Applying the quasi-steady-state approximation to the autocatalytic zymogen activation: A novel general methodology for scaling analysis

Justin Eilertsen[1], Malgorzata Anna Tyczynska[1], and Santiago Schnell[1,2,3]

[1]*Department of Molecular & Integrative Physiology, University of Michigan Medical School, Ann Arbor, Michigan, USA*
[2]*Department of Computational Medicine & Bioinformatics, University of Michigan Medical School, Ann Arbor, Michigan, USA*
[3]*E-mail: schnells@umich.edu*

Abstract

A zymogen is an inactive precursor of an enzyme that needs to go through a chemical change to become an active enzyme. The general intermolecular mechanism for the autocatalytic activation of zymogens is governed by the single-enzyme, single-substrate catalyzed reaction following the Michaelis–Menten mechanism of enzyme action, where the substrate is the zymogen and the product is the same enzyme that is catalyzing the reaction. In this article we investigate the nonlinear chemical dynamics of the intermolecular autocatalytic zymogen activation reaction mechanism. In so doing, we develop a general strategy for obtaining dimensionless parameters that, when sufficiently small, legitimize the application of the quasi-steady-state approximation. Our methodology combines energy methods and exploits the phase-plane geometry of the mathematical model, and we obtain sufficient conditions that support the validity of the standard, reverse and total quasi-steady-state approximations for the intermolecular autocatalytic zymogen activation reaction mechanism. The utility of the procedure we develop is that it circumnavigates the direct need for a priori timescale estimation, scaling, and non-dimensionalization. At the same time, a novel result emerges from our analysis: the discovery of a dynamic transcritical bifurcation that exists in the singular limit of the model equations. Moreover, associated with the dynamic transcritical bifurcation is an imperfect term. We prove that when the imperfect term vanishes and the singular vector field is perturbed, there exists a canard that follows a repulsive slow invariant manifold over timescales of $\mathcal{O}(1)$. This is the first report of such a solution for the intermolecular and autocatalytic zymogen activation reaction. By extension, our results illustrate that canards also exist in the classic single-enzyme, single-substrate reversible Michaelis–Menten reaction mechanism.

1 Introduction

Zymogens are enzyme precursors (proenzymes) that can be activated through non-catalytic [1] or catalytic reactions, and have numerous biochemical functions. For example, they play a critical role in protein digestion by converting pepsin to pepsinogen [1, 32], and trypsin to trypsinogen [19, 20, 40]. Activation can occur in three ways: (i) the inactive enzyme can be activated by another enzyme that cleaves off a peptide unit; this is the mechanism we consider in this work, (ii) the configuration of the zymogen can be changed in order to reveal the activation site, and (iii) the inactive substance is activated when a coenzyme binds to the zymogen. A simple autocatalytic reaction that utilizes mechanism (i) is the activation of trypsinogen by trypsin [48, 13], while the activation of pepsinogen to active pepsin at low pH [30] follows mechanism (ii).

Mechanism (i) is represented schematically by

$$Z + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \xrightarrow{k_2} 2E + W, \tag{1}$$

where S is a zymogen, E is an active enzyme, W is peptide, and k_1, k_2 and k_{-1} are rate constants. We refer to (1) as the intermolecular a utocatalytic zymogen activation (IAZA) reaction. Fuentes et al. [11] proposed a mechanism for zymogen activation that consists of two reaction pathways. One pathway consists of an intramolecular non-catalytic step in which the zymogen molecule disassociates into an active enzyme and a peptide

$$S \xrightarrow{k_d} E + W, \tag{2}$$

while the additional catalytic step in the second pathway follows the IAZA reaction mechanism (1).

To derive a deterministic (and finite-dimensional) model of the IAZA reaction mechanism (1), we let $z := z(t)$, $e := e(t)$ and $c := c(t)$ denote the concentrations of Z, E, and C respectively, and apply the law of mass action to (1). This yields the following set of nonlinear ordinary differential equations:

$$\frac{\mathrm{d}z}{\mathrm{d}t} = -k_1 ez + k_{-1}c, \tag{3a}$$

$$\frac{\mathrm{d}c}{\mathrm{d}t} = k_1 ez - (k_{-1} + k_2)c, \tag{3b}$$

$$\frac{\mathrm{d}e}{\mathrm{d}t} = -k_1 ez + (k_{-1} + 2k_2)c, \tag{3c}$$

$$\frac{\mathrm{d}w}{\mathrm{d}t} = k_2 c. \tag{3d}$$

The model equations (3a)–(3d) must of course be equipped with appropriate initial conditions. A typical set used in experiments is

$$z(0) = z_0, \qquad e(0) = e_0, \qquad c(0) = 0, \qquad w(0) = 0, \tag{4}$$

and, unless otherwise stated, we will assume (4) holds in the analysis that follows.

The structure of the model equations implies the existence of conservation laws. Note that equations (3a)–(3c) obey

$$\frac{\mathrm{d}z}{\mathrm{d}t} + \frac{\mathrm{d}c}{\mathrm{d}t} + \frac{\mathrm{d}w}{\mathrm{d}t} = 0, \tag{5}$$

$$\frac{\mathrm{d}e}{\mathrm{d}t} + \frac{\mathrm{d}c}{\mathrm{d}t} - \frac{\mathrm{d}w}{\mathrm{d}t} = 0, \tag{6}$$

and their sum yields an equation that is independent of w:

$$\frac{\mathrm{d}z}{\mathrm{d}t} + \frac{\mathrm{d}e}{\mathrm{d}t} + 2\frac{\mathrm{d}c}{\mathrm{d}t} = 0. \tag{7}$$

Integrating (7) with respect to (4) yields the conservation law $z + e + 2c = E_T$, where E_T is sum of initial concentrations of zymogen and enzyme: $E_T = z_0 + e_0$. Substituting $e = E_T - z - 2c$ into (3a)–(3b) yields

$$\frac{\mathrm{d}z}{\mathrm{d}t} = k_1\Big[-(E_T - z)z + (K_S + 2z)c\Big], \tag{8a}$$

$$\frac{\mathrm{d}c}{\mathrm{d}t} = k_1\Big[(E_T - z)z - (K_M + 2z)c\Big], \tag{8b}$$

$$\frac{\mathrm{d}w}{\mathrm{d}t} = k_2 c, \tag{8c}$$

where $K_S = k_{-1}/k_1$ is the equilibrium constant of the intermediate complex, $K = k_2/k_1$ is the Van Slyke–Cullen constant, and $K_M = K_S + K$ denotes the Michaelis constant of the IAZA reaction respectively. Finally, summing together (8a)–(8c) and integrating yields the additional conservation law

$$z_0 = z + c + w, \tag{9}$$

and thus we can compactly express the mass actions equations in terms of two variables: $(z, c), (w, z)$ or (w, c). Moreover, when we obtain a reduced model in any of the three two-variable coordinate systems (i.e., $(z, c), (w, z)$ or (w, c)), the conservation laws allow for the determination of the reduced model that approximates the evolution of the additional variable.

