A Low-Computational-Cost Strategy to Localize Points in the Slow Manifold Proximity for Isothermal Chemical Kinetics

A low-computational-cost strategy to localize points in the slow manifold proximity for isothermal chemical kinetics

Alessandro Ceccato[*], Paolo Nicolini[†], Diego Frezzato [‡]

Abstract

Dimensionality reduction for the modeling of reacting chemical systems can represent a fundamental achievement both for a clear understanding of the complex mechanisms under study, and also for the practical calculation of quantities of interest. To tackle the problem, different approaches have been proposed in the literature. Among them, particular attention has been devoted to the exploitation of the so-called slow manifolds (SMs). These are lower-dimensional hypersurfaces where the slow part of the evolution takes place. In this study we present a low-computational-cost algorithm (based on a previously developed theoretical framework) for the localization of candidate points in the proximity of the SM. A parallel implementation (called DRIMAK) of such an approach has been developed and the source code is made freely available. We tested the performance of the code on two model schemes for hydrogen combustion, being able to localize points that fall very close to the perceived SM with limited computational effort. The method can provide starting points for other more accurate but computationally demanding strategies; this can be a great help especially when no information about the SM is available *a priori* and very many species are involved in the reaction mechanism.

Keywords: Chemical kinetics, slow manifolds, mass-action based kinetics, dimensional reduction.

[*]Dipartimento di Scienze Chimiche, Università degli Studi di Padova, via Marzolo 1, I-35131, Padova, Italy.

[†]Department of Control Engineering - K335, Faculty of Electrical Engineering, Czech Technical University in Prague, Karlovo náměstí 13, 121 35, Prague 2, Czech Republic.

[‡]Dipartimento di Scienze Chimiche, Università degli Studi di Padova, via Marzolo 1, I-35131, Padova, Italy. Corresponding author. Email: diego.frezzato@unipd.it

Introduction

When dealing with mechanisms involving complex reaction schemes or parallel elementary reactions, simplification of the chemical kinetics description is often needed. Even in the simplest case of a constant-volume and well-stirred isothermal medium, such that the mass-action law is applicable to express the progression rate of the elementary reactions[1], the number of dynamical variables to account for (*i.e.*, the volumetric concentrations of the species involved) can be so large that the numerical integration of the evolution equations becomes a hard task (especially in the case of stiff kinetics) and, crucially, the physical understanding of the whole process is obscured. In this work we make a step forward the identification of the so-called slow manifolds (SMs in the following[a]) which are basically hypersurfaces, of *lower* dimension than that of the full concentration space, where the slow part of the system's evolution takes place. Given a global reactive process, the identification of points on its SM (if present), and their interpolation, would allow one to subsequently attain a *reduced* description of the kinetics, focusing only on the slow phase. Based on our previous theoretical works, here we provide a strategy, along with the first implementation in an open source C++ software package and related tests, to produce good candidate points to the SM proximity with very low computational cost. Other existing strategies for the SM construction (see below) could be integrated with our method in order to make a post-production screening of the solutions and to perform further refinement steps. Such a combination of strategies may be particularly useful when the dimensionality of the SM and its approximate localization in the full concentration space are unknown.

For a constant-temperature and well-stirred medium, the mean-field approach based on the so-called "mass action law" provides the mathematical description of the macroscopic chemical kinetics for a set of N species involved in M elementary reactive processes (which can be either the steps of the mechanism of a single complex reaction, or competing elementary reactions)[1]. The mathematical format consists of an autonomous set of N polynomial ordinary differential equations (ODEs) for the volumetric concentrations taken as

[a]We like to indicate that the abbreviation SIM for "slow invariant manifold" is frequently used in the specialistic literature. We prefer to use SM in continuation of our previous works on this subject.

dynamical variables. In the following, the column vector $\mathbf{x}$ collects the concentrations x_j for $j = 1, \cdots, N$. The ODE system is

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}) \tag{1}$$

where $\mathbf{F}(\mathbf{x})$ is the state-dependent "velocity field" whose components will be made explicit in the next section.

As stated above, when N becomes large, as may happen for reaction mechanisms involving radical species or in the context of biochemical networks, the need for simplification of such a description by "reducing" the dimensionality of the problem becomes urgent. A large number of strategies has been devised to achieve such a goal. The matter is quite broad and a good starting point for an interested reader may be the review made by Okino and Mavrovouniotis[2] and references therein. Of particular relevance are the sensitivity analysis,[3,4] the lumping procedures,[5,6] the application of the quasi-stationary-state and quasi-equilibrium approximations,[7,8] and the exploitation of the existence of the so-called slow manifolds, which is the subject of this work.

As anticipated, a SM can be seen as the hypersurface, of lower dimensionality than that of the whole concentration space, towards which the trajectories $\mathbf{x}(t)$ of the reactive system approach in going to the stationary points. Actually, there could be a "cascade" of manifolds of ever-reducing dimensions;[9] we stress that the SM considered here is the ultimate manifold which is approached before reaching the equilibrium manifold (EM) formed by the stationary points. The "bundling" of the trajectories on the SM is a known trait which can be exploited to formulate a reduced description of the kinetics. In fact, it usually happens that the late and slowest part of the evolution takes place in the neighborhood of a SM. Thus, if one neglects the initial and fast (with respect to the subsequent dynamics) transients, the original system of ODEs projected on the SM would suffice to describe the slow part. In this respect, the localization of the SM, or at least of good candidate points in its proximity (and possibly their interpolation with suitable parametric hypersurfaces), would provide the ingredients to build a simplified kinetics description of reduced dimensionality.

Several conceptually heterogeneous strategies have been devised to construct the SMs in the context of chemical kinetics. An interested reader may find a comprehensive presentation in the introductions of Refs.[10–13] (see also our outline in Ref.[14] and references therein). In

short, the leading idea is that close to the SM the system's evolution is slower in comparison to points far from it.

Unfortunately, such a timescale separation between fast and slow components of the evolution is unequivocally defined only for linear kinetic schemes (*i.e.*, with only elementary reactions of the first order) for which the evolution law takes the form $\dot{\mathbf{x}} = -\mathbf{K}\mathbf{x}$, with $\mathbf{K}$ being some fixed kinetic matrix. In this case, the timescale separation (if present) is manifest in a gap between the real parts of the non-null eigenvalues of $\mathbf{K}$: if an eigenvalue is well separated by the larger ones in such a sense, the SM is the hyperplane identified by the eigenvector corresponding to such an eigenvalue and by the eigenvector(s) corresponding to the null eigenvalue(s)[14].

For non-linear kinetic schemes, the fast-slow separation becomes local and, to some extent, subjectively quantified. In such a general context, the SM is formally identified within the framework of Fenichel's geometric singular perturbation (GSP) theory which deals with normally hyperbolic manifolds (not necessarily attracting[15]) in systems of ODEs with fast-slow timescale separation; see for example Ref.[16] and the concise review in Ref.[15]. Although we focus here on the case of mass-action based chemical kinetics, we wish to remark that the GSP theory and the numerical tools mentioned below are rather general and can be applied to the dimensional reduction of various kinds of dynamical systems for which the velocity field $\mathbf{F}(\mathbf{x})$ is even non-polynomial. Briefly, let ϵ be a small dimensionless parameter which quantifies the timescale separation (increasing as $\epsilon \to 0$). It is supposed that ϵ "naturally" emerges from a rescaling of the ODEs. By denoting with $\mathcal{M}_0$ the central manifold corresponding to infinite timescale separation, Fenichel's theorems assert that there exists a family of manifolds $\mathcal{M}_\epsilon$ for the given $\epsilon \neq 0$, all exponentially close to each other as $\epsilon \to 0$, and locally invariant under the dynamics (i.e., they are "persistent" in the sense of "self-preserved" by the dynamics). Note the non-uniqueness of the solution due to the possible multiplicity of $\mathcal{M}_\epsilon$. The crucial point is that while $\mathcal{M}_0$ can be obtained by solving algebraic equations, the hard task is to go beyond the mere statement of existence and construct in practice *a* manifold $\mathcal{M}_\epsilon$ to be taken (locally) as the SM. To our knowledge, the computational singular perturbation (CSP) method of Lam and Goussis[8,17] represents the most faithful numerical implementation of the GSP concepts and, in principle, is able to produce such a SM under

the sole assumption that a timescale separation between fast and slow processes does exist. The CSP tool works with a matrix format of the ODEs and, in practice, one only has to choose two initial sets of linearly independent vectors which likely span the "slow" and "fast" subspaces. By means of a two-step procedure, the route makes a refinement of the initial guess and subsequent iterations of the procedure yield improved approximations of the fast and slow subspaces.[17] The CSP approximation of the SM is then given by the points in the concentration space where the velocity field has null projection on the fast subspace generated after a chosen number of iterations. However, the implementation of the CSP tool may be not trivial: the procedure fails if the initial guess is incorrect, and the requirement of a criterion to stop the iterations introduces a degree of subjectivity. Among other popular methods for the SM construction, still based on the assumption of timescale separation but less close to the GSP concepts and built more on empiric grounds, we mention the basic quasi-stationary-state and quasi-equilibrium approximations[18], the construction of intrinsic low dimensional manifolds (ILDMs)[19] and of attracting low dimensional manifolds (ALDMs)[11], and the category of "trajectory methods"[11–13,20]. Other approaches rely on different assumptions where the timescale separation is not explicitly considered. In particular, we mention the iterative evolution of functional maps[7,18], the method of "heteroclinic connections"[21], and several optimization approaches based on concepts borrowed from nonequilibrium thermodynamics[22–25]. None of these strategies provide the SM in the sense of Fenichel's theory, but only approximations whose accuracy has to be evaluated case by case.

In this work we shall present a new strategy to produce candidate points to the SM proximity, along with its implementation in the first release of the C++ software DRIMAK (Dimensional Reduction of Isothermal Mass-Action Kinetics) developed by us.[26] The approach exploits and combines the outcomes of our recent theoretical investigations concerning the achievement of canonical (*i.e.*, universal) mathematical formats of the evolution law for mass-action based kinetics, and of their application to the localization of the SMs[14,27,28]. As demonstrated in Ref.[14], a proper change of dynamic variables leads to a universal system of ODEs in an extended space of $N \times M$ mutually constrained variables. The study of the mathematical properties of such a new format allowed us to formulate a purely geometrical

and objective definition of SM[14]. Unfortunately, the algorithmic implementation of such a definition poses a series of problems which can be hard to tackle. In Ref.[28] we have shown how a second universal format of the ODEs, that we have termed "hyperspherical representation" of the reactive system, allows one to devise an approximate but computationally efficient route to individuate points expected to be *close* to the SM. In that framework, it was discovered that the "grooves" in the multidimensional landscapes of a peculiar pair of functions (see the $Z(\mathbf{x})$ and $Z_1(\mathbf{x})$ in the following) allows one to detect the slowness of the system's progression and the persistence of such slowness. Recognizing that these traits are typical of the SM neighborhood, it was indicated that suitably designed minimization routes, followed by a screening of the produced solutions, may be used to localize points on the SM proximity. This is the idea developed in the present work.

As will be shown, the strength of the present methodology lies in its intrinsic low computational cost, in spite of the fact that the search for candidate points can even be made inside very large hyper-rectangular regions of the concentration space and without any knowledge *a priori* about dimensionality and location of the SM, nor of the equilibrium manifold. We must stress the important aspect that the strategy proposed here is not intended to replace other techniques developed for localizing the SM; rather it can be better seen as a tool to produce good starting points for a subsequent refinement procedure and/or to restrict the domain for the SM construction by means of other techniques.

