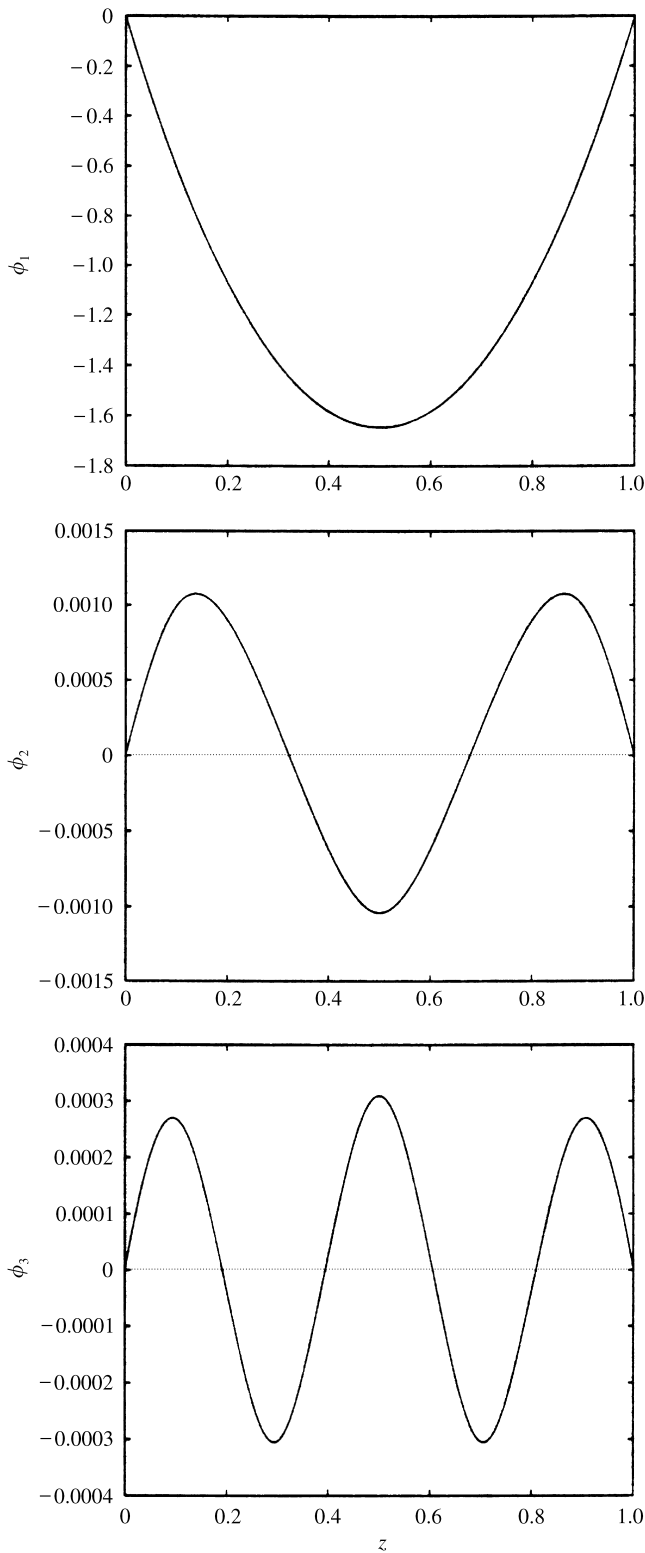$$\tau = \frac{-\beta}{\sum_{k=1}^{4} \alpha_{2k} \phi_{2k}(z) + 0.8}. \tag{42}$$

Fig. 3. First three concentration empirical eigenfunctions — catalytic rod.

Fig. 4. First three temperature empirical eigenfunctions — catalytic rod.

The above problem includes 8 state and 5 design variables, 8 equality constraints and 834 inequality constraints and was solved by SQP in 7.2 s. The results
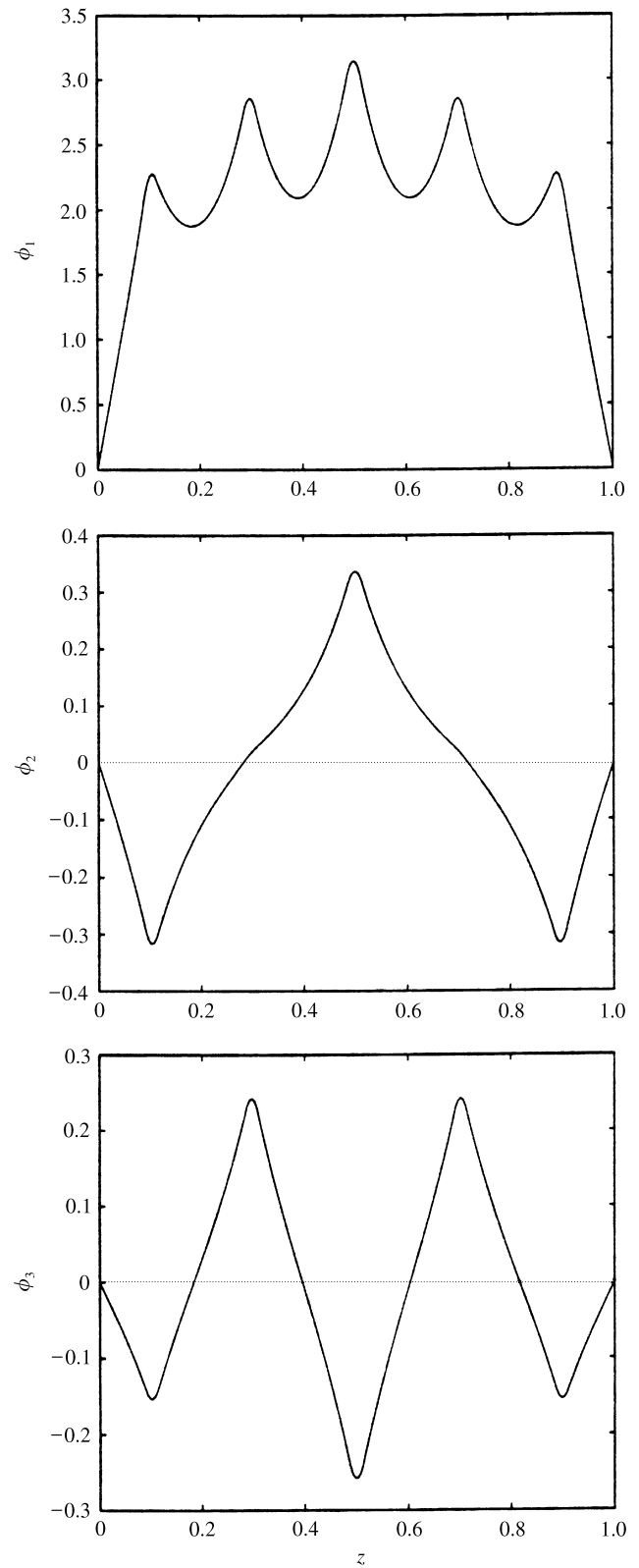
are $u_1 = 2.0$, $u_2 = 1.653$, $u_3 = 1.533$, $u_4 = 1.653$ and $u_5 = 2.0$. Using these values in the detailed finite difference model, we obtained $J_{emp} = 0.6572$. It is clear that