Although the complete system (3a)–(3d) is expressible in terms of only two variables, there is a two-part question that we need to answer: is it possible to reduce the system (8a)–(8c) further and, if so, how favorable is such a reduction with respect to a range of parameters? The answer to this question is critical for the derivation of approximate mathematical models that are more favorable (well-suited) for estimating the pertinent kinetic parameters, K_M and k_2. Several authors attempted to address this very question. An early attempt to construct a reduced model for the IAZA reaction was made by García-Moreno et al. [13]. These authors constructed what is known as a pseudo-first-order (PFO) approximation. Briefly, they assumed that $e_0 \ll z_0$, and that $z \approx z_0$ during the initial phase of the reaction. Replacing z with z_0 in (8a)–(8c) results in a set of linear equations that can be easily be solved. Unfortunately, there is a downside to the PFO approach: the approximation is only valid at the onset of the reaction, and therefore its validity is limited to very short (brief) timescales. Nevertheless, and despite the limitations of the linear approximation that arises from the PFO procedure, the analysis conducted in García-Moreno et al. [13] was essentially the first study in which considerable progress was made towards our understanding of autocatalytic zymogen activation reaction kinetics. Prior to the work of García-Moreno et al. [13], mostly "over-simplified" reaction mechanisms describing zymogen activation were studied [38, 39]. Later investigations [11, 45, 44] added complexity to the IAZA reaction mechanism (1), but the kinetic analyses conducted in these studies employed the basic methodology of García-Moreno et al. [13].

A slightly different and more rigorous approach was developed by Wu et al. [48]. Instead of constructing a reduced model by assuming PFO kinetics, these authors analyzed a reduced model that is analogous to the total quasi-steady-state approximation (tQSSA) of the Michaelis–Menten (MM) reaction mechanism. In the context of the IAZA reaction mechanism, the tQSSA approximates the temporal evolution the total enzyme concentration, $e_T = e + c$, where c is the enzyme-zymogen intermediate complex. While c is not experimentally measurable under steady-state conditions, the dynamics of e_T can be extracted from the dynamics of w, courtesy of the conservation laws. Wu et al. [48] were motivated by the inverse problem, and their aim was to estimate k_2 and K_M from experimental time course data. Amazingly, by using the tQSSA formulation, Wu et al. [48] found that their estimated k_2 and K_M values were close to those obtained in an earlier study by García-Moreno et al. [13]. Additional studies utilized the tQSSA fitting method of Wu et al. [48] to estimate kinetic parameters for ligand-induced autocatalytic reactions [24], and autocatalytic catalytic reactions that contain competitive inhibitors [46]. However, despite the promising results that emerged from the tQSSA, Wu et al. [48] did not establish a suitable criterion that guarantees the validity of the tQSSA. Thus, it is impossible to determine the appropriate bounds within which their protocol for estimating k_2 and K_M from (1) can be utilized.

Despite the prevalence of the basic IAZA reaction mechanism (1) in the biochemical literature [45, 43, 11, 12], there has been no attempt to rigorously derive reduced models and the conditions that support their legitimacy. We are interested in deriving reduced models that are valid over long timescales so as

3

not to greatly limit the domain over which experimental measurements can be taken. Thus, in the sections that follow, we will look for regions in parameter space where the IAZA reaction is comprised of disparate timescales, and use singular perturbation theory to covert our law of mass model (8a)–(8c) to a differential algebraic equation (DAE) model that is analogous to the MM equation. To obtain quasi-steady-state approximations (QSSAs) for the system (8a)–(8c), we will assume that the reaction consists of a transient or pre-quasi-steady-state phase in which the complex concentration accumulates to its maximum very quickly, followed by a quasi-steady-state (QSS) phase during which

$$\frac{\mathrm{d}c}{\mathrm{d}t} \approx 0. \tag{10}$$

If the transient phase is short in comparison to the QSS phase, then the majority of the reaction will unfold in the QSS phase. To approximate the kinetics in the QSS phase, we will assume a priori that

$$c \sim \frac{(E_T - z)z}{K_M + 2z}, \qquad \text{in } (z, c) \text{ variables, or} \tag{11a}$$

$$c \sim \frac{1}{2}\left(E_T + K_M - \sqrt{(K_M + E_T)^2 - 4(e_0 + w)(z_0 - w)} \right), \qquad \text{in } (w, c) \text{ variables.} \tag{11b}$$

The task that follows is thus to determine the condition(s) under which the approximations (11) are valid. As we will review in Section 2, singular perturbation theory has been employed to justify approximations of the form (11). Unfortunately, in order use singular perturbation theory, one must non-dimensionlize the model equations (8a)–(8c), and this requires a priori knowledge of the maximum values of the concentrations of the various chemical species, as well as accurate estimates of the transient and QSS timescales. Furthermore, timescale estimation may prove to be difficult in certain parameter regimes, and accurate timescale estimation has been shown to be necessary in order to derive conditions for the validity of a quasi-steady-state approximation (QSSA) directly from singular perturbation theory [34]. However, as we will demonstrate in this work, it is possible to avoid the impromptu scaling procedure by exploiting phase-plane geometry and utilizing energy methods to derive approximations. More importantly, our approach systematically yields the sufficient conditions that support the validity of the three quasi-steady state approximations – standard, reverse and total QSSA – for IAZA reaction mechanism.

2 Mathematical approaches for the application of the QSSA

An essential tool in the modeling of enzyme-catalyzed reactions is the QSSA. Briefly, under very specific physio-chemical conditions, enzyme-catalyzed reactions consist of two phases: a brief transient phase followed by a QSS phase. During the QSS phase, the concentration of at least one chemical species changes very slowly, and it is this comparatively slow change that permits a simplification, referred to as the QSSA, in the mathematical equations that model the reaction. As a reduced mathematical model of an enzyme-catalyzed reaction, the QSSA is a low-dimensional approximation to the full set of nonlinear model equations. The low-dimensionality of the QSSA is attractive for two reasons. First, the simplified QSSA model is typically more amenable for the estimation of kinetic parameters. Second, it relaxes the computational burden of having to numerically estimate the solution to a set of stiff equations.

Without a doubt, the best-known example of a deterministic QSSA is the MM equation[1]

$$\frac{\mathrm{d}s}{\mathrm{d}t} =\approx -\frac{Vs}{K_M + s} \approx -\frac{\mathrm{d}p}{\mathrm{d}t} \qquad \text{where} \qquad V \equiv k_2 e_0, \quad \text{and} \quad K_M = \frac{k_{-1} + k_2}{k_1}, \tag{12}$$

which provides an approximate mathematical description of the conversion of substrate, S, to product, P, when the conversion of S to P is catalyzed via the MM reaction mechanism:

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \overset{k_2}{\longrightarrow} E + P. \tag{13}$$

---

[1]Lowercase s and p denote concentrations of S and P, respectively.

In (13), C denotes the intermediate complex that consists of bound enzyme, E, and substrate molecules. k_1, k_{-1}, and k_2 are rate constants that regulate the speed of each elementary reaction. The power behind the MM equation (12) is twofold. First, it is autonomous and, after the initial transient phase, the concentrations of all of the chemical species that comprise the MM reaction mechanism (13) can be extracted from equation (12). Thus, during the QSS phase, the time courses of e, c and p can be determined solely from $s(t)$. Second, it consists of only two parameters instead of three: the Michaelis constant, K_M, and the limiting rate, V.

In the context of the MM reaction mechanism, the acronym "QSSA" refers to a family of simplified models [8]. For example, the MM equation (12) is also referred to as the *standard quasi-steady-state approximation* (sQSSA), but other QSSAs, such as the *total* (tQSSA) and the *reverse* (rQSSA), can be employed as long as long as certain conditions are met. Different experimental conditions or applications may warrant the use of a specific QSSA, and the only real obstacle in the implementation of *any* QSSA is its justification, as they are only valid when the reaction is comprised of disparate timescales [35]. The sQSSA is typically assumed to be valid when the initial substrate concentration, s_0, is much greater than the initial enzyme concentration, e_0: $e_0 \ll s_0$ [see 17, for a rigorous justification]. This separation of initial concentrations induces a timescale separation, because the complex concentration rapidly reaches its threshold value, and the time it takes c to become maximal is extremely short in comparison to the time it takes s_0 to substantially deplete. As a result, enzyme-catalyzed reactions that obey the MM reaction mechanism can easily be prepared in laboratory experiments so that (12) is legitimate, and it is this particular feature that makes the MM equation an attractive model to implement for the experimental estimation of kinetic parameters and the quantification of enzyme activity [37].