The remainder of the article is organized as follows. In the next section we summarize the theoretical background to introduce the key-functions $Z_n(\mathbf{x})$ with $n \geq 0$, which are adopted as guiding potentials; then we describe the multi-step minimization route which exploits such potentials in order to localize candidate points. In the 'Algorithmic implementation' section we illustrate the implementation of the ideas in the software DRIMAK, along with the characterization of the crucial computational steps in terms of scaling of the execution time as the number of species and reactions increases. Some technicalities are provided in the Supporting Information and in the documentation which accompanies the software. In the 'Examples' section we provide examples for two relevant cases: 1) a benchmark model of hydrogen combustion involving 6 species and 12 elementary reactions[29] also studied in Refs.[24,25] and by us in Ref.[14], and 2) a more complex mechanism of hydrogen combustion

involving 8 species and 42 elementary reactions[30]. The 'Conclusions' section provides a summary of the work presented herein, as well as perspectives for improvements of the strategy.

Theoretical background

Slow Manifolds from canonical formats of the ODEs

We shall focus on a general reactive process occurring in an isothermal and well-stirred medium with a fixed volume. The application of the mass-action law to express the rate of the elementary processes yields the j-th component of the velocity field, expressed as:

$$F_j(\mathbf{x}) = \sum_{m=1}^{M} \left(\nu_{P_j}^{(m)} - \nu_{R_j}^{(m)} \right) r_m(\mathbf{x}) \quad , \quad r_m(\mathbf{x}) = k_m \prod_i^N x_i^{\nu_{R_i}^{(m)}} \tag{2}$$

where k_m is the kinetic constant of the m-th elementary reaction with rate $r_m(\mathbf{x})$, and $\nu_{R_j}^{(m)}$ and $\nu_{P_j}^{(m)}$ are the stoichiometric coefficients of species j as reactant and product respectively (the coefficients are null if the species does not appear in the elementary reaction).

The system of ODEs $\dot{x}_j = F_j(\mathbf{x})$ specifies the evolution of $\mathbf{x}(t)$ from an initial condition $\mathbf{x}(0)$. Accordingly, any function of the actual system's state, say $f(\mathbf{x})$, evolves under the dynamics according to $f(t) \equiv f(\mathbf{x}(t))$. In what follows, time derivatives of suitable point-dependent functions will play an important role in our dimensional reduction approach. Let us introduce the notation used throughout the paper. We shall denote with $f^{(n)}(\mathbf{x})$ the point-dependent function such that

$$\frac{d^n f(\mathbf{x}(t))}{dt^n} \equiv f^{(n)}(\mathbf{x}(t)) \tag{3}$$

Explicitly, the function $f^{(n)}(\mathbf{x})$ represents the n-th time derivative of the property f, due to the dynamics, for the system in the state $\mathbf{x}$. Mathematically, $f^{(n)}(\mathbf{x}) = (\mathbf{F}(\mathbf{x}) \cdot \partial/\partial\mathbf{x})^n f(\mathbf{x})$ where the exponent n means that the operator $\mathbf{F}(\mathbf{x}) \cdot \partial/\partial\mathbf{x}$ is applied n times.[b]

[b]To see this, for the sake of notation let us introduce the operator $\hat{\mathcal{O}}(\mathbf{x}) = \mathbf{F}(\mathbf{x}) \cdot \partial/\partial\mathbf{x} = \sum_{i=1}^{N} F_i(\mathbf{x}) \partial/\partial x_i$. The first-order time derivative of $f(t) \equiv f(\mathbf{x}(t))$ is $df(t)/dt \equiv f^{(1)}(\mathbf{x}(t)) = \sum_{i=1}^{N} [F_i(\mathbf{x}) \partial f(\mathbf{x})/\partial x_i]|_{\mathbf{x}=\mathbf{x}(t)} = [\hat{\mathcal{O}}(\mathbf{x}) f(\mathbf{x})]\Big|_{\mathbf{x}=\mathbf{x}(t)}$ where it has been used $dx_i/dt = F_i(\mathbf{x})$. Note that $f^{(1)}(\mathbf{x})$, that is the first time derivative

Let us consider the following $(N \times M)^2$ quantities whose physical dimension is inverse-of-time:

$$V_{jm,j'm'}(\mathbf{x}) = M_{jm,j'm'} h_{j'm'}(\mathbf{x}) \tag{4}$$

where

$$h_{jm}(\mathbf{x}) = x_j^{-1} r_m(\mathbf{x}) \tag{5}$$

and $\mathbf{M}$ is the connectivity matrix with dimensionless elements

$$M_{jm,j'm'} = \left(\nu_{P_{j'}}^{(m')} - \nu_{R_{j'}}^{(m')}\right)\left(\delta_{j,j'} - \nu_{R_{j'}}^{(m)}\right) \tag{6}$$

where δ denotes the Kronecker Delta function. Some algebraic steps[27] show that the terms $V_{jm,j'm'}(t) \equiv V_{jm,j'm'}(\mathbf{x}(t))$ form a closed set of new dynamic variables whose evolution along a system's trajectory is governed by the following system of ODEs:

$$\dot{V}_{jm,j'm'} = -V_{jm,j'm'} \sum_{j'',m''} V_{j'm',j''m''} \tag{7}$$

The quadratic form of Eq. (7) is universal; that is, it is parameter-free and it underlies any kinetic scheme regardless of the number of species and elementary reactions.[c] All system-dependent features (*i.e.*, number of species and elementary reactions, stoichiometry, values of the kinetic constants) are borne on the dimension of such a set of new dynamical variables and on their mutual interrelations.

It was found that the key quantities in the localization of the SM are the following point-dependent "rates":

$$z_{jm}(\mathbf{x}) = \sum_{j'm'} V_{jm,j'm'}(\mathbf{x}) \tag{8}$$

---

under the flow, is the so-called Lie derivative. The second-order derivative is then $d^2 f(t)/dt^2 \equiv f^{(2)}(\mathbf{x}(t)) = df^{(1)}(\mathbf{x}(t))/dt = [\hat{\mathcal{O}}(\mathbf{x}) f^{(1)}(\mathbf{x})]\Big|_{\mathbf{x}=\mathbf{x}(t)} = [\hat{\mathcal{O}}(\mathbf{x})(\hat{\mathcal{O}}(\mathbf{x}) f(\mathbf{x}))]\Big|_{\mathbf{x}=\mathbf{x}(t)} \equiv [\hat{\mathcal{O}}(\mathbf{x})^2 f(\mathbf{x})]\Big|_{\mathbf{x}=\mathbf{x}(t)}$. By iterating, $d^n f(t)/dt^n \equiv f^{(n)}(\mathbf{x}(t)) = [\hat{\mathcal{O}}(\mathbf{x})^n f(\mathbf{x})]\Big|_{\mathbf{x}=\mathbf{x}(t)}$.

[c] Although it was derived by us in Ref.[27], the kind of transformation from $\mathbf{x}$ to the set of $h_{jm}(\mathbf{x})$ in Eq. 5 was already known for decades and was even re-discovered independently by several authors with minor variations. For example, it should be mentioned that is was applied by Brenig and Goriely in the context of general transformations amongst equivalence classes of representation for continuous-time systems[31], by Fairén and Hernández-Bermejo[32,33] and by Gouzé[34].

As shown in the Supporting Information of Ref.[27], these $N \times M$ rates are mutually linked by a number of linear interrelations so that only N of them are independent (the same number of interrelations, but of non-linear type, links the h_{jm} functions defined in Eq. 5). What emerged from the combined formal-heuristic inspection illustrated in Ref.[14], is that the SM can be defined by operating with the point-dependent time derivatives of n-th order, $z_{jm}^{(n)}(\mathbf{x})$, as outlined below.

For the sake of brevity, let us introduce the cumulative index Q to label the species-step pair from now on:

$$Q = (j, m) \quad , \quad Q = 1, 2, \cdots, Q_s \quad , \quad Q_s = N \times M \tag{9}$$

In Ref.[14] we formulated the conjecture that a trajectory $\mathbf{x}(t)$ enters an "Attractiveness Region" (AR) of the concentration space, within which the high-order time-derivatives $z_Q^{(n)}(\mathbf{x}(t))$ tend to become multiples of one another and monotonically decay to zero towards the equilibrium. The SM is then defined as the hyper-surface formed by the points $\mathbf{x}_{\mathrm{SM}}$ within the AR such that $z_Q^{(n)}(\mathbf{x}_{\mathrm{SM}}) = 0$ for *all* Q as $n \to \infty$. On the EM, one has the stronger and exact condition $z_Q^{(n \geq 1)}(\mathbf{x}_{\mathrm{EM}}) = 0$. This provides a geometric *definition* of SM as a global object in the concentration space. The implementation of this definition allowed us to detect SMs in a series of simple case models[14]. However, the practical application to produce points $\mathbf{x}_{\mathrm{SM}}$ poses two kinds of problem: 1) there is actually no way to know in advance the dimensionality and the boundaries of the AR within which the search has to be performed; 2) this definition of SM requires the computation of derivatives $z_Q^{(n)}(\mathbf{x})$ of very high order. While the quadratic structure of the ODEs in Eq. 7 offers the possibility to easily compute high-order derivatives via recursive formulae (see the Appendix A), the problem of circumscribing the AR still remains the crucial one.

It should be noted that several model reduction criteria based on time derivatives have been proposed in the past. However, those methods employ the ($\mathbf{x}$-dependent) time derivatives of the concentration vector $\mathbf{x}$, while here we deal with derivatives of the rate functions $z_Q(\mathbf{x})$; the connection between the two sets of derivatives is not trivial. In fact, by combining Eqs. 4 - 8 it can be verified that $z_{jm}(\mathbf{x}) = \sum_{j'} w_{jm,j'}(\mathbf{x}) x_{j'}^{(1)}(\mathbf{x})$ with the point-dependent factors $w_{jm,j'}(\mathbf{x}) = (\delta_{j,j'} - \nu_{R_{j'}}^{(m)}) x_{j'}^{-1}$. By taking successive time derivatives of both members,

it can be seen the n-th time derivative of a rate $z_Q(\mathbf{x})$ is related, in a quite intricate way, to the components of $\mathbf{x}$, $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$, ..., $\mathbf{x}^{(n+1)}$. In Appendix C we give only a brief and qualitative outline of the main approaches aimed at localizing the SM by employing time derivatives of the state vector $\mathbf{x}$. Formal connections between our approach and these other strategies are still to be established on formal grounds.

Proximity to the Slow Manifold

In Ref.[28] we have made some progress in localizing points which *likely* fall in the neighborhood of the SM, rather than search for the true $\mathbf{x}_{\mathrm{SM}}$ points according to our definition of SM given in Ref.[14]. The initial step was to turn to a new representation of the state of the reactive system in another $(N \times M)^2$ abstract space. We have termed such a representation as "hyperspherical", since the actual state is specified by a positive-valued "radial" coordinate with physical dimension of inverse-of-time, and by a set of dimensionless "angular" coordinates.