the proposed approach yields an accurate solution to (PE), while requiring a significantly smaller computational time compared to the conventional approach (spatial discretization with finite differences). This drastic reduction in CPU time is due to the fact that the dimensionality of the model obtained by Galerkin's method with empirical eigenfunctions is much smaller than that of the model obtained by finite differences. The optimal

rod temperature, concentration of A and reaction rate profiles obtained by using the two different approaches are given in Fig. 5. In both cases, the maximum temperature constraints are satisfied and the results are almost identical.

Remark 10. We note that in the case of point actuation which influences the system at z_0, the function $\delta(z - z_0)$ is assumed to have the finite value $1/2\mu$ in the interval $[z_0 - \mu, z_0 + \mu]$ (where μ is a small positive real number) and be zero elsewhere in the domain of definition of z.

6. Application to a packed-bed reactor

We consider a non-isothermal packed-bed reactor, where an elementary exothermic reaction of the form $A \to B$ takes place. Due to the exothermicity of the reaction, a jacket with three independent cooling zones is used around the reactor to remove heat (see Fig. 6 for a schematic of the process). Under standard modeling assumptions, a steady-state model of the process expressed in dimensionless variables takes the form in Ray (1981).

$$0 = -\rho_f c_{pf} v \frac{\partial T}{\partial \bar{z}} + k \frac{\partial^2 T}{\partial \bar{z}^2} + (-\Delta H) k_0 \exp^{-E/RT} C_A$$
$$+ U_T (T_w - T), \tag{43}$$

$$0 = -v \frac{\partial C_A}{\partial \bar{z}} + D \frac{\partial^2 C_A}{\partial \bar{z}^2} - k_0 \exp^{-E/RT} C_A$$

subject to the boundary conditions

$$\bar{z} = 0, \quad \rho_f c_{pf} v (T - T_f) = k \frac{\partial T}{\partial \bar{z}},$$

$$v(C_A - C_{AF}) = D \frac{\partial C_A}{\partial \bar{z}}, \tag{44}$$

$$\bar{z} = l, \quad \frac{\partial T}{\partial \bar{z}} = \frac{\partial C_A}{\partial \bar{z}} = 0,$$

where T, C_A denote the temperature and concentration in the reactor, ρ_f, c_{pf}, v denote the density, heat capacity and velocity of the fluid, k, D denote the conductivity and diffusivity coefficients, $k_0, E, \Delta H$ denote the pre-exponential factor, the activation energy and the heat of the
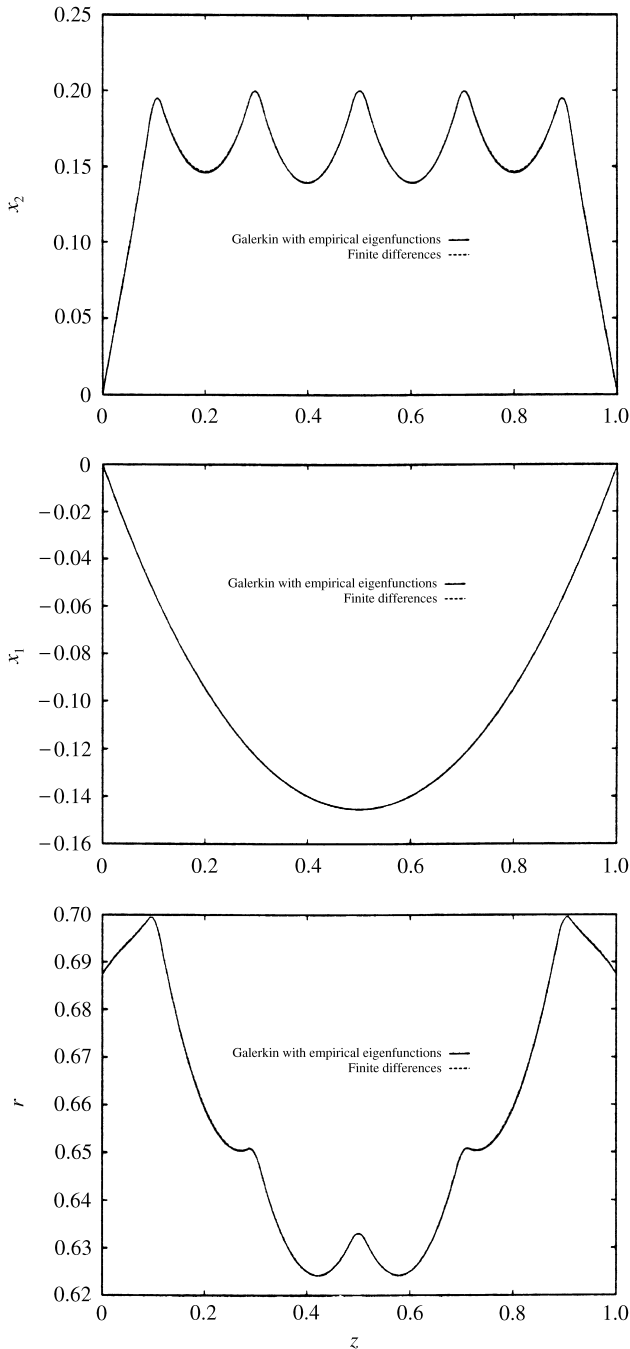