Although the qualifier that certifies the validity of the MM equation is easy to enforce in a controlled experimental assay (i.e., $e_0 \ll s_0$), the MM equation is often the subject of criticism when it comes to modeling the enzyme catalyzed reaction outside the $e_0 \ll s_0$ domain [6, 42]. However, this does not mean that the QSSA cannot be applied if the condition $e_0 \ll s_0$ does not hold: enter the tQSSA [23, 4]. In the tQSSA, the substrate and complex concentrations are lumped together to form a single dynamic variable called the *total* substrate, $s_T = s + c$. Just like the sQSSA, the approximate time-dependency of the chemical species that constitute the MM reaction mechanism can be determined uniquely from s_T:

$$\frac{\mathrm{d}s_T}{\mathrm{d}t} \approx -\frac{k_2}{2}\left(e_0 + K_M + s_T - \sqrt{(e_0 + K_M + s_T)^2 - 4e_0 s_T}\right). \tag{14}$$

It is important to note that the variable s_T is not experimentally measurable. However, like the sQSSA, preliminary studies suggest that the tQSSA is also a useful model from which to quantify enzyme activity and estimate kinetic parameters by exploiting the reaction conservation laws [6]. The power of the approximation (14) lies on the wide range of physical conditions under which is valid [29, 4, 42, 3].

The versatility of the QSSA should not be understated: it essential in the experimental quantification of enzyme activity, as well as in the computational modeling of biochemical reactions. The real question therefore pertains to the procedure by which scientists determine the extent of its validity. Historically, scaling methods have been used to determine the precise qualifiers that safeguard the use of the QSSA. This is because the QSSA is derived from Tikhonov theory [41], and the validity of the QSSA coincides with regions in parameter space where the model equations are singularly perturbed. Thus, scaling, in the context of the QSSA, is really the art of uncovering the singular perturbation parameter that permits the application of the QSSA. However, scaling is mostly an ad hoc procedure that relies heavily on a priori knowledge of the specific dynamical behavior of the reaction. And, if the history of the validity of the sQSSA has taught us anything, it is that different scales often yield different singular perturbation parameters [17, 31, 33, 28, 35]. Thus, it might be said that, due to the non-uniqueness of singular perturbation parameters, the scaling procedure is in some sense an ill-posed methodology, since it has the capacity to permit more than one solution. Moreover, even when a singular perturbation parameter derived from scaling analysis is agreed upon by scientists to be correct,[2] the scaling procedure can often obfuscate important nonlinear phenomena such as dynamic bifurcations. For example, although mathematical analyses of the MM reaction mechanism have been carried out for well over sixty years, it was only recently shown that a dynamic transcritical bifurcation is present in the phase-plane of the model equations [8]. The significance of this bifurcation is

---

[2]For example, it is well-established that the condition $e_0 \ll s_0 + K_M$ derived by Segel [35] is the "correct" condition that supports the validity of the sQSSA. However, Palsson [27] noted that the true condition is $e_0/k_M \ll 1$, and was absolutely correct in his assessment.

that it is responsible for the precise partitioning in parameter space that separates the validity of the sQSSA and rQSSA. Until its discovery, the domain of the rQSSA was thought to be much more restrictive than it actually is. Most importantly, even if a singular perturbation parameter is known, the presence of a dynamic bifurcation is important, because it can signal a temporary loss of asymptotic accuracy.

Even when improvised scaling methods are successfully employed, timescales are the hardest scales to obtain, but also the most important, since the criteria for establishing the validity of the QSSA is directly related to the ratio of fast and slow timescales. For example, in the deterministic domain, the tQSSA is assumed to be valid as long as the fast and slow timescales are widely separated. There are however two potential drawbacks in the utilization of timescales to determine the validity of the QSSA. First, while timescales are extremely well-defined when the model equations are linear, it is not always clear how to obtain accurate timescale estimates when the model equations are nonlinear. Accordingly, one generally has to rely heavily on ad hoc methods in the nonlinear regime and, while such methods have had moderate success in the analysis of low-dimensional systems, the degree of success timescale methods will have when applied to higher-dimensional problems is unclear. In short, the techniques developed to analyze low-dimensional problems may loose their efficacy when applied to higher dimensional systems. Second, timescales are by definition *local*, meaning they depend strongly on initial conditions. The potential problem with this is that timescales may not adequately reflect the accuracy of the QSSA downstream of the initial conditions.

An alternative to scaling and direct perturbation techniques are geometric methodologies that combine phase-plane geometry with energy methods. This combination relies on the mathematical analysis of the phase-plane geometry, and thus avoids the need for a priori scaling and non-dimensionalization. Our purpose in this work is to develop a geometric methodology that is capable of determining conditions that authorize the validity of the QSSA for enzyme-catalyzed reactions.

3 Investigating the phase–plane geometry of the IAZA reaction mechanism

In this section, our aim is to capture relevant phase-plane geometry that can be used to construct QSSAs. To start, we write the mass action equations in (w, c) coordinates by invoking the conservation law (9):

$$\dot{c} = k_1(e_0 + w - c)(z_0 - c - w) - k_1 K_M c, \tag{15a}$$

$$\dot{w} = k_2 c. \tag{15b}$$

Note that there are two nullclines in the (w, c) phase–plane

$$c = h^-(w) = \frac{1}{2}(E_T + K_M) - \frac{1}{2}\sqrt{(E_T + K_M)^2 - 4(z_0 - w)(e_0 + w)}, \tag{16a}$$

$$c = h^+(w) = \frac{1}{2}(E_T + K_M) + \frac{1}{2}\sqrt{(E_T + K_M)^2 - 4(z_0 - w)(e_0 + w)}. \tag{16b}$$

Since the time derivative of w vanishes along the curve $c = 0$, and time derivative of c vanishes along nullclines (16), the intersection of (16) with $c = 0$ provides a set of equilibra. The nullcline $c = h^-(w)$ intersects with $c = 0$ at the points

$$x^{(1)} = (z_0, 0), \quad x^{(2)} = (-e_0, 0). \tag{17}$$

Linearization reveals that $x^{(1)}$ is a stable equilibrium point and $x^{(2)}$ is a saddle point. Since $x^{(2)}$ is a saddle point, it comes equipped with an invariant stable manifold, $W^s(x^{(2)})$, as well as an invariant unstable manifold, $W^u(x^{(2)})$,

$$W^s(x^{(2)}) = \{(w, c) \in \mathbb{R}^2 : \lim_{t \to \infty} \varphi^t(x) = x^{(2)}\}, \tag{18a}$$

$$W^u(x^{(2)}) = \{w, c) \in \mathbb{R}^2 : \lim_{t \to -\infty} \varphi^t(x) = x^{(2)}\}, \tag{18b}$$

where $\varphi^t(x)$ denotes the evolution operator of the vector field, and satisfies:

$$\varphi^0(x) = x, \qquad \varphi^{-t}\varphi^t(x) = \varphi^0(x) = x, \qquad \varphi^t(\varphi^s(x)) = \varphi^{t+s}(x). \tag{19}$$