The analysis of the dynamics for these new state variables (which are clearly mutually interrelated) led us to individuate tentative mathematical formulations to express the conditions of "slowness" and "persistence of the slowness" when a trajectory is close to the SM. Namely, argumentation in Ref.[28] led us to indicate that the following scalar functions might serve as "guiding potentials" to drive the search for candidate points in the proximity of the SM:

$$Z_n(\mathbf{x}) = \sqrt{Q_s^{-1} \sum_Q z_Q^{(n)}(\mathbf{x})^2} \tag{10}$$

The division by Q_s, which is immaterial in practice, is introduced only to interpret the $Z_n(\mathbf{x})$ functions as the root-mean-square averages of the $z_Q^{(n)}(\mathbf{x})$ derivatives. If a number N^{irr} of species are irreversibly produced (*i.e.*, they do not appear as reactants in any of the elementary steps), then the SM hypersurface is orthogonal to the concentration subspace of the reactant species. In this situation, it is convenient to exploit such a dimensional reduction *a priori* and operate with the "reduced" guiding potentials $Z_n(\mathbf{x})$ computed by restricting the summation in Eq. 10 to the subset of $(N - N^{\mathrm{irr}}) \times M$ values $Q = (j, m)$ with j referring to reactant species. Clearly, the z_{jm} components involved are functions only of the concentrations of these species.

In particular, the lowest-order functions, *i.e.*, $Z(\mathbf{x}) \equiv Z_0(\mathbf{x})$ and $Z_1(\mathbf{x})$, prove to be sufficient to localize the proximity of the SM. As we have shown in Ref.[28] for a model case (the Lindemann-Hinshelwood scheme[1]), the landscapes of these functions display characteristic "grooves" within which the condition of slowness (grooves of $Z(\mathbf{x})$) and of its persistence (grooves of $Z_1(\mathbf{x})$) are expected to be met. A two-step minimization route along chosen paths (see below) was proposed to detect points for which both conditions are likely fulfilled. Starting from some randomly drawn point $\mathbf{x}_0$, a first minimization of $Z(\mathbf{x})$ leads to a point $\mathbf{x}_1$ into the "slowness region", while a subsequent minimization of $Z_1(\mathbf{x})$ starting from $\mathbf{x}_1$ leads to a point $\mathbf{x}_2$ (supposed to be close to $\mathbf{x}_1$) eventually taken as a candidate point to the SM proximity. The procedure can be then continued to higher orders of derivatives, that is, by considering the functions $Z_n(\mathbf{x})$ and performing an $(n+1)$-step minimization. Continuation to higher derivatives, however, was found to yield (at least in a series of preliminary tests) little improvement at the price of increasing computational time.[d]

To perform the multi-step minimization, we opt for paths in which the concentration of a species is fixed and the minimization of the functions is performed with respect to the other components of the set $\mathbf{x}$. The motivation of such a choice relies on the fact that, without any constraint, the minimization process would probably produce only points $\mathbf{x}_{\mathrm{EM}}$ on the EM, since $Z_n(\mathbf{x}_{\mathrm{EM}}) = 0$ for any order n. However, if the dimension of the EM is smaller than $N-2$, then the $(N-1)$-dimensional hyperplanes (*i.e.*, the search sections at fixed concentration of one of the species) have a very low chance to intersect the EM, even if a portion of it falls within the domain of inspection. This is the situation which is likely encountered in the cases of interest where the SM, and hence also the EM, have a dimension much lower than N. If the "active space" is reduced to a number $\tilde{N} < N$ of concentrations

[d]Interestingly, there seems to be some connection (albeit qualitative at this stage) between our two-step minimization route and the SM construction via the variational trajectory-based method with objective function $\Phi(\mathbf{x}) = ||\mathbf{x}^{(2)}||^2$ (see the Appendix C for notation and details). As indicated by Lebiedz and coworkers in Ref.[20], the choice of such basic objective function in the early implementations of the strategy was motivated by the fact that low values of $\Phi(\mathbf{x})$ likely catch, as a whole, the slowness of the dynamics on the SM and the attractiveness of the SM. Notably, both approaches are based on constrained minimization routes, work with time derivatives of the velocity field at most of second order, and employ objective functions which are supposed to catch the same features of the evolution on the SM.

of independent species (because of the enforcement of linear constraints and/or neglection *a priori* of the species only produced, see the next section), the considerations made above still hold regarding the search in the $\tilde{N}$-dimensional subspace. Finally, once several minimizations for different (fixed) values of the species concentrations have been performed, the solutions are then merged.

In the Supporting Information of Ref.[28] we have shown that an early implementation of the basic two-step strategy (*i.e.*, the use of only Z and Z_1) is effective in localizing the SM neighborhood for two model cases, namely the Lindemann-Hinshelwood scheme and a highly non-linear scheme with elementary steps up to the fourth order. However, a number of issues made clear that several improvements were required: 1) to assure the quick localization of the candidate points within a given multidimensional box in the concentration space, possibly under enforcement of linear constraints among the concentrations, 2) to remove "spurious solutions",[e] and 3) to establish a ranking for the likelihood that, according to the chosen approach, the remaining points are believed to be close to the SM. The constraints mentioned above may be the intrinsic stoichiometric ones (*i.e.*, those related to mass-conservation along the trajectories) or even arbitrary constraints which fix linear combinations of the species concentrations to given values (see the 'Examples' section). These constraints allow one to focus on sections of the full concentration space in order to simplify the visualization and the presentation of the outcomes. The technical solutions that we propose to face the issue 1) are presented in the 'Algorithmic implementation' section, along with the description of how they are implemented in the software DRIMAK. Concerning the *a posteriori* check on the candidate points (issues 2) and 3) above) we opt to employ a screening based on the ILDM approach mentioned in the Introduction[19] and implemented as described in Appendix B. Such an analysis is performed by means of an independent program which reads the output from DRIMAK and yields the filtered results. A DRIMAK user may choose to employ a different motivated strategy to assess, case by case, the quality of the raw outcome and make a sensible selection of the points produced.

[e] As shown in Ref.[28], the strategy leads also to the localization of points far from the perceived SM. This trait seems to be almost unavoidable depending on the features of the specific kinetic scheme.

Algorithmic implementation

Computational strategy as employed in DRIMAK C++ code

The central idea depicted in the 'Proximity to the Slow Manifold' subsection is implemented in the C++ software DRIMAK, the pseudo-code of which is given in the box 'Algorithm 1'. The algorithm employs the arrays of species concentrations specified hereafter. First, let $\mathbf{x}$ be the array made of the complete set of concentrations of the N species. The user is allowed to specify a number $N^{\mathrm{con}} \geq 0$ of linear constraints among the species concentrations. In this case, DRIMAK also requires the specification of an equal number of "dependent" species (this automatically fixes the number $N^{\mathrm{ind}} = N - N^{\mathrm{con}}$ of "independent" species). The concentration array $\mathbf{x}$ is then split into the two subsets $\mathbf{x}^{\mathrm{dep}}$ and $\mathbf{x}^{\mathrm{ind}}$ corresponding to the dependent and independent species respectively. If $N^{\mathrm{con}} > 0$, the full set $\mathbf{x}$ is retrieved from the independent concentrations $\mathbf{x}^{\mathrm{ind}}$ by employing the procedure described in the Supporting Information. Finally, it might be the case that, among the N^{ind} species, a fraction N^{irr} of them does not enter as reactants in any elementary step.[f] The concentration array $\tilde{\mathbf{x}} \subseteq \mathbf{x}^{\mathrm{ind}}$, made of $\tilde{N} = N^{\mathrm{ind}} - N^{\mathrm{irr}}$ elements and obtained by removing the N^{irr} species concentrations from $\mathbf{x}^{\mathrm{ind}}$, constitutes the active space of the minimization procedure.

The user is also asked to input the borders ($\mathbf{x}^{\mathrm{ind}}_{\mathrm{min}}$ and $\mathbf{x}^{\mathrm{ind}}_{\mathrm{max}}$) of the N^{ind}-dimensional region to be inspected for the SM search. The N-dimensional region I indicated in 'Algorithm 1' is then defined by $\mathbf{x}^{\mathrm{ind}}_{\mathrm{min}} < \mathbf{x}^{\mathrm{ind}} < \mathbf{x}^{\mathrm{ind}}_{\mathrm{max}}$ for the independent species, along with $\mathbf{x}^{\mathrm{dep}} > \mathbf{0}$ for the dependent ones.

The total number of requested points is equally distributed among the $\tilde{N}$ species whose concentrations span the active space of the search. For each one of these species, one at a time, the concentration is kept fixed while doing the multi-minimization of the functions $Z_n(\mathbf{x}(\mathbf{x}^{\mathrm{ind}}))$ with respect to the concentrations of the remaining $\tilde{N} - 1$ species inside the

[f] Although not explicitly reported in 'Algorithm 1', at the beginning of the algorithm, a check is made to ascertain whether some species are irreversibly produced. As mentioned in the section 'Proximity to the Slow Manifold', in this case the computation of the $Z_{n \leq n_{\mathrm{max}}}(\mathbf{x})$ functions is made only with the "reduced set" of $(N - N^{\mathrm{irr}}) \times M$ components z_{jm} where the label jm refers to the pair made of reactant species and elementary step. In addition, these z_{jm} components are functions only of the concentrations of the reactant species.

Algorithm 1 DRIMAK pseudo-code

Require: From input file: reaction mechanism (number N and list of species, number M and stoichiometry of the elementary steps, values of the kinetic constants); possible N^{dep} linear constraints and their specification; list of the dependent species; boundaries of the inspected N^{ind}-dimensional subregion of I. Prompt input: maximum number TOT_POINTS of points to be produced; initial seed for random number generation; maximum order $n_{\max} \geq 1$ for the $Z_n(\mathbf{x})$ functions; initial trust region radius (ρ_{beg}) and final trust region radius (ρ_{end}) for the minimization procedures.

Ensure:

for $k = 1$ to $\tilde{N}$ **do**
  for pts $= 1$ to TOT_POINTS$/\tilde{N}$ **do**
    Draw a point $\mathbf{x}_0^{\mathrm{ind}}$ at random in the N^{ind}-dimensional subdomain of I
    for $n = 0$ to $n_{\max}$ **do**
      Find $\tilde{\mathbf{x}}_{\min} = \arg\min_{\tilde{\mathbf{x}} \subseteq \mathbf{x}^{\mathrm{ind}}} \{Z_n(\mathbf{x}(\mathbf{x}^{\mathrm{ind}}))\}$ starting from the initial point $\mathbf{x}_n^{\mathrm{ind}}$ and under the constraint that the k-th component of $\tilde{\mathbf{x}}$ remains fixed
      Fill $\mathbf{x}_{n+1}^{\mathrm{ind}}$ with the $\tilde{\mathbf{x}}_{\min}$ values (the remaining N^{irr} entries are taken from $\mathbf{x}_n^{\mathrm{ind}}$)
      Retrieve the full point $\mathbf{x} = \mathbf{x}(\mathbf{x}_{n+1}^{\mathrm{ind}})$
      If $(\mathbf{x} \notin I)$ **goto** 3
    end for
    Store the candidate point $\mathbf{x}$
  end for
end for
return Produced points

user-defined hyper-rectangle embedded in the region I. In the current implementation, the initial point $\mathbf{x}_0^{\mathrm{ind}}$ is drawn at random from the uniform distribution on the logarithm of the concentrations; such a selection is made by employing the standard C++ function *rand()* to generate random numbers (one per coordinate) from the uniform distribution between 0 and 1, and then performing a rescaling according to the dimensions of the hyper-rectangle.