Fig. 5. Optimal temperature (top plot), concentration (middle plot) and reaction rate (bottom plot) profiles — catalytic rod. Comparison between Galerkin (dashed line) and finite-difference (solid line) models.
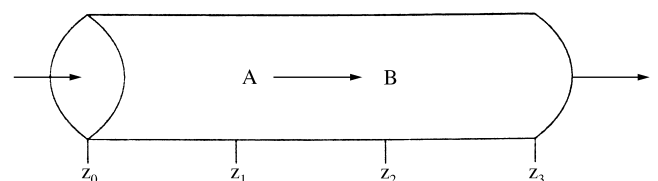

Fig. 6. Packed-bed reactor with three distributed actuators.

reaction, U_T denotes the heat-transfer coefficient between the wall and the bulk of the reactor, l denotes the length of the reactor, T_w denotes the wall temperature, and C_{AF}, T_f denote some reference values for concentration and temperature.

The optimization problem is to maximize the rate of production of species B by choosing the temperature of the three independent cooling zones. The reactor temperature and the temperatures of the cooling zones at all positions are required to be lower than certain maximum values. We note that the even though T_w cannot be adjusted directly in practice, but indirectly through adjustment of the jacket inlet flow rate, the computation of the value of the jacket inlet flow rate which yields the optimal T_w can be easily done by utilizing a steady-state energy balance (which is an algebraic equation) for the jacket. Mathematically, this optimization problem is formulated as follows:

$$\min C_A(l),$$

s.t.

$$0 = -v\frac{\partial C_A}{\partial \bar{z}} + D\frac{\partial^2 C_A}{\partial \bar{z}^2} - k_0 \exp^{-E/RT} C_A,$$

$$0 = -\rho_f c_{pf} v\frac{\partial T}{\partial \bar{z}} + k\frac{\partial^2 T}{\partial \bar{z}^2}$$
$$+ (-\Delta H)k_0 \exp^{-E/RT} C_A + U_T(T_w - T),$$

$$T_J = \sum_{j=1}^{5} [H(\bar{z}_j - \bar{z}_{j-1}) - H(\bar{z}_{j+1} - \bar{z}_j)]T_{Jj},$$

at $\bar{z} = 0$, $\quad \rho_f c_{pf} v(T - T_f) = k\frac{\partial T}{\partial \bar{z}}$,

$$v(C_A - C_{AF}) = D\frac{\partial C_A}{\partial \bar{z}},$$

at $\bar{z} = l$, $\quad \dfrac{\partial T}{\partial \bar{z}} = \dfrac{\partial C_A}{\partial \bar{z}} = 0,$

$$T_{\min} \leqslant T(\bar{z}) \leqslant T_{\max}, \quad \forall \bar{z},$$

$$T_{J\min} \leqslant T_{Jj} \leqslant T_{J\max} \quad (j = 1, \ldots, 3),$$

where $H(\cdot)$ is the standard Heaviside function.

To facilitate our development, we will use the dimensionless variables

$$z = \frac{\bar{z}}{l}, \quad Pe_T = \frac{\rho_f c_{pf} vl}{k}, \quad Pe_C = \frac{vl}{D},$$

$$\gamma = \frac{E}{RT_0}, \quad \beta_T = \frac{U_T l}{\rho_f c_{pf} v}, \quad B_C = \frac{lk_0 \exp^{-\gamma}}{v},$$

$$B_T = \frac{(-\Delta H)C_{Af} lk_0 \exp^{-\gamma}}{\rho_f c_{pf} T_f v}, \quad (46)$$

$$x_2 = \frac{T - T_f}{T_f}, \quad x_1 = \frac{C_A - C_{AF}}{C_{AF}}, \quad u = \frac{T_w - T_{ws}}{T_f},$$

where T_{ws} denotes some steady-state profile for the wall temperature, to put the above optimization problem into the following dimensionless form:

$$\min x_1(1)$$

s.t.