While the saddle point $x^{(2)}$ is *unphysical* in the sense that it represents a negative concentration, its presence indicates the presence of an invariant manifold on which $\lim_{t\to\infty}\varphi^t(x) = x^{(1)}$, and one can easily deduce that the unstable manifold of $x^{(2)}$ "connects" with $x^{(2)}$, forming a heteroclinic orbit, $\mathcal{H}$ (see Figure 1).
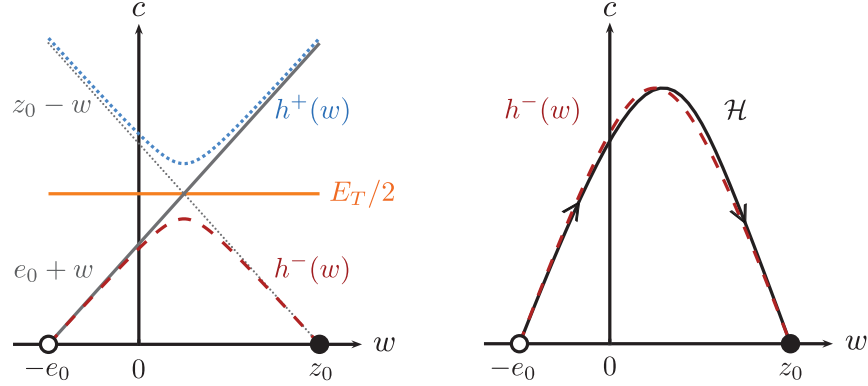

Figure 1: **The phase plane geometry associated with the IAZA reaction mechanism (1) contains two equilibrium points and a heteroclinic orbit, $\mathcal{H}$.** Left: The dashed red curve is the c-nullcline given by $c = h^-(w)$; the dotted blue curve corresponds to $c = h^+(w)$. The horizontal orange line is the constant $c = E_T/2$. The gray diagonal dotted line is $c = z_0 - w$; the gray diagonal solid line is $c = w + e_0$. The c-nullcline coincides with the w-axis at two equilibrium points. The black circle is the stable equilibrium point $x^{(1)}$, located at $(z_0, 0)$ and the saddle equilibrium $x^{(2)}$ is located at $(-e_0, 0)$ and marked by a white circle. Right: The dashed red curve is the c-nullcline $h^-(w)$. The black circle corresponds to the stable fixed point $x^{(1)}$ located at $(z_0, 0)$; the white circle corresponds to the saddle point $x^{(2)}$ located at $(-e_0, 0)$. Although the saddle point is unphysical, the two equilibria are connected via a heteroclinic orbit $\mathcal{H}$, which is indicated by the thin black curve. Note that $\mathcal{H}$ lies just below c=$h^-(w)$ when to the left of the apex, and just above $c = h^-(w)$ to the right of the apex.

When typical initial conditions for experiments (4) are imposed, the nullcline given by $c = h^-(w)$ provides an upper bound on the complex concentration, c. There are two cases we need to work out. The first case we need to consider is $e_0 \leq z_0$. In this situation, the complex concentration has a supremum given by the apex of the curve $c = h^-(w)$. Differentiating this expression with respect to w yields a critical point at $w = (z_0 - e_0)/2 \equiv w^*$, and thus

$$\sup c \equiv c^* = h^-(w^*). \tag{20}$$

In contrast, if $z_0 < e_0$, then the critical point at w^* is negative and therefore nonphysical. Thus, when experimental initial conditions are prescribed and $z_0 < e_0$, the supremum of c is given by $h^-(0)$, since $c = h^-(w)$ is a monotonically decreasing function on the interval $0 \leq w \leq z_0$ (see Figure 2). From this point forward, we will simply refer to the supremum of c as λ_Z, and take this to be $h^-(w^*)$ if $e_0 \leq z_0$, or $h^-(0)$ if $z_0 < e_0$.

There is additional structure on top of the fixed points and nullclines. Central to our analysis will be the notion of a *forward invariant set*, Λ. Formally, we call a set Λ forward invariant[3] if

$$\varphi^t : \Lambda \mapsto \Lambda \ \ \forall t \in \mathbb{R}^+. \tag{21}$$

Forward invariant sets are common in mass action models of chemical reactions, as they arise quite naturally from the conservation of mass. Note carefully that this is different than a trapping region since we do not require φ^t to map boundary points of Λ to the interior of Λ. Note also that this set is not *invariant*, because we do not require trajectories to be remain on or inside of Λ as $t \to -\infty$.

---

[3]Sometimes these sets are called *positively invariant*, which is meant to stress the fact that they are invariant with respect to positive time.
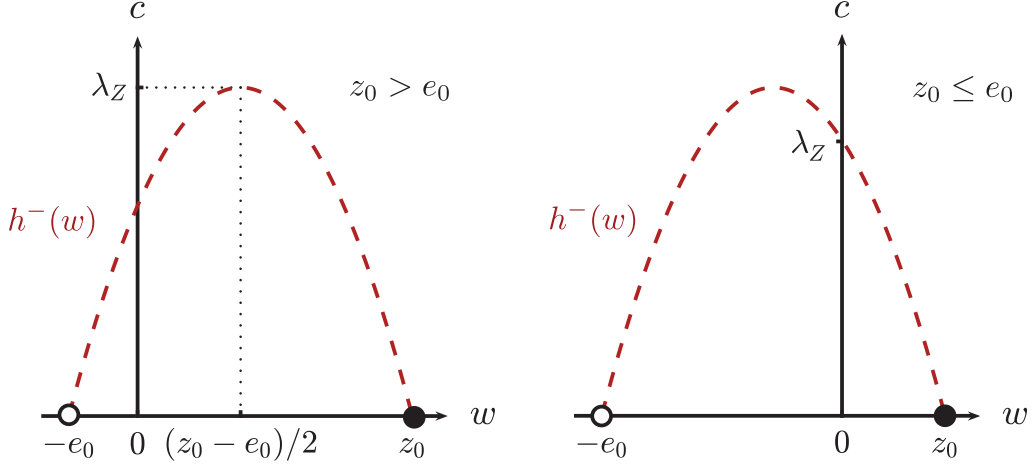
Figure 2: **The supremum of c depends on the initial ratio of enzyme to zymogen concentration in the IAZA reaction mechanism (1).** In both panels, the dashed, red curve is the c-nullcline given by $c = h^-(w)$. LEFT: This is an illustration of the phase-plane when $e_0 < z_0$. In this case, the apex of the nullcline given by $c = h^-(w)$ occurs at $w^* = (z_0 - e_0)/2$, which is positive. Consequently, the complex concentration is bounded by $c \leq \lambda_Z = h^-(w^*)$. RIGHT: In this illustration, the initial enzyme concentration is greater than the initial zymogen concentration, i.e., $z_0 < e_0$. The apex of the c-nullcline now lies in quadrant II, and corresponds to a negative and unphysical w^*. Consequently, c is bounded above by $h^-(0)$ when $z_0 < e_0$.

To construct a forward invariant set, note that due to conservation, any trajectory that starts inside or on the region "Λ" bounded by the curves

$$c = E_T/2, \tag{22a}$$

$$c = e_0 + w, \tag{22b}$$

$$c = z_0 - w, \tag{22c}$$

$$c = 0, \tag{22d}$$

$$w = 0, \tag{22e}$$

must stay on or in Λ for all positive time. In fact, we can be even more restrictive, and define Λ^* to be Λ with (22a) replaced by $c = \lambda_Z$ (see FIGURE 3 for an illustration of Λ^*).