The minimization of the functions $Z_n(\mathbf{x})$ is performed by means of a FORTRAN77 routine written by Michael J. D. Powell[35]. Such a routine, called LINCOA ("LINearly Constrained Optimization Algorithm") belongs to the category of the so-called trust region methods[36,37]. It allows one to efficiently find a *local* minimum of a function without explicit computation of its derivatives. The routine requires the initial and the final values of the trust region radius, ρ_{beg} and $\rho_{\mathrm{end}} \leq \rho_{\mathrm{beg}}$ respectively (from the name of the parameters in the LINCOA code). The search for a minimum terminates when the trust region radius, which can not increase during the iterations, reaches the lower bound ρ_{end}. While ρ_{beg} should be chosen to be of the order of one tenth of the greatest expected change of variables at the beginning, a trial value of ρ_{end} should be the required accuracy for the localization of the minimum point in the concentration space. However, there is no direct connection between ρ_{end} and the actual accuracy of the produced point of minimum. Remarkably, LINCOA also allows one to enforce a number of linear constraints among the independent variables. We exploited such a feature in order to confine the minimization outcomes within the user-specified domain (for more details see the software documentation). After each call to LINCOA, a check is made to ensure that the concentrations of the dependent species are non-negative. The usage of LINCOA within DRIMAK requires that the dimensionality of the active space is greater than two, *i.e.*, the reaction mechanism needs to have at least three independent species that enter some elementary step as reactants.

Finally, given the need to work with concentrations that span several orders of magnitude, we decided to perform the minimization by using the base-ten logarithm of the concentrations (in place of their actual values) as independent variables. Preliminary calculations revealed that such a choice does not significantly affect the overall computation time, while it seems to improve the accuracy of the results for the example schemes studied.

The likelihood of the produced points being close to the SM may eventually be evaluated

by resorting to the ILDM strategy as described in Appendix B. This allows one to rank the points and, possibly, to exclude the highly "unreliable" ones.

The execution of DRIMAK requires a user-provided input file; for a detailed description of such a file and some examples see the software documentation. In brief, the input file contains the chemical mechanism to be inspected (encoded in a specific format), the numerical values of the kinetic constants and, possibly, a number of linear constraints to be applied to the concentrations of the species in order to explore sections of the full concentration space; it suffices that the concentrations of at least three species (not irreversibly formed) remain unconstrained.

DRIMAK is an embarrassingly parallel code which implements the MPI paradigm. If the number of processes chosen by the user is greater than one, then the number of points to be found is equally distributed among the fixed processes. It is worth to pointing out that the multi-step minimization route may repeatedly fail in localizing points. This may happen when the specific section of the concentration space does not intersect the SM inside the selected domain I, or even if no portion of the SM falls in such a domain. In these situations, giving priority to end the computation after a maximum number of iterations, the total number of points produced could be lower than the requested number. In the worst case, in which no points are detected, DRIMAK throws an instance claiming there are no candidate points to the SM proximity and stops its execution.

Performance scaling versus N and M in the computation of $Z(\mathbf{x})$ and $Z_1(\mathbf{x})$

Much of the computational time is spent on evaluating the functions $Z_n(\mathbf{x})$ during the multi-step minimizations.[g] For the basic case $n_{\max} = 1$, we have faced the problem of

[g]We should stress that the exploitation of the sparsity of the connectivity matrix $\mathbf{M}$ is crucial to the reduction of the computational times of the functions $z_Q^{(n)}$ required in calculations of Z_n (see Appendix A). We have also tested the effectiveness of GPU (Graphic Processor Units) programming to speed up the matrix-vector operations. Preliminary checks have shown that a negligible gain is obtained; however, it will be worthwhile to continue the inspection of GPU programming, especially to develop kernels for the evaluation of the h_Q functions which requires the computation of powers of the species concentrations.

establishing how the times required to compute $Z(\mathbf{x})$ and $Z_1(\mathbf{x})$, for a tested point $\mathbf{x}$, scale with the dimension of the system under inspection, that is with N (number of species) and M (number of elementary steps), but regardless of the peculiarity of the kinetic scheme.

For this purpose, we opted to generate randomly an ensemble of kinetic schemes, with N ranging from 2 to 50 and M from N to $3N$. Each scheme is created by drawing at random, for each elementary step m, its molecularity $\mathcal{M}_m = 1, 2, 3$. For each step, reactant species and related stoichiometric coefficients are also generated randomly according to $\sum_j \nu_{R_j}^{(m)} = \mathcal{M}_m$. Then, the product species and the related coefficients are also drawn at random under the constraint $\sum_j \nu_{P_j}^{(m)} = \mathcal{M}_m$. The last constraint is imposed in order to preserve mass-conservation globally, that is, to confer some realism to the randomly generated scheme. After generation of the elementary reactions, a check is made to exclude possible "identities" and replicated reactions (in this case, new reactions are generated and the check is repeated). In addition, a final check is made to assure that all the generated schemes are distinct, that is, made by steps which are not mere permutations. For each scheme, the values of the kinetic constants and the species concentrations were generated at random in the intervals from 10^{-4} to 10^4 and from 10^{-6} to 1 respectively (units of measure are immaterial in this context). Finally, for each pair (N, M), 50 different kinetic schemes have been created. The computational times needed for calculating Z and Z_1 were stored, along with their averages made upon the 50 schemes. The code was compiled with no optimization flags (the "-O0" flag was used under Linux environment) in order to have an optimization-independent output. These tests (as well as the other calculations to produce the results presented in this paper) were performed on a workstation whose characteristics are specified in the footnotei.

First of all, the spread of computational times over the ensemble of 50 schemes per each (N, M) pair, was found to reach at most 30% of the average time; thus, being interested only in the scaling of the order of magnitude of the computational time, we shall focus on the average values. The results are presented in Figure 1. The average times for computing Z and Z_1 are shown with blue marks. It turned out that the following function

$$\tau_{\boldsymbol{\alpha}}(N, M) = \alpha_1 + \alpha_2 N + \alpha_3 M + \alpha_4 N^2 + \alpha_5 M^2 + \alpha_6 NM \tag{11}$$

can fit adequately the average computational times of Z and Z_1. In both cases, the array $\boldsymbol{\alpha}$

was obtained by minimizing the objective function

$$\Psi(\boldsymbol{\alpha}) = \sqrt{\sum_{N,M} \left(\frac{\tau_{\boldsymbol{\alpha}}(N,M) - \overline{\tau}(N,M)}{\overline{\tau}(N,M)} \right)^2} \tag{12}$$

where $\overline{\tau}(N,M)$ is the average time actually required for the pair (N,M). The interpolating surfaces are shown in light-grey and the best sets of parameters are given in the figure caption. In the figure we also report the average computational times for the models of hydrogen combustion, as illustrated in the next section. These values are in good agreement with Eq. 11. This is particularly significant for the extended hydrogen combustion model (Scheme B in the following) which falls outside the explored range species/steps used to derive the parametric expression in Eq. 11. This means that such an equation may be used to make predictions about the computational time needed for the calculation of Z and Z_1 on different schemes.

Furthermore, by repeating the same tests on different computers operating with different processors (but of the same typology of the reference one indicated in footnote i) and clock frequencies, we noted that the offset α_1 depends on the specific machine, while the coefficients from α_2 to α_6 roughly scale with the inverse of the clock frequency. Thus, by taking into account the fact that the clock frequency here was 1.80 GHz, from Eq. 11 one could estimate the computational time of $Z(\mathbf{x})$ and $Z_1(\mathbf{x})$ as $\tau(N,M) \simeq \tau_0 + [\tau_{\boldsymbol{\alpha}\times 1.80/f}(N,M) - \tau_{\boldsymbol{\alpha}\times 1.80/f}(N_0,M_0))]$ where τ_0 stands for the computational time required for a single low-dimension test mechanism with N_0 species and M_0 elementary steps, and f is the clock frequency in GHz of the specific computer (the machine-dependent offset cancels).

Examples

In this section we present the results of the application of DRIMAK on two kinetic models of hydrogen combustion. The assumptions of a well-stirred medium and isothermal conditions are clearly unrealistic. However, our purpose is only to test the effectiveness of DRIMAK regardless of the realism of the specific example. The first model, Scheme A in the following, is a basic scheme with 6 species and 12 elementary reactions[29]; such a scheme is often taken

as a benchmark in studies regarding the simplification of chemical kinetics. The second model, Scheme B, is a much more elaborate mechanism[30] which features 8 species and 21 reversible elementary steps, two pairs of which are actually the same reactions with different rate constants.

As detailed below, some constraints are applied to confine the reacting systems (both the trajectories and the candidate points produced by DRIMAK) over sections of the full 6-dimensional or 8-dimensional concentration spaces.

Basic scheme of hydrogen combustion

The basic scheme of hydrogen combustion is reported below:

$$\begin{aligned}
\mathrm{H_2} &\rightleftarrows 2\mathrm{H} && k_1 = 2,\ k_{-1} = 216\\
\mathrm{O_2} &\rightleftarrows 2\mathrm{O} && k_2 = 1,\ k_{-2} = 337.5\\
\mathrm{H_2O} &\rightleftarrows \mathrm{H} + \mathrm{OH} && k_3 = 1,\ k_{-3} = 1400\\
\mathrm{H_2} + \mathrm{O} &\rightleftarrows \mathrm{H} + \mathrm{OH} && k_4 = 1000,\ k_{-4} = 10800\\
\mathrm{O_2} + \mathrm{H} &\rightleftarrows \mathrm{O} + \mathrm{OH} && k_5 = 1000,\ k_{-5} = 33750\\
\mathrm{H_2} + \mathrm{O} &\rightleftarrows \mathrm{H_2O} && k_6 = 100,\ k_{-6} = 0.7714
\end{aligned}$$

(Scheme A)

It is implicit that the time variable and the volumetric concentrations are expressed in some units of measure, here immaterial, which should be fixed by comparing the progression rate of such a fictional reactive system with experimental observations (see for example Ref.[29]).

Two linear constraints are applied, namely

$$\begin{aligned}
2[\mathrm{H_2}] + 2[\mathrm{H_2O}] + [\mathrm{H}] + [\mathrm{OH}] &= 2\\
2[\mathrm{O_2}] + [\mathrm{H_2O}] + [\mathrm{O}] + [\mathrm{OH}] &= 1
\end{aligned} \tag{13}$$

By imposing these constraints one fixes the total concentrations of hydrogen atoms *and* of oxygen atoms which, in addition, will remain in a stoichiometric ratio of 2:1. Correspondingly, the number of independent species concentrations reduces to four. As dependent species we chose $\mathrm{H_2O}$ and $\mathrm{O_2}$. In such a 4-dimensional section of the full space, the SM

appears to be 1-dimensional, while the EM reduces to a point at concentrations $[H_2]_{eq} = 0.7$, $[H]_{eq} = 0.05$, $[O_2]_{eq} = 0.135$, $[O]_{eq} = 0.02$, $[OH]_{eq} = 0.01$. Furthermore, from previous studies[14], it is also known that such a SM is embedded in a 2-dimensional surface which is approached by the trajectories before they reach the proximity of the SM itself. This surface can be "glimpsed" in Figure 2 by looking at the behaviour of the ensemble of trajectories.

The search for candidate points to the SM proximity is performed within the domain $5\cdot10^{-3} < [H_2] < 1$, $10^{-3} < [H] < 9\cdot10^{-2}$, $2.5\cdot10^{-3} < [O] < 9\cdot10^{-2}$, $5\cdot10^{-4} < [OH] < 9\cdot10^{-2}$, $[H_2O] > 0$, $[O_2] > 0$.[h] The results of the calculation are shown in Figures 2 and 3 where the produced points are displayed with blue dots. The production of 2000 candidate points by DRIMAK requested roughly 20 seconds on our workstation.[i]

The main outcome is that the proximity of the perceived 1-dimensional SM is successfully localized by the software, but a non-negligible amount of "spurious" solutions is also produced. It might be the case that such points belong to the 2-dimensional surface which embeds the SM. Indeed, trajectories which start from these points are found to remain within the thin region which encloses the majority of the spurious solutions. Further investigations are needed to shed light on such a phenomenology.