$$0 = -\frac{\partial x_2}{\partial z} + \frac{1}{Pe_T}\frac{\partial^2 x_2}{\partial z^2} + B_T B_C \exp^{\gamma x_2/1 + x_2} x_1$$
$$+ \beta_T(u - x_2),$$

$$0 = -\frac{\partial x_1}{\partial z} + \frac{1}{Pe_C}\frac{\partial^2 x_1}{\partial z^2} - B_C \exp^{\gamma x_2/1 + x_2} x_1,$$
$$\text{(PE2)}$$

$$u(z) = \sum_{j=1}^{5} [H(z_j - z_{j-1}) - H(z_{j+1} - z_j)]u_j,$$

at $z = 0$, $\quad Pe_T x_2 = \dfrac{\partial x_2}{\partial z}, \quad Pe_C x_1 = \dfrac{\partial x_1}{\partial z}$,

at $z = 1$, $\quad \dfrac{\partial x_2}{\partial z} = \dfrac{\partial x_1}{\partial z} = 0,$

$$x_{1\min} \leqslant x_1(z) \leqslant x_{1\max}, \quad \forall z,$$

$$x_{2\min} \leqslant x_2(z) \leqslant x_{2\max}, \quad \forall z,$$

$$u_{\min} \leqslant u_j \leqslant u_{\max} \quad (j = 1, \ldots, 3).$$

The values of the dimensionless constants were chosen as $B_T = 1.6$, $B_C = 0.1$, $Pe_T = Pe_C = 7.0$, $\beta_T = 2.0$, and $\gamma = 10.0$, while the values of the constraints were chosen as $x_{1\min} = -1.0$, $x_{1\max} = 0.0$, $x_{2\min} = -1.0$, $x_{2\max} = 0.8$, $u_{\min} = 0.0$ and $u_{\max} = 0.4$.

Following the same approach as in the catalytic rod example, we initially solve the optimization problem (PE2) by directly applying a second-order accurate finite-difference scheme to reduce it to a finite-dimensional problem, which we then solve using SQP. The quadratic problems in the SQP method were solved by using MINOS. Specifically, we used 102 discretization points to derive a finite-dimensional program which includes 204 state and 3 design variables, 204 equality constraints and 414 inequality constraints. We verified that further increase in the number of discretization points leads to negligible improvements in the accuracy of the results. The solution to this problem required 2 min 5 s and led to $u_1 = 0.2062$, $u_2 = 0.0783$, $u_3 = 0.40$, and $J_{opt} = -0.9772$.

We now use the proposed method to solve (PE2). We initially construct an ensemble of solutions (snapshots) of the process model (equality constraints of (PE2)) by taking a snapshot for each design variable at 9 equally spaced values between its maximum and minimum value; this leads to $9^3 = 729$ snapshots which give us 729 sets of values for x_1 and x_2. We apply Karhunen–Loéve expansion to this set of snapshots to compute 10 empirical

eigenfunctions for concentration ($\phi_{1k}(z)$) and 10 for temperature ($\phi_{2k}(z)$). The first three concentration and temperature empirical eigenfunctions are shown in Figs. 7 and 8, respectively. Both sets of eigenfunctions are orthogonal. Expanding $x_1(z)$ and $x_2(z)$ as linear combinations of the empirical eigenfunctions

$$x_1(z) = \sum_{k=1}^{10} a_{1k}\phi_{1k}(z), \quad x_2(z) = \sum_{k=1}^{10} a_{2k}\phi_{2k}(z), \qquad (47)$$

where a_{1k} and a_{2k} are constant coefficients, substituting into the optimization problem and taking the inner product of the PDE equality constraints with the empirical eigenfunctions, we obtain

$$\min \sum_{k=1}^{10} a_{1k}\phi_{1k}(1),$$

s.t.

$$0 = - \int_0^1 A_1(z)\phi_{1k}(z)\,\mathrm{d}z + \frac{1}{Pe_C}\int_0^1 B_1(z)\phi_{1k}(z)\,\mathrm{d}z$$

$$- B_C \int_0^1 Q(z)\phi_{1k}(z)\,\mathrm{d}z \quad (k = 1,\dots,10),$$

$$0 = - \int_0^1 A_2(z)\phi_{2k}(z)\,\mathrm{d}z + \frac{1}{Pe_T}\int_0^1 B_2(z)\phi_{2k}(z)\,\mathrm{d}z$$

$$+ B_T B_C \int_0^1 Q(z)\phi_{2k}(z)\,\mathrm{d}z \qquad (48)$$

$$- a_{2k}\beta_T \int_0^1 \phi_{2k}^2(z)\,\mathrm{d}z + \beta_T \int_0^1 u(z)\phi_{2k}(z)\,\mathrm{d}z$$

$$\times (k = 1,\dots,10)$$

$$-1.0 \leqslant \sum_{k=1}^{10} \alpha_{1k}\phi_{1k}(z) \leqslant 0, \quad \forall z$$

$$-1.0 \leqslant \sum_{k=1}^{10} \alpha_{2k}\phi_{2k}(z) \leqslant 0.8, \quad \forall z$$

$$0 \leqslant u_j \leqslant 0.4 \quad (j = 1,\dots,3),$$

where

$$A_1(z) = \sum_{k=1}^{10}\left(\alpha_{1k}\frac{\mathrm{d}\phi_{1k}(z)}{\mathrm{d}z}\right), \quad B_1(z) = \sum_{k=1}^{10}\left(\alpha_{1k}\frac{\mathrm{d}^2\phi_{1k}(z)}{\mathrm{d}z^2}\right),$$

$$A_2(z) = \sum_{k=1}^{10}\left(\alpha_{2k}\frac{\mathrm{d}\phi_{2k}(z)}{\mathrm{d}z}\right), \quad B_2(z) = \sum_{k=1}^{10}\left(\alpha_{2k}\frac{\mathrm{d}^2\phi_{2k}(z)}{\mathrm{d}z^2}\right),$$

$$Q(z) = \left(\sum_{k=1}^{10}\alpha_{1k}\phi_{1k}(z)\right)e^\tau, \quad \tau = \frac{\gamma\sum_{k=1}^{10}\alpha_{2k}\phi_{2k}(z)}{\sum_{k=1}^{10}\alpha_{2k}\phi_{2k}(z)+1}. \qquad (49)$$