So, why are we interested in forward invariant sets? For two reasons: First, our initial condition belongs to Λ^*, so our solution trajectory will lie inside or on the boundary of Λ^*. Thus, we only *need* to consider points that belong to Λ^* in order to describe a physical reaction that satisfies the conservation laws. Second, the set Λ^* is bounded. Since the components of our vector field are polynomial functions, they will have well-defined minimum and maximum values on Λ^*. This simplifies our analysis greatly. For example, the maximum value of c is λ_Z if we only consider initial conditions that belong to Λ^*, and this implies

$$0 < \frac{dw}{dt} \leq k_2 \lambda_Z. \tag{23}$$

In a certain sense, forward invariant sets are homologous to the state space of a stochastic process in that they provide a set of *possible* states of the system at a given point in time. Unless we start outside of Λ^*, there is no need to consider points outside of Λ^*, since they are unreachable. The only caveat with forward invariant sets is that they are not unique, and there may be situations when it is advantageous to define a less-restrictive forward invariant set, and we will say more about this in later on.

This concludes our section of phase-plane geometry. Our strategy moving forward will be to work out conditions that support the validity of the reduced models that are valid over long timescales. As we shall
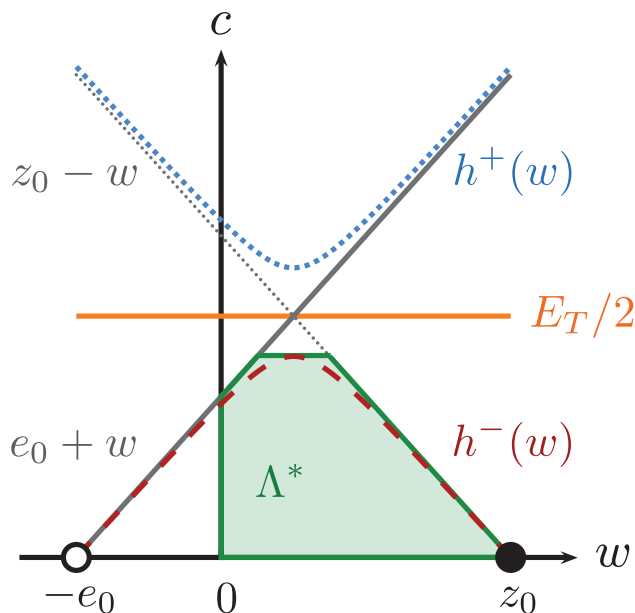
Figure 3: **Due to the presence of conservation laws and attractors, the phase-plane contains forward invariant sets in the IAZA reaction mechanism (1)**. The set Λ^* is filled with green shade and outlined by the green lines. Any trajectory that starts on or in the Λ^* stays on or in it for all $t > 0$. Obviously, there are an infinite number of forward invariant sets, but we will work primarily with Λ^* in the forthcoming analysis.

see, the conditions that permit the usage of reduced models will vary with the choice of kinetic parameters. Our aim will be to derive set of dimensionless parameters that collectively support the validity of reduced models by utilizing Fenichel theory and phase-plane geometry. To set the stage, we will review the essential basics of Fenichel theory in the next section 4.

4 Fenichel theory: Normally hyperbolic and compact manifolds with *possible* boundary

In this section we review some basic results of geometric singular perturbation theory (GSPT). We also introduce the reader to Tikhonov-Fenichel parameters (TFPs), the theory of which has been developed extensively by Goeke et al. [15]. Both GSPT and TFPs are tantamount towards the development of a geometric methodology and, even if the reader is already familiar with GSPT, we encourage them read the subsection covering TFPs, as they were developed in the context of model reduction in chemical kinetics.

4.1 Geometric singular perturbation theory

As mentioned in Section 3, we can utilize Fenichel theory to derive approximate rate equations for the IAZA reaction mechanism (2). Fenichel theory examines the persistence of normally hyperbolic, invariant manifolds. While normally hyperbolic invariant manifolds are quite general, they arise naturally in physical systems the exhibit disparate fast and slow timescales. Let x belong to $\mathbb{R}^n$, y belong to $\mathbb{R}^m$, $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, and $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$. Model equations (of the ordinary differential equation variety) that possess multiple

timescales can often be written in the form

$$\dot{x} = \varepsilon f(x, y; \varepsilon), \tag{24a}$$

$$\dot{y} = g(x, y; \varepsilon), \tag{24b}$$

where $0 < \varepsilon \ll 1$. Setting ε to zero in (24a) implies x is a constant, and the set defined by

$$\mathcal{M}_0 = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : g(x, y; 0) = 0\} \tag{25}$$

is filled with non-isolated fixed equilibrium points. We will refer to this set as the *critical set*. For simplicity, let us assume that $m = n = 1$, and that there exists a function[4] $h : \mathbb{R} \to \mathbb{R}$ such that $g(x, h(x)) = 0$. If $h(x)$ is continuous on $\mathbb{R}$, then the set $(x, h(x))$ defines a connected one-dimensional manifold embedded in $\mathbb{R}^2$. We will refer to this manifold as the *critical manifold*. The critical manifold is clearly invariant, since it is comprised of fixed points. Moreover, whenever

$$\mathrm{Re}\left(\left.\frac{\partial g(x, y)}{\partial y}\right|_{y=h_0(x)}\right) \neq 0, \tag{26}$$

the manifold defined by $y = h_0(x)$ is said to be *normally hyperbolic*; this means that the commanding component of the linearized flow is along directions that are normal to the critical manifold, $\mathcal{M}_0$. This is rather intuitive in the special case of a manifold that consists of fixed points, since the flow in the tangential direction is identically zero.

The question we need to answer is: what happens when the vector field is perturbed so that ε is not longer zero but $0 < \varepsilon \ll 1$? A now-famous theorem from Fenichel [10] provides the answer:[5]

Theorem 1 (Fenichel's First Theorem). *If $\varepsilon > 0$, but sufficiently small, there exists a manifold $\mathcal{M}_\varepsilon$ that lies within $\mathcal{O}(\varepsilon)$ of $\mathcal{M}_0$ and is diffeomorphic to $\mathcal{M}_0$. Moreover, it is locally invariant under the flow of φ^t, including in ε and C^r for any $r < +\infty$.*

There are two apparent caveats in Theorem 1. First, $\mathcal{M}_0$ must be compact, but can possibly have a boundary. Even if it constitutes a manifold, the critical set, $\mathcal{M}_0$, is generally not compact. The terminology can be tricky, because normally one thinks of a compact manifold as automatically having a boundary, since the term "closed manifold" refers to the special case of a compact manifold without boundary. In fact, if our system is planar, then every connected, one-dimensional compact manifold embedded in $\mathbb{R}^2$ is homeomorphic to a circle, since manifolds by definition are *without boundary*. Luckily, Fenichel's First Theorem allows for manifolds *with* boundary, which means we can define a *critical manifold* to be the *critical set* defined over some compact subset, S.