The employment of the ILDM-based strategy as described in Appendix B finally yields quite good results; the 1-dimensional SM is in fact caught efficiently while almost all the unlikely solutions are removed. It is worth stressing that this step also requires low computational cost; the "filtering" of the 2000 candidate points produced by DRIMAK required only a few seconds on our workstation. The points which passed the ILDM ranking-plus-screening, totalling 734 points, are shown with larger red marks. With reference to the parameters reported in Appendix B, the ranking of the solutions has been done by employing $\epsilon_{ILDM} = 0.5$, followed by deletion of points if $\eta < 10$. These are obviously subjective choices and the application of different parameters would modify the outcome. Nonetheless this example shows that, with some caution and insight case by case, it is possible to "filter" the results in a sensible way.

[h] The initial trust region radius was fixed to 10^{-1}, while ρ_{end} was set to 10^{-10}.

[i] Computations were performed on a workstation with an Intel(R) Xeon(R) CPU E5-2603 v2 @ 1.80 GHz and 32 GB of RAM.

Extended scheme of hydrogen combustion

The extended kinetic model of hydrogen combustion[30] is reported below. In this case, volumetric concentrations and time variables are expressed in units mol/L and sec., respectively. For the reference temperature, we chose 1000 K. The forward kinetic constants at this temperature were obtained using data from Ref.[30] while the backward constants derive from microscopic reversibility (see the Supporting Information for details and actual values of the kinetic constants).

1) $H + O_2 \rightleftarrows O + OH$

2) $O + H_2 \rightleftarrows H + OH$

3) $H_2 + OH \rightleftarrows H_2O + H$

4) $O + H_2O \rightleftarrows 2OH$

5) $H_2 + N_2 \rightleftarrows 2H + N_2$

6) $2O + N_2 \rightleftarrows O_2 + N_2$

7) $O + H + N_2 \rightleftarrows OH + N_2$

8) $H + OH + N_2 \rightleftarrows H_2O + N_2$

9) $H + O_2 + N_2 \rightleftarrows HO_2 + N_2$

10) $HO_2 + H \rightleftarrows H_2 + O_2$

11) $HO_2 + H \rightleftarrows 2OH$

12) $HO_2 + O \rightleftarrows OH + O_2$

13) $HO_2 + OH \rightleftarrows H_2O + O_2$

14) $2HO_2 \rightleftarrows H_2O_2 + O_2$ (duplicated)

15) $H_2O_2 + N_2 \rightleftarrows 2OH + N_2$

16) $H_2O_2 + H \rightleftarrows H_2O + OH$

17) $H_2O_2 + H \rightleftarrows H_2 + HO_2$

18) $H_2O_2 + O \rightleftarrows OH + HO_2$

19) $H_2O_2 + OH \rightleftarrows H_2O + HO_2$ (duplicated)

(Scheme B)

Two linear constraints are applied, as in Scheme A, to the total concentrations of hydrogen and oxygen atoms in the system:

$$\begin{aligned} [H] + [OH] + 2[H_2] + 2[H_2O] + [HO_2] + 2[H_2O_2] &= 0.09 \text{ mol/L} \\ [O] + [OH] + 2[O_2] + [H_2O] + 2[HO_2] + 2[H_2O_2] &= 0.045 \text{ mol/L} \end{aligned} \tag{14}$$

Accordingly, the concentrations of 6 species constitute the independent variables; as dependent variables, here we opt to take the concentrations of the species H_2O_2 and H_2O. The molar concentration of the buffer species N_2 was set to 0.2025 mol/L. Similarly to Scheme A, under the mass constraints, a 1-dimensional SM emerges and the EM reduces to a single point.

The search for candidate points to the SM proximity has been conducted within the

following domain: $10^{-9} < [H] < 10^{-2}$, $10^{-12} < [O] < 10^{-2}$, $10^{-9} < [OH] < 10^{-2}$, $5 \cdot 10^{-8} < [H_2] < 10^{-2}$, $5 \cdot 10^{-8} < [O_2] < 10^{-2}$, $10^{-13} < [HO_2] < 10^{-2}$, $[H_2O] > 0$, $[H_2O_2] > 0$ (all values are in mol/L).[j] We like to stress the remarkable extension of such a domain, whose shorter dimension spans almost six orders of magnitude, while the larger one spans eleven orders of magnitude.

The Figures 4 and 5 show one three-dimensional and three two-dimensional projections of the whole concentration space. Such projections refer to the three main species involved in the reaction, namely H_2, O_2 and H_2O. Because of the relatively high complexity of this scheme, we chose to present only the plots referring to these species. The total number of candidate points produced by the software is 2000, and it took roughly 35 seconds to complete the execution. The ILDM-based "filtering" strategy retains very few points, in fact only 9 of the 2000points produced, but it is worth noting that they appear to be among the ones closest to the perceived SM. The ILDM ranking/screening of the outcomes has been done with $\epsilon_{\mathrm{ILDM}} = 0.5$ (as for Scheme A), while η is just required to be greater than 1. The latter condition is milder than that applied to remove spurious solutions for Scheme A, but anyway consistent with the idea that the velocity vector should have the main projection on the lower set of eigenvectors (the "slow" subset) of the local kinetic matrix, as outlined in Appendix B. Once again, it must be stressed that such choices are (to some extent) subjective, but nonetheless necessary in order to remove spurious solutions. At any rate, even taking into account the "unfiltered" points, the results could be considered satisfactory for the three important species; indeed there is an evident accumulation of points just in the proximity of the perceived SM.

Conclusions

In this paper we have presented an algorithm developed by us for the production of candidate points to be in the proximity of the slow manifold in the species concentration space. The

[j] In this case the initial trust region was set to 10^{-1}, while ρ_{end} was set to 10^{-2}. Lower values assigned to these parameters are shown to lead, for this scheme, to numerical problems causing DRIMAK to stop its execution.

approach is based on the theoretical framework previously derived by us and presented in detail elsewhere[14,27,28]. We have implemented the method into the code DRIMAK[26] written in C++ with exploitation of the MPI paradigm.

We have tested the software on two model schemes for hydrogen combustion, obtaining 2000 candidate points and then "filtering" them by using a strategy based on the ILDM method[19]. For both schemes here presented, the software was able to produce candidate points for the SM proximity in a very effective way. By considering that the inspected regions span several order of magnitudes in the species concentrations (and, most importantly, that such a huge extension of the research domain in logarithmic scale does not affect significantly the performance for the studied models), these achievements seem to be even more valuable. This means that the software can be potentially applied to systems where the *a priori* knowledge on the existence and localization of slow manifolds is limited. Furthermore, the computational performance shown in this study (less than ten chemical species, few tens of elementary steps, tens of seconds on a standard computer to produce thousands of candidate points to the SM neighborhood) discloses a promising scenario for the application of DRIMAK to more complex mechanisms.

We like to stress again the importance of a "filtering" procedure in a post-production ranking and screening of the DRIMAK outcomes. A sound procedure not only permits one to neglect evidently spurious solutions, but it would also provide a measure (through the ranking of the points) of the proximity to the target slow manifold. The ILDMs-based criterion employed here proved to be effective (although also a large number of evident "good points" are removed) and to require a low computational cost, at least for the present examples. On the other hand, more effective routes for the post-production selection may be developed and a DRIMAK user even has the freedom to devise a personal strategy to tackle the problem.

Finally, we must underline the fact that our algorithm does not compete with other methods to construct the slow manifolds. Rather, our strategy is aimed at providing "likely good points" from which other methods (possibly of heavier computational cost) could start the localization of the SM. In this sense, ours and other methodologies are complementary and their synergy could be very useful especially for high-dimensional kinetic schemes.

Appendix A: Recursive formulae for the time derivatives $z_Q^{(n)}$

The time evolution of the terms $h_Q(t) = h_Q(\mathbf{x}(t))$ defined in Eq. 5 is specified by[27]

$$\dot{h}_Q = -h_Q \sum_{Q'} M_{Q,Q'} h_{Q'} \tag{A1}$$

This is indeed the basic equation which yields Eq. 7 once $V_{Q,Q'}(t) = M_{Q,Q'} h_{Q'}(t)$ is considered. By deriving n times both members (using the rule of multiple derivative of a product of functions), the following recursive relation is obtained:

$$h_Q^{(n+1)}(\mathbf{x}) = -\sum_{Q'} M_{Q,Q'} \sum_{m=0}^{n} \binom{n}{m} h_Q^{(m)}(\mathbf{x})\, h_{Q'}^{(n-m)}(\mathbf{x})\,, \qquad \binom{n}{m} = \frac{n!}{m!(n-m)!} \tag{A2}$$

Such a relation allows one to get the $(n+1)$-th derivatives at the specific point once all derivatives of lower order have been determined for all Q starting from the set $h_Q^{(0)}(\mathbf{x}) \equiv h_Q(\mathbf{x})$. Then, from Eq. 8 it follows

$$z_Q^{(n)}(\mathbf{x}) = \sum_{Q'} M_{Q,Q'}\, h_{Q'}^{(n)}(\mathbf{x}) \tag{A3}$$

for any order $n \geq 0$.

Appendix B: ILDMs construction

Let us first introduce the matrix $\mathbf{K}(\mathbf{x}) = -\mathbf{J}(\mathbf{x})$ where $\mathbf{J}(\mathbf{x})$ is the point-dependent Jacobian of the velocity field $\mathbf{F}(\mathbf{x})$. In what follows, the eigenspace of $\mathbf{K}(\mathbf{x})$ will play an important role. The eigenspace is determined through the solution of $\mathbf{K}(\mathbf{x})\mathbf{W}(\mathbf{x}) = \mathbf{W}(\mathbf{x})\mathbf{\Lambda}(\mathbf{x})$ with respect to the matrix $\mathbf{W}(\mathbf{x})$, whose columns are the right-eigenvectors of $\mathbf{K}(\mathbf{x})$, and to the diagonal matrix $\mathbf{\Lambda}(\mathbf{x})$ whose real or complex (but pair-conjugated) entries are the associated eigenvalues. The m-th eigenvalue and m-th eigenvector are denoted as $\lambda_m(\mathbf{x})$ and $\mathbf{w}_m(\mathbf{x})$ respectively. Finally, let the eigenvalues (and the corresponding eigenvectors) be listed according to the ascending order of their real parts, $\lambda_m^r(\mathbf{x})$.

Let $\mathbf{x}_c(t)$ be a reference trajectory, and $\mathbf{x}(t)$ a trajectory close to it; if mass-conservation constraints are present, we also require that $\mathbf{x}_c(t)$ and $\mathbf{x}(t)$ correspond to the same mass-conservation constants. The displacement vector is $\delta\mathbf{x}(t) = \mathbf{x}(t) - \mathbf{x}_c(t)$. A reference trajectory is considered to lie on an ILDM if the trajectories in its neighborhood rapidly converge

to it. Namely, for any choice of $\delta\mathbf{x}(0)$, a trajectory on an ILDM is such that, in a time window $0 \leq t \leq \Delta t$ with Δt sufficiently small, $\delta\mathbf{x}(t)$ evolves (in the sense of rotation and length's variation) in the way that the point $\mathbf{x}(\Delta t)$ falls *almost* on the reference trajectory and it is proximal to the point $\mathbf{x}_c(\Delta t)$. Let us elaborate such a picture.