The above problem includes 20 state and 3 design variables, 20 equality constraints and 414 inequality constraints and was solved by SQP in 41 s. The results are
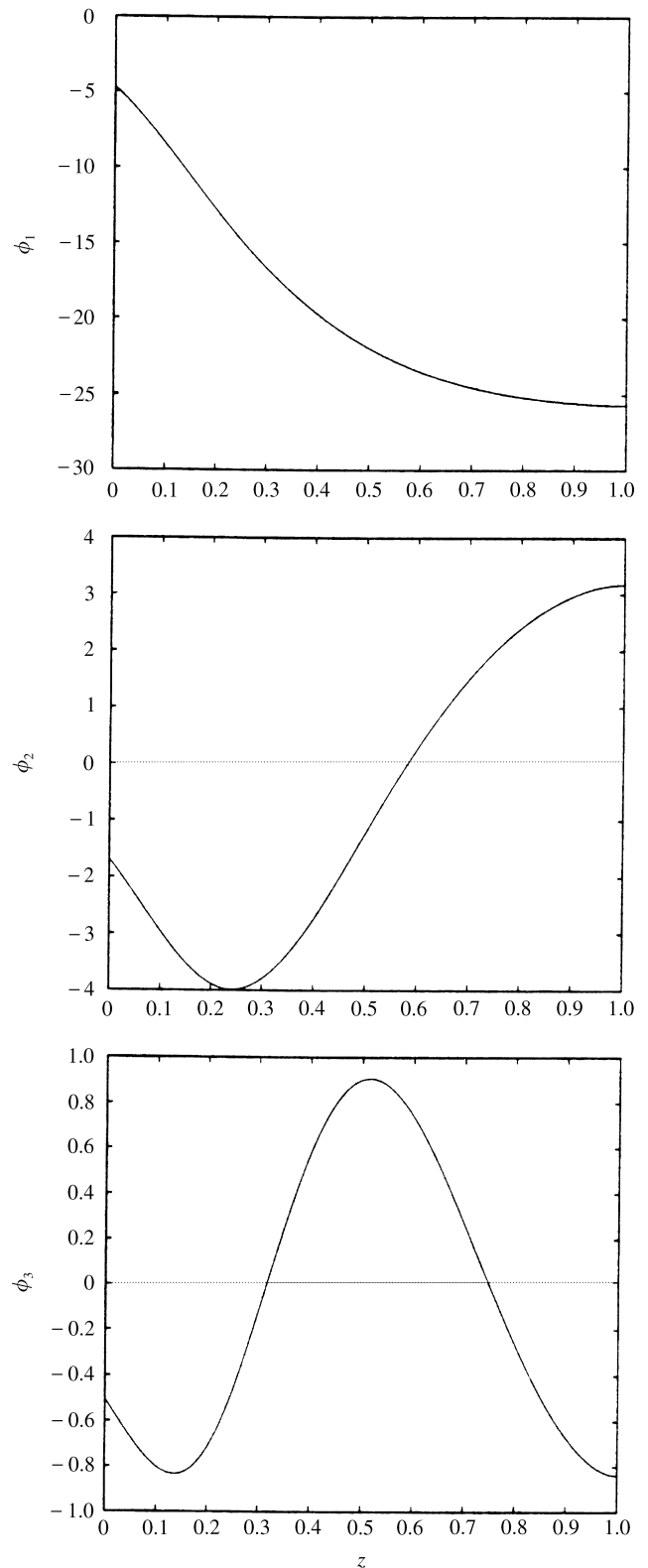

Fig. 7. First three concentration empirical eigenfunctions — packed-bed reactor.