The second caveat is the appearance of the term *locally invariant*. Briefly, this means that if the manifold has a boundary, then trajectories are allowed to enter and exit the manifold through the boundaries. Note this is more restrictive than *invariant* since, by definition, invariant implies that any trajectory that starts on the manifold cannot enter or exit the manifold. With that said, the First Theorem of Fenichel (Theorem 1) tells us that normally hyperbolic and invariant compact subsets of the critical manifold persist under small perturbation. Thus, if $0 < \varepsilon \ll 1$, then there will exist a slow manifold, $\mathcal{M}_\varepsilon$, and the flow on $\mathcal{M}_\varepsilon$ will be *slow* but non-trivial. From this point on, we will simply refer to the slow manifold as the slow, invariant manifold (SIM). If the real part of the eigenvalues obtained via linearization about $\mathcal{M}_0$ is less than zero, then the SIM, $\mathcal{M}_\varepsilon$, will attract nearby trajectories. In fact, nearby trajectories will contract exponentially fast towards $\mathcal{M}_\varepsilon$, and stay close to it as long as it remains attractive. Thus, on the slow time $T = \varepsilon t$, the DAE given by

$$x' = f(x, h(x)), \tag{27a}$$

$$y = h(x), \tag{27b}$$

where " ${}'$ " denotes differentiation with respect to the slow time, T, provides a very good approximation to (24a)–(24b) when $0 < \varepsilon \ll 1$. Generally speaking, we call an approximation of the form (27a)–(27b) a

---

[4]One can always show that such a function exists locally if the implicit function theorem is satisfied.
[5]This version of Fenichel's first theorem can be found in [18].

reduced model, since its dimension is less than the dimension of our original model (24a)–(24b). In chemical kinetics, the approximation (27) *is* a QSSA. Our quest is therefore to a suitable "ε" that, when small, permits a reduced model for (8a)–(8c) in the form of (27a)–(27b).

Although the more recent work of Eldering [9] has extended the theorems of Fenichel to non-compact manifolds, we will keep in line with Fenichel's theorems in this paper, and only concern ourselves with compact subsets of the critical set. This is primarily because we are only interested in physical domains and bounded trajectories. We note that it is possible to describe the behavior of trajectories at "infinity" through use of the Poincaré sphere [7, 26], but we will omit such an excursion from this work since it is not a necessary component in the derivation of approximate rate laws and reduced models.

4.2   Tikhonov-Fenichel parameters

To formulate a reduced model, we require two elements: a small parameter, ε, and an invariant, attracting manifold, $\mathcal{M}_\varepsilon$. Due to the form of the phase-plane geometry, we have a "candidate" invariant manifold, $\mathcal{H}$. However, we know nothing about its structure in terms of a suitable function $c = h(w)$; we simply know it exists and that it links $x^{(2)}$ with $x^{(1)}$. Thus, knowledge of the existence of $\mathcal{H}$ is not very useful unless we can somehow implement it to formulate a meaningful QSSA.

For low-dimensional systems, there often is an easy way to spot critical manifolds by determining appropriate Tikhonov-Fenichel parameters (TPFs). These parameters were conceptualized by Goeke et al. [14, 15]. For planar systems, a TPF is any physical parameter whose vanishing results in the formation of a normally hyperbolic manifold of equilibrium points. For example, if we set $k_1 = 0$ in (8a)–(8c), then the vector field contains a manifold of equilibria given by $c = 0$, which we will denote as $\mathcal{M}_{k_1}^0$. Moreover, linearization about $\mathcal{M}_{k_1}^0$ reveals the manifold is normally hyperbolic and attracting. Thus, if k_1 is made to be sufficiently small, then **Theorem 1** suggests that there will exist a slow, invariant manifold "$\mathcal{M}_{k_1}$" within the vector field when k_1 is small but nonzero (see FIGURE 4).
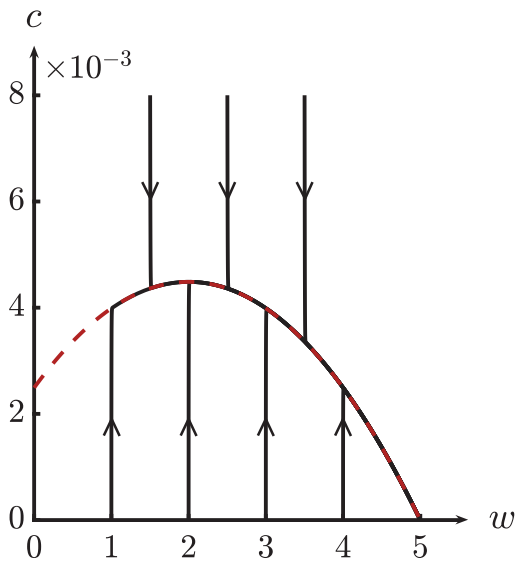

Figure 4: **The zeroing of a Tikhonov-Fenichel parameter gives rise to critical sets of equilibria.** The solid black curves are the numerical solutions to the mass action equations (8a)–(8c). The dashed curve is the c-nullcline. In all simulations $k_1 = 0.01$, $k_{-1} = k_2 = 10$, $e_0 = 1$ and $z_0 = 5$. Since k_{-1} is small, the phase-plane trajectories are attracted to a slow manifold $\mathcal{M}_{k_1}$; it appears as though the c-nullcline serves as good approximation to the slow manifold. Also note that $\max c \ll \min\{e_0, z_0, E_T/2\}$, and thus $\mathcal{M}_{k_1}$ appears to be "close" to $\mathcal{M}_{k_1}^0$ in some sense. For illustrative purposes, the units of all parameters and concentrations are arbitrary.

It is clear that k_1 is a TFP. Another TPF is k_2. Setting $k_2 = 0$ results in *two* disjoint critical sets of

equilibria: the c-nullclines $h^-(w)$ and $h^+(w)$. Thus, if k_2 is sufficiently small, we expect the existence of two slow manifolds, $\mathcal{M}^{0,-}_{k_2}$ and $\mathcal{M}^{0,+}_{k_2}$. There is however a dynamic difference between $\mathcal{M}^{0,-}_{k_2}$ and $\mathcal{M}^{0,+}_{k_2}$. First, when $k_2 = 0$, the nullclines are given by the curves

$$h^{\pm}_{k_2}(w) \equiv c = \frac{1}{2}\left(E_T + K_S \pm \sqrt{(E_T + K_S)^2 - 4(e0 + w)(z_0 - w)} \right), \tag{28}$$

and the differential equation for the system is given by

$$\dot{c} = k_1(e_0 + w - c)(z_0 - w - c) - k_1 K_S c, \tag{29a}$$
$$\dot{w} = 0. \tag{29b}$$

Differentiating the right hand side of (29a) with respect to c yields

$$\frac{\partial}{\partial c}[k_1(e_0 + w - c)(z_0 - w - c) - k_1 K_S c] = -k_1(E_T + K_M - 2c) \equiv g_c(c). \tag{30}$$

Finally, since

$$g_c(h^-_{k_2}(w)) < 0 \quad \text{and} \quad g_c(h^+_{k_2}(w)) > 0, \tag{31}$$

is follows that $\mathcal{M}^{0,-}_{k_2}$ is *attractive*, while $\mathcal{M}^{0,+}_{k_2}$ is *repulsive*. Thus, perturbing the vector field in a way that k_2 is non-zero but small, should give rise to a slow, invariant manifold $\mathcal{M}^-_{k_2}$ that attracts nearby trajectories.