For displacements that are small enough, the evolution of $\delta\mathbf{x}(t)$ can be described by $d\delta\mathbf{x}(t)/dt \simeq -\mathbf{K}(\mathbf{x}_c(t))\delta\mathbf{x}(t)$. Then consider a sufficiently small Δt, such that for $0 \leq t \leq \Delta t$ it is likely to assume that 1) $\mathbf{K}(\mathbf{x}_c(t)) \simeq \mathbf{K}(\mathbf{x}_c(0))$ is almost constant along the reference trajectory, and 2) the displacement $\delta\mathbf{x}(t)$ remains small. Under the fulfillment of conditions 1) and 2), the approximate evolution equation for $\delta\mathbf{x}(t)$ remains accurate and its integration is explicit: $\delta\mathbf{x}(\Delta t) \simeq \sum_m c_m(0) e^{-\lambda_m(\mathbf{x}_c(0))\Delta t}\, \mathbf{w}_m(\mathbf{x}_c(0))$, where the coefficients $c_m(0)$ are the components of the chosen initial $\delta\mathbf{x}(0)$ on the (non-orthogonal) eigenvectors. The ILDM assumption corresponds to having $\delta\mathbf{x}(\Delta t)$ *essentially* collinear to the velocity vector $\mathbf{F}(\mathbf{x}_c(\Delta t)) \simeq \mathbf{F}(\mathbf{x}_c(0))$, where it is assumed the smoothness of the velocity variation along the reference trajectory. It follows that $\sum_m c_m(0) e^{-\lambda_m(\mathbf{x}_c(0))\Delta t}\, \mathbf{w}_m(\mathbf{x}_c(0)) \propto \mathbf{F}(\mathbf{x}_c(0))$. By dropping the subscript "c" for the reference trajectory, a point $\mathbf{x}$ is considered to lie on an ILDM if $\mathbf{F}(\mathbf{x}) \simeq \kappa \sum_m c_m(0) e^{-\lambda_m(\mathbf{x})\Delta t}\, \mathbf{w}_m(\mathbf{x})$, with κ a proportionality factor. Now suppose that the eigenvalues can be partitioned into two subsets, one corresponding to the "low" eigenvalues labeled by the index m_l, and one to the "high" eigenvalues labeled by the index m_h. The separation between the two sets is established by the presence of an eigenvalue $\lambda_{m^*}(\mathbf{x})$ (or by a group of eigenvalues with equal real part as discussed below) such that

$$\lambda_1^r(\mathbf{x}) \leq \cdots \leq \lambda_{m^*-2}^r(\mathbf{x}) \leq \lambda_{m^*-1}^r(\mathbf{x}) \leq \lambda_{m^*}^r(\mathbf{x}) \;\ll\; \lambda_{m^*+1}^r(\mathbf{x}) \leq \lambda_{m^*+2}^r(\mathbf{x}) \leq \cdots \leq \lambda_N^r(\mathbf{x}) \tag{15}$$

All eigenvalues with $m_l \leq m^*$ form the "low" set, while the eigenvalues with $m_h > m^*$ form the "high" set. Such a sequence of inequalities *may be* converted, depending on the time-interval Δt, into inequalities between the exponential factors which enter the summation given above: if Δt is such that $e^{-\lambda_{m^*}^r(\mathbf{x})\Delta t} \gg e^{-\lambda_{m^*+1}^r(\mathbf{x})\Delta t}$, then the "high" terms in the summation are negligible and $\mathbf{F}(\mathbf{x})$ has a relevant component only on the "low" subspace (here it is assumed that the corresponding $c_{m_l}(0)$ are not null).

In summary, the conditions for $\mathbf{x}$ belonging to an ILDM are: a) existence of a spectral

gap in the real parts of the eigenvalues of $\mathbf{K}(\mathbf{x})$, and b) if a) is fulfilled, the components of the velocity vector $\mathbf{F}(\mathbf{x})$ on the "high" subspace must be negligible with respect to that on the "low" subspace.

Concerning the leading condition a), the possible gap is detected as follows. By taking into account the fact that there may exist degeneracies on the real parts of the eigenvalues, let us collect the degenerate eigenvalues into groups labeled by the index $i = 1, 2, \ldots$. The notation $\lambda^{(i)}(\mathbf{x})$ here below stands for the real part of the degenerate eigenvalues that belong to the i-th group (hence $\lambda^{(1)} < \lambda^{(2)} < \ldots$). Let us now consider a triad of consecutive groups, and the associated exponential factors $p_{i-1} = e^{-\lambda^{(i-1)}(\mathbf{x})\Delta t}$, $p_i = e^{-\lambda^{(i)}(\mathbf{x})\Delta t}$ and $p_{i+1} = e^{-\lambda^{(i+1)}(\mathbf{x})\Delta t}$. We say that a gap exists between the groups i and $(i+1)$ if $p_{i+1}/p_i \ll p_i/p_{i-1}$. This is equivalent to stating that while the exponential factors associated to the group i still have a relevant weight if compared to those of the group $(i-1)$, the exponential factors of the group $(i+1)$ (and also all higher factors taken as a whole) are negligible with respect to those of the i-th group. By introducing the parameter

$$\epsilon_i(\mathbf{x}) = 2\frac{\lambda^{(i)}(\mathbf{x}) - \lambda^{(i-1)}(\mathbf{x})}{\lambda^{(i+1)}(\mathbf{x}) - \lambda^{(i-1)}(\mathbf{x})} \quad \text{for } i \geq 2 \tag{16}$$

the inequality given above can be expressed as $e^{(\epsilon_i(\mathbf{X})-1)[\lambda^{(i+1)}(\mathbf{X})-\lambda^{(i-1)}(\mathbf{X})]\Delta t} \ll 1$. Recognizing that $\lambda^{(i+1)}(\mathbf{x}) - \lambda^{(i-1)}(\mathbf{x}) > 0$, a gap between the groups i and $(i+1)$ exists if $\epsilon_i(\mathbf{x}) \ll 1$. The fulfillment of condition a) hence corresponds to finding the (possible) *lowest* group i^* such that

$$\epsilon_{i^*}(\mathbf{x}) \leq \epsilon_{\text{ILDM}} \ll 1 \tag{17}$$

where the threshold value ϵ_{ILDM} has to be, unfortunately, subjectively chosen. If such a group is found, then the "low" set corresponds to all eigenvalues/eigenvectors of the groups from 1 to i^* (the "high" set is then defined by the eigenvalues/eigenvectors of the groups starting from $i^* + 1$). If none of the $\epsilon_i(\mathbf{x})$ fulfill the condition in Eq. 17, then the "low" set is constituted, by default, by the eigenvalues/eigenvectors of the group 1.

Concerning the condition b), the components of $\mathbf{F}(\mathbf{x})$ on the high and low sets are given by

$$\mathbf{F}_{l(h)}(\mathbf{x}) = \sum_{m_{l(h)}} [\mathbf{W}(\mathbf{x})^{-1}\mathbf{F}(\mathbf{x})]_{m_{l(h)}}\, \mathbf{w}_{m_{l(h)}}(\mathbf{x}) \quad , \quad \mathbf{F}(\mathbf{x}) = \mathbf{F}_l(\mathbf{x}) + \mathbf{F}_h(\mathbf{x}) \tag{18}$$

The fulfillment of condition b) is assessed by computing the following ratio between the Euclidean norms:

$$\eta(\mathbf{x}) = \frac{||\mathbf{F}_l(\mathbf{x})||}{||\mathbf{F}_h(\mathbf{x})||} \tag{19}$$

The points which pass the check of condition a) are then ranked according to the magnitude of $\eta(\mathbf{x})$, which should be greater than one to be consistent with the ILDM picture: as $\eta(\mathbf{x})$ is larger, the likelihood of the point $\mathbf{x}$ belonging to the SM proximity increases.

Note that the ILDM defined above is nothing but a locally attracting low-dimensional manifold without the specification "slow"; indeed it may even be a "fast" manifold in the presence of "low" eigenvalues with a negative and large real part. The characteristic "slow" is attributed by checking if the following condition holds:

$$\text{slow ILDM if} \quad |\lambda^{(i^*)}(\mathbf{x})| < |\lambda^{(i^*+1)}(\mathbf{x})| \tag{20}$$

This condition is equivalent to saying that the dominant exponential factors of the "low" set evolve slower, regardless of the fact that they decrease or increase, than each of the terms of the "high" set. If this condition is not fulfilled, we attribute the characteristic "fast" to the ILDM.

Appendix C: Mention of other strategies employing time derivatives to approximate the Slow Manifold

In what follows, $\mathbf{x}^{(n)}$ will denote the n-th time derivative of the state vector $\mathbf{x}$. Even if not indicated for sake of notation, it should be kept in mind that the components of $\mathbf{x}^{(n)}$ depend on $\mathbf{x}$. Although we shall refer to $\mathbf{x}$ as the concentration vector, we remark that all the methodologies mentioned below are applicable to the construction of the SM, even for dynamical systems different from mass-action based chemical kinetics.

We begin this brief overview by mentioning the zero-derivative principle (ZDP) of Gear et al.[38] By splitting the vector $\mathbf{x}$ in $\mathbf{x}_r$ and $\mathbf{x}_i$, where $\mathbf{x}_r$ stands for a subset of "relevant" (or observable) variables adopted to parametrize the SM, the ZDP approximation at the m-th order consists of searching for points in the concentration space where the $(m+1)$-th time derivatives of the remaining components are all null, that is $\mathbf{x}_i^{(m+1)} = \mathbf{0}$. Such a criterion

relies on the assumption that *some* suitable change of variables would convert the original system of ODEs into a singular perturbation format. The ZDP at order m is then equivalent to find the manifold where all the first $(m+1)$ terms of the "inner solution" (i.e., the fast-evolving component of the singular perturbation solution) are identically null. Remarkably, as m increases, the manifolds generated by the ZDP tend to the SM in the sense of Fenichel's definition (see 'Theorem 2.1' in Ref.[39])

Another approach is the flow curvature method (FCM) of Ginoux et al.[40] where a slow $(N-1)$-dimensional manifold is identified by the points of null "flow curvature" of the trajectories in the N-dimensional space. In our notation, the constitutive equation of such a manifold results in $\det(\mathbf{C}(\mathbf{x}))=0$, where $\mathbf{C}(\mathbf{x})$ is the $N \times N$ matrix whose n-th column is the vector of time derivatives $\mathbf{x}^{(n)}$ (compare with the original 'Proposition 2.1' in Ref.[40] and with the formulation in Ref.[13]). The iteration of the FCM by replacing the flow curvature with its successive time derivatives yields further dimensional reductions towards the SM.