$u_1 = 0.2001$, $u_2 = 0.0876$, $u_3 = 0.40$. Using these values in the detailed finite-difference model, we obtained $J_{\mathrm{emp}} = -0.9770$. It is clear that the proposed approach
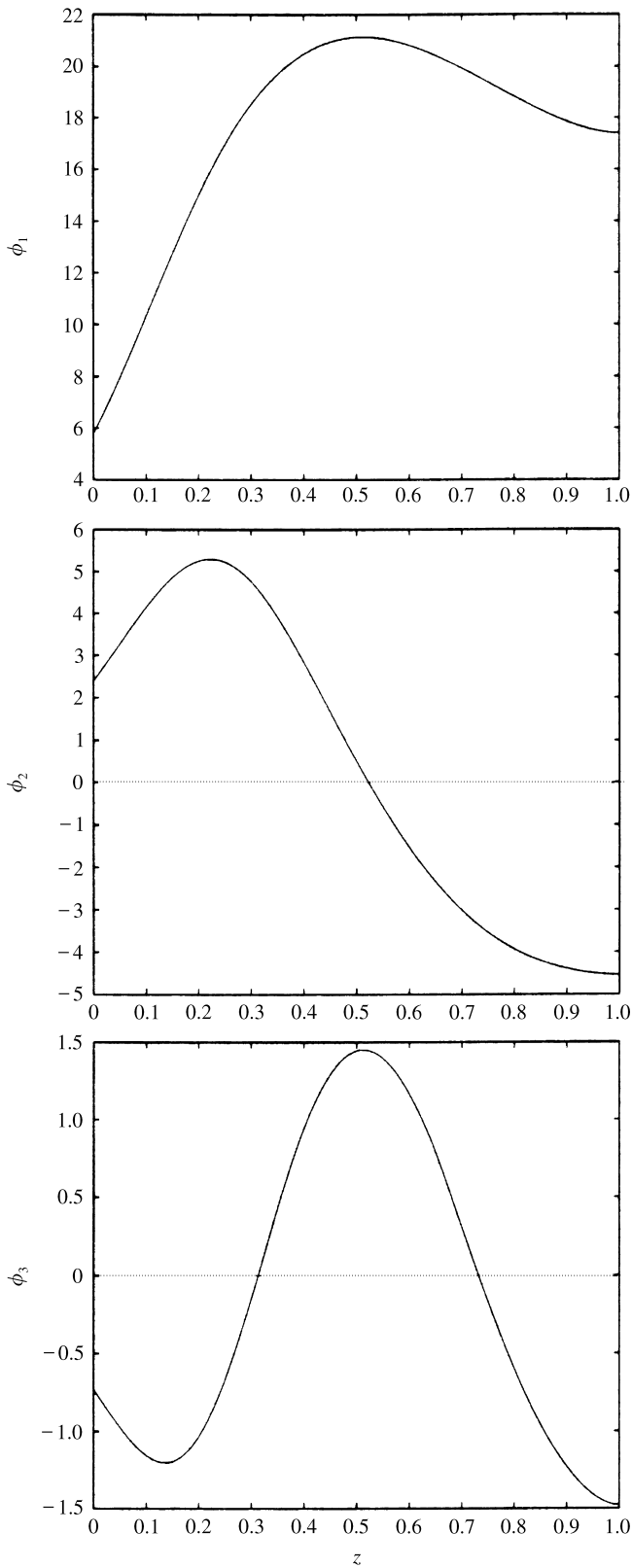
Fig. 8. First three temperature empirical eigenfunctions — packed-bed reactor.
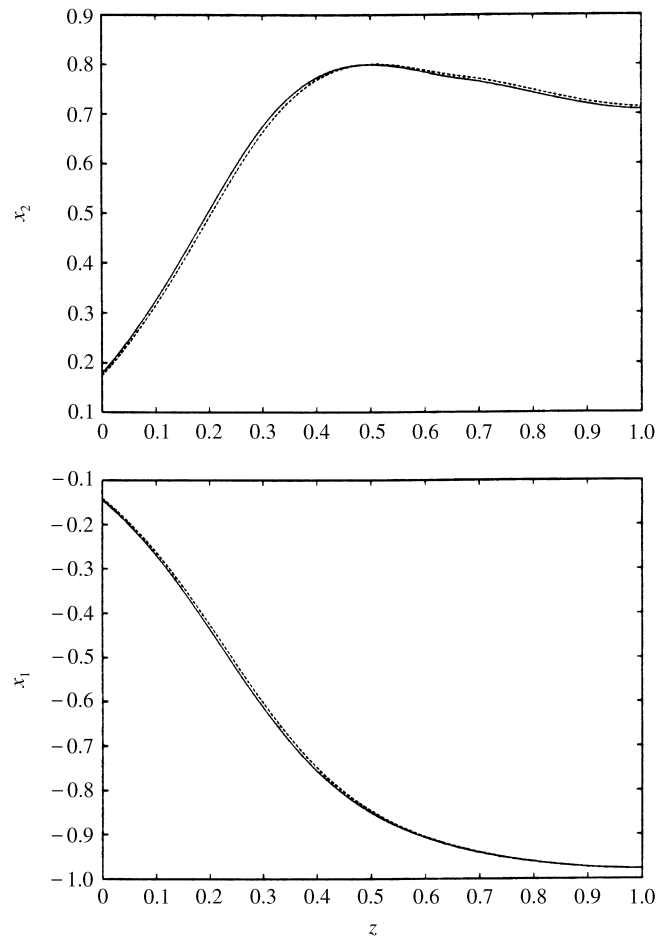

Fig. 9. Optimal temperature (top plot) and concentration (bottom plot) profiles — packed-bed reactor. Comparison between Galerkin (dashed line) and finite-difference (solid line) models.

yields an accurate solution to (PE), while requiring a significantly smaller computational time compared to the conventional approach (spatial discretization with finite

differences). The optimal reactor temperature and concentration of species A obtained by using these two different approaches are given in Fig. 9. In both cases, the maximum reactor and jacket temperature constraints are satisfied and the results are very close.

Remark 11. Referring to both simulation studies, the following remarks are in order: (a) the nonlinear finite-dimensional approximate programs obtained through finite-difference discretization and empirical eigenfunction-based Galerkin discretization were solved with the same successive quadratic programming algorithm (no sparse matrix techniques were employed) to provide a clear picture of the advantage obtained by using eigenfunction-based Galerkin discretization, and (b) the time needed to construct the ensemble of snapshots was not included in the calculation of the time needed to solve the optimization problem through empirical eigenfunction-based Galerkin method because we only use the computed empirical eigenfunctions to solve one optimization problem (in both applications,

this time was found to be smaller than the time needed to solve the infinite-dimensional program using finite differences).

Remark 12. We note that even though in both examples we treat the concentration and temperature data separately to compute the sets of empirical eigenfunctions for concentration and temperature, one could also construct empirical eigenfunctions based on the whole vector of dependent variables (appropriately scaled), rather than concentration and temperature separately. However, the significant reduction achieved in the computational time needed to solve the optimization problem with the computed eigenfunctions (note that the resulting finite-dimensional programs have just eight and 20 equality constraints, respectively) does not leave much room for further reduction of the computational time by reducing the number of eigenfunctions required to obtain an accurate solution through an alternative treatment of the ensemble of snapshots, and therefore, such an alternative treatment is not studied here. In addition, the separate treatment of concentration and temperature data and the computation of separate sets of eigenfunctions for concentration and temperature is more intuitive physically and less complex computationally.