The final parameter, k_{-1}, is the only rate constant that is not a TFP. The reasoning is a bit subtle, since the vanishing of k_{-1} does give rise to an invariant manifold: $c = z_0 - w$. However, the manifold may not attract nearby trajectories at an exponential rate (if at all), and thus its utility in the formulation of a reduced model may be minimal (see FIGURE 5).
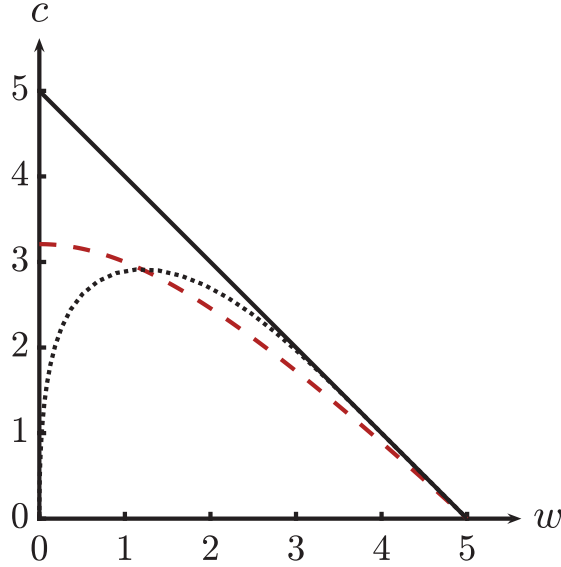

Figure 5: **The invariant manifold $c = z_0 - w$ obtained by setting $k_{-1} = 0$ may not attract nearby trajectories. Hence, k_{-1} is not a TFP.** In this figure, $k_1 = k_2 = 10$, $z_0 = e_0 = 5$, and $k_{-1} = 0$. The very thick black curve is the numerically approximated solution to (8a)–(8c) with $c(0) = z_0$ and $w(0) = 0$. The thin black curve is the numerical solution to (8a)–(8c) with $c(0) = w(0) = 0$. The dashed red curve is the c nullcline given by $c = h^-(w)$. For illustrative purposes, the units of all parameters and concentrations are arbitrary. Clearly, trajectories do not *necessarily* approach $c = z_0 - w$ when $k_{-1} = 0$ in a rapid fashion.

We have now established that attracting and invariant slow manifolds will be present in the phase-plane when either k_2 or k_1 is very small. There may be other parameters (or combinations of parameters) whose

vanishing implies the existence of a normally hyperbolic invariant manifold. However, the question we must address is: how small must TPFs be in comparison to other parameters in order to validate the use of a reduced model? Obviously, when a parameter is non-zero, the term "small" takes on a relative meaning. We must also determine the nature of the long-time dynamics when more than one TPF is small. For example, suppose $k_2 \sim k_1 \sim 10^{-6}$, but all other parameters are bounded away from 0. What reduced model is appropriate in this parameter regime? This answer to this question is not entirely obvious. Thus, it is necessary to determine a *dimensionless* parameter (i.e., a suitable ε) that, when small, ensures the existence of an attractive slow manifold when the TFPs are non-zero. The establishment of such a parameter will also allow us to quantify the asymptotic error of a specific reduced model, and the derivation of this dimensionless parameter is the subject of next section.

5 Singular perturbation parameters

Under the assumption that

$$c \approx h^-(w) \tag{32}$$

holds for the effective duration of the reaction, our objective is then to establish a criterion for validity of (32). Intuitively, (32) holds as long as the difference between c and $h^-(w)$ remains small and bounded as the reaction proceeds. This can be expressed in terms of a limit supremum:

$$\limsup_{t \to \infty} [c - h^-(w)]^2 \le \varepsilon_L^2 \lambda_Z^2. \tag{33}$$

The relationship between e_0 and z_0 alters the value of λ_Z according to the phase-plane geometry described in Section 3. As a reminder, λ_Z is given by:

$$\sup c \equiv \lambda_Z = \begin{cases} h^-((z_0 - e_0)/2) & \text{if} \quad e_0 < z_0, \\ h^-(0) & \text{if} \quad z_0 \le e_0. \end{cases} \tag{34}$$

Let us next define

$$\mu \equiv \sqrt{(K_M + E_T)^2 - 4e_0 z_0} \quad \text{and} \quad \gamma \equiv \sqrt{K_M(K_M + 2E_T)}, \tag{35}$$

so λ_Z can be written as

$$\lambda_Z = \begin{cases} (K_M + E_T - \gamma)/2 & \text{if} \quad e_0 < z_0, \\ (K_M + E_T - \mu)/2 & \text{if} \quad z_0 \le e_0. \end{cases}$$

Next, we will denote $[c - h^-(w)] \equiv \mathcal{E}_Z$. The derivative of $\mathcal{E}_Z^2$ is:

$$\frac{1}{2}\frac{d\mathcal{E}_Z^2}{dt} = \left[\frac{dc}{dt} - \frac{dh^-}{dw}\frac{dw}{dt} \right][c - h^-(w)] \tag{36a}$$

$$= \left[k_1[c - h^+(w)][c - h^-(w)] - \frac{dh^-}{dw}\frac{dw}{dt} \right][c - h^-(w)]. \tag{36b}$$

Phase-plane geometry dictates that $c - h^+(w) < 0$ for any trajectory starting in or on Λ^*. It follows from (36) that

$$\frac{1}{2}\frac{d\mathcal{E}_Z^2}{dt} \le k_1 \max[c - h^+(w)]\mathcal{E}_Z^2 + \max\left|\frac{dh^-}{dw}\right| \sup\left|\frac{dw}{dt}\right| |\mathcal{E}_Z|. \tag{37}$$

The term "$\max[c - h^+(w)]$" is easily bounded above by[6]

$$\lambda_Z - \min_{0 \le w \le z_0} h^+(w), \tag{38}$$

---

[6]Recall that $[c - h^+(w)] < 0 \ \forall x \in \Lambda^*$. Thus, the "maximum" of this quantity is the value of $[c - h^+(w)]$ that is negative, but close to zero.

where $\min h^+(w)$ is given by:

$$\min_{0 \le w \le z_0} h^+(w) = \begin{cases} h^+((z_0 - e_0)/2) & \text{if } e_0 < z_0, \\ h^+(0) & \text{if } z_0 \le e_0. \end{cases} \tag{39}$$

In FIGURE 6, we show an illustration of dynamical behaviour of (39). Using definition (39), we calculate:

$$\max[c - h^+(w)] = \begin{cases} -\sqrt{K_M(K_M + 2E_T)} = -\gamma & \text{if } e_0 \le z_0, \\[2ex] -\sqrt{(E_T + K_M)^2 - 4e_0 z_0} = -\mu & \text{if } z_0 < e_0. \end{cases} \tag{40}$$
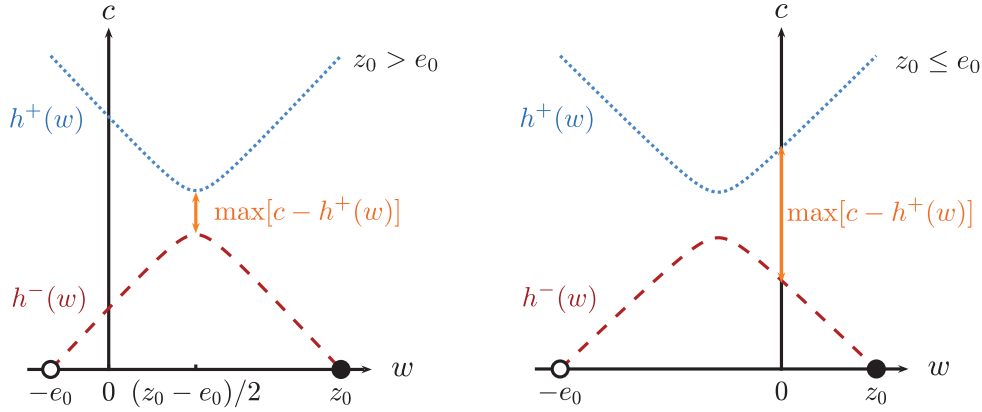

Figure 6: **We can compute an upper bound of the term $[\mathbf{c} - \mathbf{h}^+(\mathbf{w})]$ by exploiting phase-plane geometry**. Due to phase-plane geometry, $[c - h^+(w)] \le 0$ if the initial condition lies on or in Λ^*. LEFT: This is an illustration of $\max[c - h^+(w)]$ when $e_0 < z_0$. In this case, the maximum of the function occurs at the "throat" of the conic section and is given by $[c^* - h^+(w^*)]$. RIGHT: This is an illustration of $\max[c - h^+(w)]$ when $z_0 < e_0$. In this scenario, the maximum is simply $h^-(0) - h^+(0) \equiv \lambda_Z - h^+(0)$.