Time derivatives of $\mathbf{x}$ have also been employed to build functionals for the localization of the SM via the trajectory-based variational principles of Lebiedz and coworkers.[12,13,20] A functional is constructed by taking the line integral of a *suitably chosen* function $\Phi(\mathbf{x})$ (the "objective function") along a trial trajectory piece of fixed time duration. A subset of relevant variables $\mathbf{x}_r$, as stated above, is adopted to parametrize the SM. At fixed $\mathbf{x}_r$, the target is to "reconstruct" the whole components of a candidate point to the SM. To this aim, the functional is minimized with respect to the trajectory piece, possibly enforcing conservation constraints, under the condition that for an intermediate point on the trajectory (which will be the produced point) the relevant variables $\mathbf{x}_r$ take the fixed values. Among the choices of $\Phi(\mathbf{x})$, the form $\Phi(\mathbf{x}) = ||\mathbf{x}^{(n)}||^2$ has been recently proposed[13]; here, $||\cdot||$ stands for the Euclidean norm of the vector at argument. The functional of order $n=2$ was employed in the early implementations of the method.[12,20] In such a case, the objective function takes the explicit form $\Phi(\mathbf{x}) = ||\mathbf{x}^{(2)}||^2 = ||\mathbf{J}(\mathbf{x})\mathbf{F}(\mathbf{x})||^2$ where $\mathbf{J}(\mathbf{x})$ is the Jacobian matrix of the velocity field $\mathbf{F}(\mathbf{x})$. As indicated in Ref.[20], the motivation underlying the choice of such an objective function is that $\Phi(\mathbf{x}) \leq ||\mathbf{J}(\mathbf{x})||\,||\mathbf{F}(\mathbf{x})||$ where $||\mathbf{J}(\mathbf{x})||$ stands for the 2-norm of the Jacobian matrix. Since low values of $||\mathbf{J}(\mathbf{x})||$ can be intuitively associated with "attractiveness of the SM" and low values of $||\mathbf{F}(\mathbf{x})||$ can be associated with the "slowness"

of the dynamics, the minimization of the functional should catch both these relevant features of the SM. The direct minimization of the functions $||\mathbf{x}^{(n)}||$, with respect to $\mathbf{x}_i$ at fixed $\mathbf{x}_r$, was also proposed by Girimaji[41] as a likely strategy to obtain approximations of the SM.

ACKNOWLEDGMENTS

This work has been funded by Università degli Studi di Padova through "Progetti di Ricerca di Ateneo" - PRAT2013. This work has been supported by the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070), funded by the European Regional Development Fund and the national budget of the Czech Republic via the Research and Development for Innovations Operational Programme, as well as Czech Ministry of Education, Youth and Sports via the project Large Research, Development and Innovations Infrastructures (LM2011033). We acknowledge PRACE for awarding us access to resource Abel based in Norway at Norwegian metacenter for High Performance Computing (NOTUR). We thank Dr. Mirco Zerbetto (Università degli Studi di Padova) for the preliminary tests of GPU programming and Dr. Benjamin J. Irving (Czech Technical University in Prague) for carefully reading the manuscript. Finally we thank a Reviewer for helpful suggestions and for having addressed us to relevant references.

Additional Supporting Information may be found in the online version of this article.

References

1. K. J. Laidler. *Chemical Kinetics*. 3^{rd} ed., Harper Collins Publishers, New York, 1987.

2. M. S. Okino and M. L. Mavrovouniotis. Simplification of mathematical models of chemical reaction systems. *Chem. Rev.*, 98(2):391, 1998.

3. S. Vajda, P. Valko, and T. Turnyi. Principal component analysis of kinetic models. *Int. J. Kinet. Chem.*, 17(1):55–81, 1985.

4. T. Turnyi, T. Brces, and S. Vajda. Reaction rate analysis of complex kinetic systems. *Int. J. Kinet. Chem.*, 21(2):83–99, 1989.

5. T. C. Ho and B. S. White. A general analysis of approximate nonlinear lumping in chemical kinetics. i. unconstrained lumping. *J. Chem. Phys.*, 101(2):1172–1187, 1994.

6. G. Li, A. S. Tomlin, H. Rabitz, and J. Tth. On the continuum approximation of large reaction mixtures. *AIChE J.*, 56(7):1894–1906, 2010.

7. S. J. Fraser. The steady state and equilibrium approximations: A geometric picture. *J. Chem. Phys.*, 88(8):4732–4738, 1988.

8. S. H. Lam and D. A. Goussis. The csp method for simplifying kinetics. *Int. J. Chem. Kinet.*, 26(4):461, 1994.

9. M. R. Roussel and S. J. Fraser. On the geometry of transient relaxation. *J. Chem. Phys.*, 94(11):7106–7113, 1991.

10. A. N. Al-Khateeb, J. M. Powers, S. Paolucci, A. J. Sommese, J. A. Diller, J. D. Hauenstein, and J. D. Mengers. One-dimensional slow invariant manifolds for spatially homogenous reactive system. *J. Chem. Phys.*, 131(2):024118, 2009.

11. R. T. Skodje and M. J. Davis. Geometrical simplification of complex kinetic systems. *J. Phys. Chem. A*, 105(45):10356–10365, 2001.

12. D. Lebiedz, J. Siehr, and J. Unger. A variational principle for computing slow invariant manifolds in dissipative dynamical systems. *SIAM J. Sci. Comput.*, 33(2):703–720, 2011.

13. D. Lebiedz and J. Unger. On fundamental unifying concepts for trajectory-based slow invariant attracting manifold computation in multiscale models of chemical kinetics. *Math. Comput. Model. Dyn.*, 22:87–112, 2016.

14. P. Nicolini and D. Frezzato. Features in chemical kinetics. ii. a self-emerging definition of slow manifolds. *J. Chem. Phys.*, 138(23):234102, 2013.

15. H. G. Kaper and T. J. Kaper. Asymptotic analysis of two reduction methods for systems of chemical reactions. *Physica D*, 165:66–93, 2002.

16. C. K. R. T. Jones. *"Geometric singular perturbation theory" in Dynamical Systems (Vol. 1609, edited by L. Arnold).* Springer-Verlag, Berlin, 1994, 1994.

17. A. Zagaris, H. G. Kaper, and T. J. Kaper. Analysis of the computational singular perturbation reduction method for chemical kinetics. *J. Nonlinear Sci.*, 14(1):59, 2004.

18. M. R. Roussel and S. J. Fraser. Invariant manifold methods for metabolic model reduction. *Chaos*, 11(1):196, 2001.

19. U. Maas and S. B. Pope. Symplifying chemical kinetics: Intrinsic low-dimensional manifolds in composition space. *Combust. Flame*, 88:239–264, 1992.

20. D. Lebiedz and J. Siehr. A continuation method for the efficient solution of parametric optimization problems in kinetic model reduction. *SIAM J. Sci. Comput.*, 35(3):A1584–A1603, 2013.

21. M. J. Davis and R. T. Skodje. Geometric investigation of low-dimensional manifolds in systems approaching equilibrium. *J. Chem. Phys.*, 111(3):859–874, 1999.

22. A. N. Gorban and I. V. Karlin. Method of invariant manifold for chemical kinetics. *Chem. Eng. Sci.*, 58:4751, 2003.

23. D. Lebiedz. Computing minimal entropy production trajectories: an approach to model reduction in chemical kinetics. *J. Chem. Phys.*, 120(15):6890, 2004.

24. V. Reinhardt, M. Winckler, and D. Lebiedz. Approximation of slow attracting manifolds in chemical kinetics by trajectory-based optimization approaches. *J. Phys. Chem. A*, 112(8):1712, 2008.

25. D. Lebiedz. Entropy-related extremum principles for model reduction of dissipative dynamical systems. *Entropy*, 12(4):706, 2010.

26. DRIMAK is distributed under the General Public License v2.0. Software and documentation are vailable at: `http://www.chimica.unipd.it/licc/software.html`.

27. P. Nicolini and D. Frezzato. Features in chemical kinetics. i. signatures of self-emerging dimensional reduction from a general formal of the evolution law. *J. Chem. Phys.*, 138(23):234101, 2013.

28. A. Ceccato, P. Nicolini, and D. Frezzato. Features in chemical kinetics. iii. attracting subspaces in a hyper-spherical representation of the reactive system. *J. Chem. Phys.*, 143(22):224109, 2015.

29. A. N. Gorban, I. V. Karlin, and A. Y. Zynovyev. Constructive methods of invariant manifolds for kinetic problems. *Physics Reports*, 396(4-6):197–403, 2004.

30. J. Li, Z. Zhao, A. Kazakov, and F. L. Dreyer. An updated comprehensive kinetic model of hydrogen combustion. *Int. J. Chem. Kin.*, 36(10):566–575, 2004.

31. L. Brenig and A. Goriely. Universal canonical forms for time-continuous dynamical systems. *Phys. Rev. A*, 40(7):4119–4122, 1989.

32. B. Hernández-Bermejo and V. Fairén. Nonpolynomial vector fields under the lotka-volterra normal form. *Phys. Lett. A*, 206(1-2):31–37, 1995.

33. V. Fairén and B. Hernández-Bermejo. Mass action law conjugate representation for general chemical mechanisms. *J. Phys. Chem.*, 100(49):19023–19028, 1996.

34. J. L. Gouzé. Transformation of polynomial differential systems in the positive orthant. Technical report, INRIA, Sophia-Antipolis, 06561, Valbonne, France, 1996.

35. Download link for LINCOA. `http://mat.uc.pt/~zhang/software.html`. Last view: 23^{th} September 2016.

36. M. J. D. Powell. On trust region methods for unconstrained minimization without derivatives. *Math. Program., Ser. B*, 97:605–623, 2003.

37. M. J. D. Powell. On fast trust region methods for quadratic models with linear constraints. *Mathematical Programming Computation*, 7(3):237–267, 2015.

38. C. W. Gear, T. J. Kaper, I. G. Kevrekidis, and A. Zagaris. Projecting to a slow manifold: singularly perturbed systems and legacy codes. *SIAM J. Applied Dynamical Systems*, 4(3):711–732, 2005.

39. A. Zagaris, C. W. Gear, T. J. Kaper, and Y. G. Kevrekidis. Analysis of the accuracy and convergence of equationfree projection to a slow manifold. *ESAIM: Math. Model. Num.*, 43:757–784, 2009.

40. J.-M. Ginoux, B. Rossetto, and L. Chua. Slow invariant manifolds as curvature of the flow of dynamical systems. *International Journal of Bifurcation and Chaos*, 18(11):3409–3430, 2007.

41. S. S. Girimaji. Reduction of large dynamical systems by minimization of evolution rate. *Phys. Rev. Lett.*, 82:2282 – 2285, 1999.

Figure 1: a) Average time for the computation of Z as function of N and M. Blue marks are the calculated points whilst the surface is obtained by interpolating these points with the expression of $\tau_{\boldsymbol{\alpha}}(N, M)$ in Eq. 11 (fit parameters: $\alpha_1 = 4.24 \cdot 10^{-7}$ s, $\alpha_2 = 1.72 \cdot 10^{-7}$ s, $\alpha_3 = -3.48 \cdot 10^{-7}$ s, $\alpha_4 = -1.28 \cdot 10^{-8}$ s, $\alpha_5 = 4.89 \cdot 10^{-8}$ s, $\alpha_6 = 6.98 \cdot 10^{-8}$ s). b) The same as in panel a), here for the computation of Z_1 (fit parameters $\alpha_1 = 1.02 \cdot 10^{-6}$ s, $\alpha_2 = 3.26 \cdot 10^{-7}$ s, $\alpha_3 = -7.44 \cdot 10^{-7}$ s, $\alpha_4 = -2.44 \cdot 10^{-8}$ s, $\alpha_5 = 9.79 \cdot 10^{-8}$ s, $\alpha_6 = 1.11 \cdot 10^{-7}$ s). Large red marks correspond to the computational times for the two models of hydrogen combustion considered in this study.