7. Conclusions

This article presented a computationally efficient method for the solution of optimization problems arising in the context of transport-reaction processes governed by nonlinear elliptic partial differential equations. The central idea of the method was to discretize the infinite-dimensional optimization problem by utilizing the method of weighted residuals with empirical eigenfunctions obtained by applying Karhunen–Loéve expansion to an appropriately constructed ensemble of solutions of the PDE equality constraints for different values of the design variables. This model reduction procedure leads to low-dimensional, nonlinear programs which represent accurate approximations of the infinite-dimensional nonlinear program and for which the solution can be obtained through standard optimization algorithms with small computational effort. The key issues of construction of the ensemble used for the computation of the empirical eigenfunctions and validity of the optimal solutions computed from the finite-dimensional programs were also addressed. The proposed method was successfully applied to two chemical process examples and was found to be more efficient computationally with respect to conventional optimization approaches based on spatial discretization with the finite difference method.

Acknowledgements

Appendix

Proof of Theorem 1. We initially define the error variable $e_{kN} = a_k - a_{kN}$ and rewrite the nonlinear program of Eq. (36) in the following form:

$$J_{\text{opt}}^N = \min \int_\Omega G\left(\sum_{k=1}^N (a_k + e_{kN})\phi_k(z), d\right)\mathrm{d}z,$$

s.t

$$0 = \int_\Omega \psi_v(z)\mathscr{A}\left(\sum_{k=1}^N (a_k + e_{kN})\phi_k(z)\right)\mathrm{d}z$$

$$+ \int_\Omega \psi_v(z)f\left(\sum_{k=1}^N (a_k + e_{kN})\phi_k(z), d\right)\mathrm{d}z, \quad (A.1)$$

$$v = 1, \ldots, N,$$

$$\int_\Omega \psi_v(z)g\left(\sum_{k=1}^N (a_k + e_{kN})\phi_k(z), d\right)\mathrm{d}z \leqslant 0,$$

$$v = 1, \ldots, N.$$

Since the set of functions, $\phi_1(z), \phi_2(z), \ldots, \phi_k(z)$, is orthogonal and complete and the functions $\mathscr{A}(x), f(x, d)$ are smooth, the following result holds for any $d \in \mathbb{R}^p$:

$$\lim_{N \to \infty} \int_\Omega \psi_v(z)\left(\left[\mathscr{A}\left(\sum_{k=1}^\infty a_k\phi_k(z)\right) + f\left(\sum_{k=1}^\infty a_k\phi_k(z), d\right)\right]\right.$$

$$\left. - \left[\mathscr{A}\left(\sum_{k=1}^N a_{kN}\phi_k(z)\right) + f\left(\sum_{k=1}^N a_{kN}\phi_k(z), d\right)\right]\right)\mathrm{d}z = 0$$

$$(A.2)$$

which implies that the following limit holds for any $d \in \mathbb{R}^p$:

$$\lim_{N \to \infty} a_{kN} = a_k, \quad \forall k = 1, \ldots, \infty. \quad (A.3)$$

Taking the limit as $N \to \infty$ of the nonlinear program of Eq. (A.1), we obtain

$$\lim_{N \to \infty} J_{\text{opt}}^N = \lim_{N \to \infty} \min \int_\Omega G\left(\sum_{k=1}^N (a_k + e_{kN})\phi_k(z), d\right)\mathrm{d}z,$$

s.t

$$0 = \lim_{N \to \infty} \int_\Omega \psi_v(z)\mathscr{A}\left(\sum_{k=1}^N (a_k + e_{kN})\phi_k(z)\right)\mathrm{d}z$$

$$+ \int_\Omega \psi_v(z)f\left(\sum_{k=1}^N (a_k + e_{kN})\phi_k(z), d\right)\mathrm{d}z,$$

$$v = 1, \ldots, N,$$

$$\lim_{N \to \infty} \int_\Omega \psi_v(z) g\left(\sum_{k=1}^{N} (a_k + e_{kN})\phi_k(z), d \right) dz \leqslant 0,$$

$$v = 1, \dots, N \tag{A.4}$$

or

$$\lim_{N \to \infty} J_{opt}^N = \min \int_\Omega G\left(\sum_{k=1}^{\infty} \left(a_k + \lim_{N \to \infty} e_{kN} \right)\phi_k(z), d \right) dz,$$

s.t

$$0 = \int_\Omega \psi_v(z) \mathscr{A}\left(\sum_{k=1}^{\infty} \left(a_k + \lim_{N \to \infty} e_{kN} \right)\phi_k(z) \right) dz$$

$$+ \int_\Omega \psi_v(z) f\left(\sum_{k=1}^{\infty} \left(a_k + \lim_{N \to \infty} e_{kN} \right)\phi_k(z), d \right) dz, \quad \text{(A.5)}$$

$$v = 1, \dots, \infty$$

$$\int_\Omega \psi_v(z) g\left(\sum_{k=1}^{\infty} \left(a_k + \lim_{N \to \infty} e_{kN} \right)\phi_k(z), d \right) dz \leqslant 0,$$

$$v = 1, \dots, \infty$$

and using Eq. (A.3), we finally obtain

$$\lim_{N \to \infty} J_{opt}^N = \min \int_\Omega G\left(\sum_{k=1}^{\infty} (a_k)\phi_k(z), d \right) dz = J_{opt}$$

s.t

$$0 = \int_\Omega \psi_v(z) \mathscr{A}\left(\sum_{k=1}^{\infty} a_k \phi_k(z) \right) dz \tag{A.6}$$

$$+ \int_\Omega \psi_v(z) f\left(\sum_{k=1}^{\infty} a_k \phi_k(z), d \right) dz, \quad v = 1, \dots, \infty,$$

$$\int_\Omega \psi_v(z) g\left(\sum_{k=1}^{\infty} a_k \phi_k(z), d \right) dz \leqslant 0, \quad v = 1, \dots, \infty,$$

from which the result of the theorem follows directly. □

References

Al-Said, E. A., & Noor, M. A. (1998). Numerical solutions of a system of fourth order boundary value problems. *International Journal of Computer Mathematics*, *70*, 347–355.

Al-Said, E. A., Noor, M. A., & Rassias, T. M. (1998). Numerical solutions of third-order obstacle problems. *International Journal of Computer Mathematics*, *69*, 75–84.

Alexandrov, N. M., Dennis, J. E., Lewis, R. M., & Torczon, V. (1998). A trust-region framework for managing the use of approximation models in optimization. *Structural Optimization*, *15*, 16–23.

Baker, J., & Christofides, P. D. (1999). Nonlinear control of rapid thermal chemical vapor desposition under uncertainty. *Computers & Chemical Engineering*, *23*(s), 233–236.

Baker, J., & Christofides, P. D. (2000). Finite dimensional approximation and control of nonlinear parabolic PDE systems. *International Journal of Control*, *73*, 439–456.

Bangia, A. K., Batcho, P. F., Kevrekidis, I. G., & Karniadakis, G. E. (1997). Unsteady 2-D flows in complex geometries: Comparative bifurcation studies with global eigenfunction expansion. *SIAM Journal of Scientific Computing*, *18*, 775–805.

Bendersky, E., & Christofides, P. D. (1999). A computationally efficient method for optimization of transport-reaction processes. *Computers & Chemical Engineering*, *23*(s), 447–450.

Biegler, L. T., Grossman, I. E., & Westerberg, A. W. (1997). *Systematic methods of chemical process design*. Englewood cliffs, NJ: Prentice-Hall.

Biegler, L. T., Nocedal, J., & Schmid, C. (1995). Reduced hessian strategies for large-scale nonlinear programming. *SIAM Journal of Optimization*, *5*, 314.

Chou, S. H., & Porsching, T. A. (1998). A note on the restarted cg method and reduced space additive correction. *Computers & Mathematics with Applications*, *35*, 129–136.

Dennis, J. E., ElAlem, M., & Maciel, M. C. (1997). A global convergence theory for general trust-region-based algorithms for equality constrained optimization. *SIAM Journal on Optimization*, *7*, 177–207.

Dennis, J. E., ElAlem, M., & Williamson, K. (1999). A trust-region approach to nonlinear systems of equalities and inequalities. *SIAM Journal of Optimization*, *9*, 291–315.

Esposito, R. W., & Floudas, C. A., (1998). Global optimization of nonconvex problems with differential & algebraic constraints. *AICHE Annual meeting, paper 236b, Miami Beach, FL*.

Finlayson, B. A. (1980). *Nonlinear analysis in chemical engineering.* New York: McGraw-Hill.

Floudas, C. A., & Panos, M. P. (1992). *Recent advances in global optimization*. Princeton, NJ: Princeton University Press.

Fukunaga, K. (1990). *Introduction to statistical pattern recognition*. New York: Academic Press.

Graham, M. D., & Kevrekidis, I. G. (1996). Alternative approaches to the Karhunen-Loève decomposition for model reduction and data analysis. *Computers & Chemical Engineering*, *20*, 495–506.

Hall, C. A., Porsching, T. A., & Mesina, G. L. (1992). On a network method for unsteady incompressible fluid-flow on triangular grids. *International Journal of Numerical Methods in Fluids*, *15*, 1383–1406.

Holmes, P., Lumley, J. L., & Berkooz, G. (1996). *Turbulence, coherent structures, dynamical systems and symmetry*. New York: Cambridge University Press.

Lumley, J. L. (1981). Coherent structures in turbulence. In *Transition and turbulence* (pp. 215–242). New York: Academic Press.

Manousiouthakis, V., & Sourlas, D. (1992). Global optimization approach to rationally constrained rational programming problems. *Chemical Engineering Communications*, *115*, 127–147.

Park, H. M., & Cho, D. H. (1996). The use of the Karhunen-Loeve decomposition for the modeling of distributed parameter systems. *Chemical Engineering Science*, *51*, 81–98.

Park, H. M., & Lee, J. H. (1998). A method of solving inverse convection problems by means of model recuction. *Chemical Engineering Science*, *53*, 1731–1744.

Rabier, P. J., & Rheinboldt, W. C. (1995). On the numerical solution of the euler-lagrange equations. *SIAM Journal on Numerical Analysis*, *32*, 318–329.

Ray, W. H. (1981). *Advanced process control.*. New York: McGraw-Hill.

Rheinboldt, W. C. (1993). On the theory and error estimation of the reduced basis method of multiparameter problems. *Nonlinear Analysis-Theory Methods & Applications*, *21*, 849–858.

Shvartsman, S. Y., & Kevrekidis, I. G. (1998). Nonlinear model reduction for control of distributed parameter systems: A computer assisted study. *A.I.Ch.E. Journal*, *44*, 1579–1595.

Sirovich, L. (1987a). Turbulence and the dynamics of coherent structures: Part I: Coherent Structures. *Quarterly of Applied Mathematics*, *XLV*, 561–571.

Sirovich, L. (1987b). Turbulence and the dynamics of coherent structures: Part II: Symmetries and transformations. *Quarterly of Applied Mathematics*, *XLV*, 573–582.

Teman, R. (1988). *Infinite-dimensional dynamical systems in mechanics and physics*. New York: Springer.

Vasantharajan, S., Viswanathan, J., & Biegler, L. T. (1990). Reduced successive quadratic programming implementation for large-scale optimization problems with smaller degrees of freedom. *Computers & Chemical Engineering*, *14*, 907–915.

Yu, T. K., & Seinfeld, J. H. (1973). Observability and optimal measurement location in linear distribution parameter systems. *International Journal of Control*, *18*, 785–799.