Figure 2: Projection on the subspace of the radical species for the basic hydrogen combustion mechanism, Scheme A. Blue dots are 2000 candidate points produced by DRIMAK and the larger red marks are the "filtered" results according to the ILDM-based strategy. The large green circle corresponds to the equilibrium point.

Figure 3: Two dimensional projections for the basic hydrogen combustion mechanism, Scheme A. Blue dots are 2000 candidate points produced by DRIMAK and the larger red marks are the "filtered" results according to the ILDM-based strategy. The large green circle corresponds to the equilibrium point.

Figure 4: Three dimensional subspace of the three main species H_2, O_2 and H_2O of Scheme B. Blue dots are 2000 candidates points produced by DRIMAK and the larger red marks are the "filtered" results according to the ILDM-based strategy. Concentrations are expressed in mol/L.

Figure 5: Two dimensional projections for the main species of Scheme B. Blue dots are 2000 candidate points produced by DRIMAK and the larger red marks are the "filtered" results according to the ILDM-based strategy. Concentrations are expressed in mol/L.

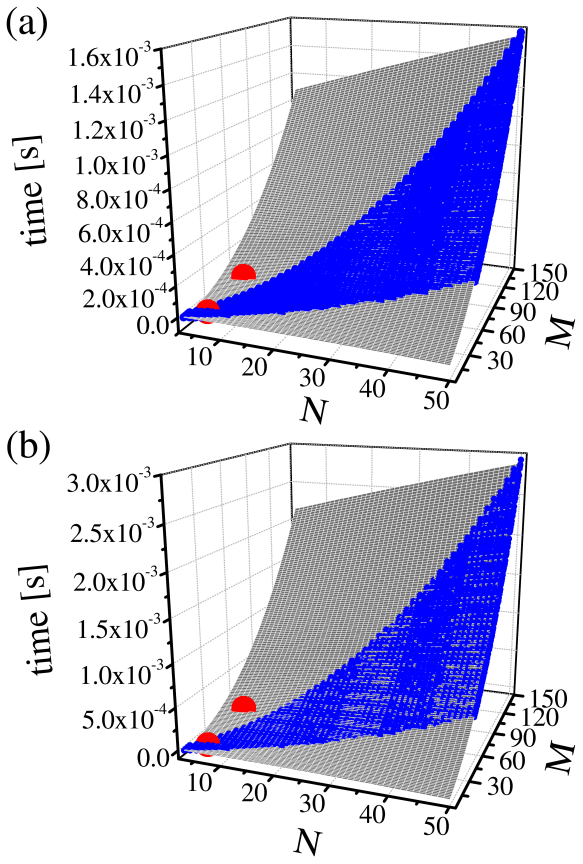
(a)
time [s]
1.6x10-3
1.4x10-3
1.2x10-3
1.0x10-3
8.0x10-4
6.0x10-4
4.0x10-4
2.0x10-4
0.0
10
20
30
40
50
N
150
120
90
60
30
M
(b)
time [s]
3.0x10-3
2.5x10-3
2.0x10-3
1.5x10-3
1.0x10-3
5.0x10-4
0.0
10
20
30
40
50
N
150
120
90
60
30
M

Figure 1

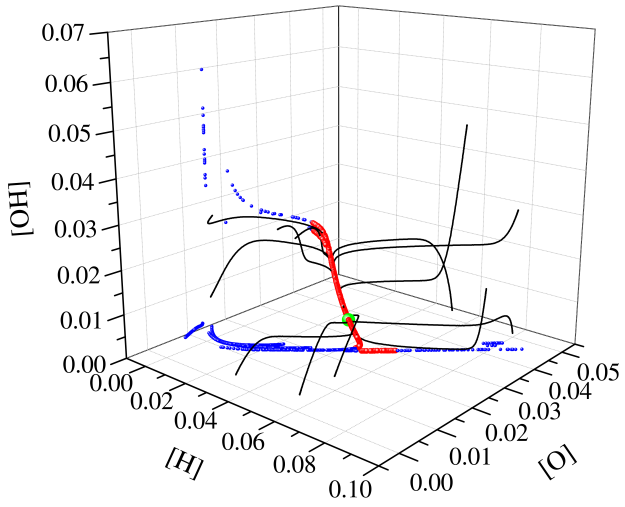
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
[OH]
0.00
0.02
0.04
0.06
0.08
0.10
[H]
0.00
0.01
0.02
0.03
0.04
0.05
[O]

Figure 2

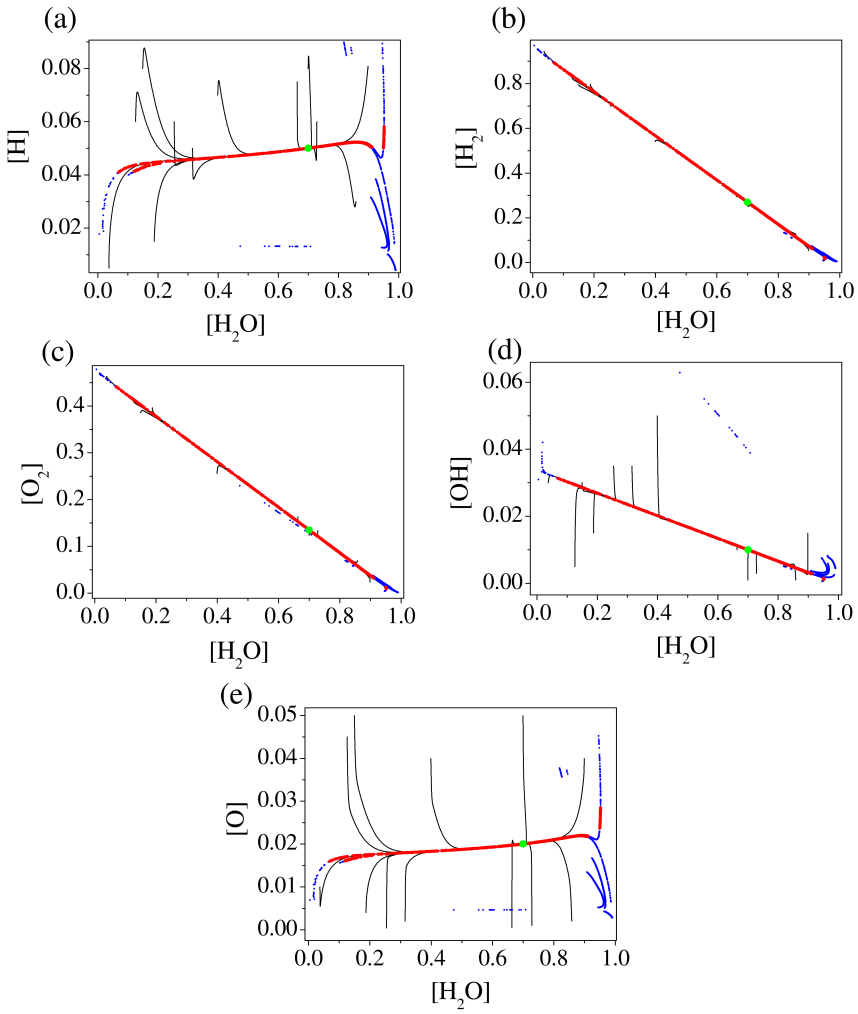

Figure 3

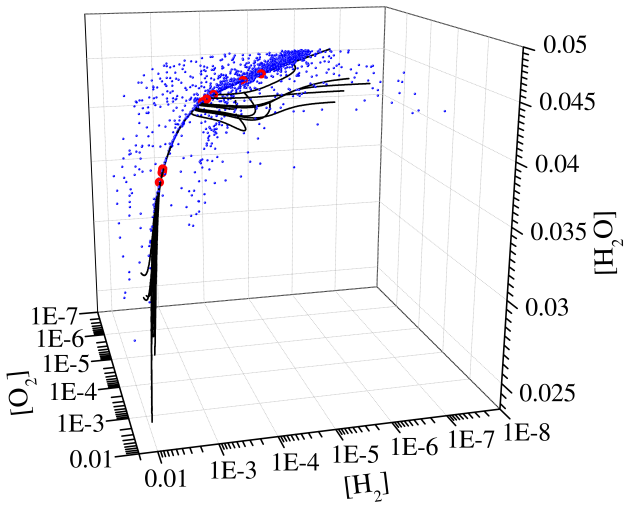
0.05
0.045
0.04
0.035
0.03
0.025
[H2O]
1E-7
1E-6
1E-5
1E-4
1E-3
0.01
[O2]
0.01
1E-3
1E-4
1E-5
1E-6
1E-7
1E-8
[H2]

Figure 4

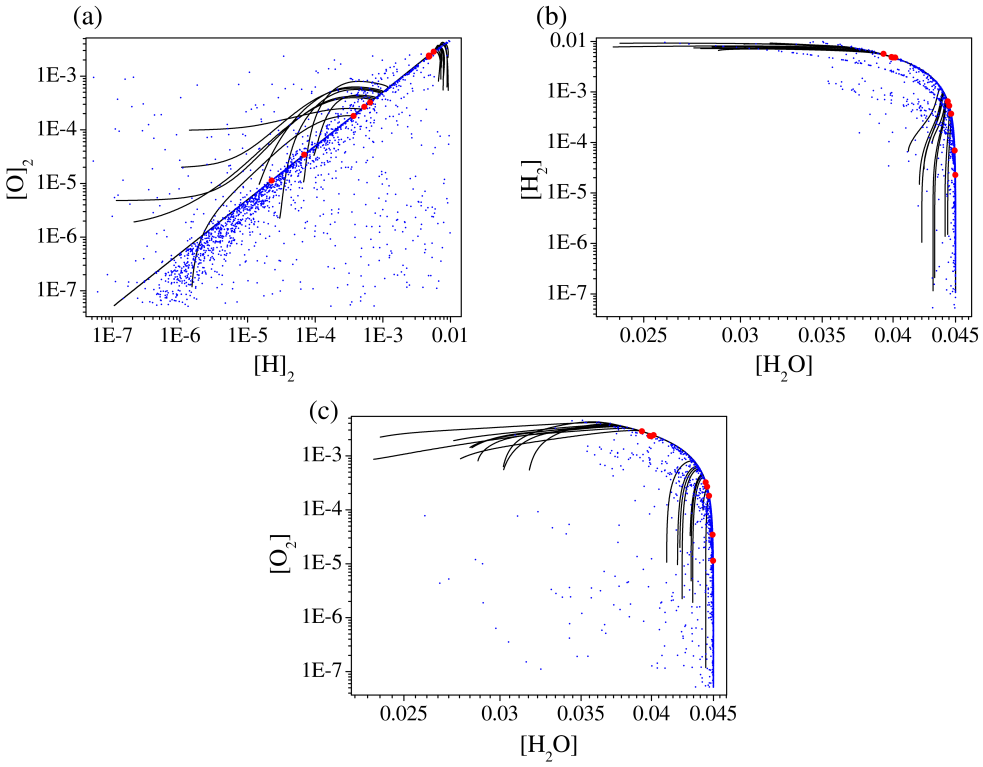
(a)
1E-3
1E-4
1E-5
1E-6
1E-7
[O]2
1E-7
1E-6
1E-5
1E-4
1E-3
0.01
[H]2
(b)
0.01
1E-3
1E-4
1E-5
1E-6
1E-7
[H2]
0.025
0.03
0.035
0.04
0.045
[H2O]
(c)
1E-3
1E-4
1E-5
1E-6
1E-7
[O2]
0.025
0.03
0.035
0.04
0.045
[H2O]

Figure 5