Our aim will now be to manipulate the differential inequality (37) so that the differential form of Grönwall's lemma can be applied. The last term on the right hand side of (37) is easily bounded in terms of $\mathcal{E}_Z^2$,

$$\max\left|\frac{dh}{dw}\right| \sup\left|\frac{dw}{dt}\right| |\mathcal{E}_Z| \le \delta \mathcal{E}_Z^2 + \frac{\left(\max\left|\frac{dh^-}{dw}\right|\right)^2 \left(\max\left|\frac{dw}{dt}\right|\right)^2}{4\delta}, \tag{41}$$

where δ is any positive constant. If we choose $\delta = k_1 \gamma/2$, it holds that

$$\frac{d\mathcal{E}_Z^2}{dt} \le -k_1 \gamma \mathcal{E}_Z^2 + \frac{\left(\max\left|\frac{dh^-}{dw}\right|\right)^2 \left(\max\left|\frac{dw}{dt}\right|\right)^2}{k_1 \gamma}. \tag{42}$$

Grönwall's lemma implies

$$\mathcal{E}_Z^2 \le \mathcal{E}(0)_Z^2 e^{-k_1 \gamma t} + \frac{\left(\max\left|\frac{dh^-}{dw}\right|\right)^2 \left(\max\left|\frac{dw}{dt}\right|\right)^2}{(k_1 \gamma)^2}\left(1 - e^{-\gamma t}\right), \tag{43a}$$

$$\le \mathcal{E}(0)_Z^2 e^{-k_1 \gamma t} + \frac{\left(\max\left|\frac{dh^-}{dw}\right|\right)^2 \left(\max\left|\frac{dw}{dt}\right|\right)^2}{(k_1 \gamma)^2}, \tag{43b}$$

14

and it follows directly from (43) that

$$\limsup_{t\to\infty}\mathcal{E}_Z^2 \equiv \limsup_{t\to\infty}(c-h^-(w))^2 \le \frac{\left(\max\left|\dfrac{\mathrm{d}h^-}{\mathrm{d}w}\right|\right)^2\left(\max\left|\dfrac{\mathrm{d}w}{\mathrm{d}t}\right|\right)^2}{(k_1\gamma)^2}. \tag{44}$$

The constants that bound the $\limsup$ are easily computed over Λ^* using standard calculus techniques:

$$\max\left|\frac{\mathrm{d}h^-}{\mathrm{d}w}\right| = \frac{E_T}{K_M + E_T}, \qquad \max\left|\frac{\mathrm{d}w}{\mathrm{d}t}\right| = k_2\lambda_Z. \tag{45}$$

Hence, we see that

$$\limsup_{t\to\infty}(c-h^-(w))^2 \le \varepsilon_L^2\lambda_Z^2, \tag{46}$$

where ε_L is given by

$$\varepsilon_L \equiv \left(\frac{E_T}{K_M+E_T}\right)\left(\frac{K}{K_S+K}\right)\left(\frac{K_M}{\sqrt{K_M(K_M+2E_T)}}\right). \tag{47}$$

The dimensionless parameter ε_L is essential, because it ensures the validity of the QSSA. To see this, carefully observe that the bound (43) is satisfied *pointwise* in time. At any time point, t, it holds that

$$\mathcal{E}_Z^2(t) \le \mathcal{E}^2(0)e^{-k_1\gamma t} + \varepsilon_L^2\lambda_Z^2, \tag{48}$$

which means that the *number* on the left hand side of the inequality is less than or equal to the *number* on the right of the inequality. Taking the principal square root of both sides yields

$$\sqrt{\mathcal{E}_Z^2(t)} \le \sqrt{\mathcal{E}^2(0)e^{-k_1\gamma t} + \varepsilon_L^2\lambda_Z^2}, \tag{49}$$

which holds since the principal "$\sqrt{\cdot}$" is a monotonically increasing function. Next, since $\sqrt{x+y} \le \sqrt{x} + \sqrt{y}$ for all $0 \le x, y$, we have

$$\sqrt{\mathcal{E}_Z^2} \le \sqrt{\mathcal{E}^2(0)e^{-k_1\gamma t} + \varepsilon_L^2\lambda_Z^2} \le \sqrt{\mathcal{E}^2(0)e^{-k_1\gamma t}} + \sqrt{\varepsilon_L^2\lambda_Z^2}. \tag{50}$$

Finally, it follows from (50) that

$$|c-h^-(w)|(t) \le |(c-h^-(w))(0)|e^{-k_1\gamma t/2} + \varepsilon_L\lambda_Z, \tag{51}$$

and thus the long-time error in the QSSA is ensured by demanding $\varepsilon_L \ll 1$, since λ_Z is bounded.

With three different forms of QSSAs, namely sQSSA, rQSSA and tQSSA, the question follows is: which QSSA is validated by $\varepsilon_L \ll 1$? To answer this, note first that ε_L factors into three small parameters:

$$\varepsilon_Z \equiv \frac{E_T}{K_M+E_T}, \qquad \nu_Z \equiv \frac{K}{K_S+K}, \qquad \epsilon_Z \equiv \frac{K_M}{\sqrt{K_M(K_M+2E_T)}}. \tag{52}$$

The vanishing of each small parameter in (52) gives rise a manifold of equilibrium points. Specifically, if $\varepsilon_Z = 0$, then the w-axis (i.e., the curve $c = 0$) is filled with equilibrium points. This suggests that ε_Z regulates the validity of the sQSSA. In contrast, if $\epsilon_Z \ll 1$, then $\varepsilon_Z \sim 1$, and this implies that the parameter ϵ_Z controls the validity of the rQSSA. Thus, we anticipate that ε_Z, ν_Z, and ϵ_Z control the validity of the sQSSA, tQSSA and rQSSA for the mass action equations (15a)–(15b) of the IAZA reaction mechanism.

For the case when $z_0 \le e_0$, we express the upper bound on $|c-h^-(w)|$ in equivalent form but with modified parameter values. Since (40) indicates $\max[c-h^+(w)] \le -\mu$ (again, see FIGURE 6, RIGHT Panel), then using μ instead of γ implies that the upper bound for $|c-h^-(w)|$ is

$$|c-h^-(w)| \le |(c-h^-(w))(0)|e^{-k_1\mu t/2} + \hat{\varepsilon}_L\lambda_Z. \tag{53}$$

The parameter $\hat{\varepsilon}_L = \varepsilon_Z \nu_Z \upsilon_Z$, and υ_Z, which is the only term that differs from the terms in ε_L, is given by

$$\upsilon_Z \equiv \frac{K_M}{\sqrt{(E_T + K_M)^2 - 4e_0 z_0}} = \frac{K_M}{E_T + K_M - 2\lambda_Z}. \tag{54}$$

What (54) reveals is that the conditions for the validity of sQSSA and tQSSA do not change regardless of whether $e_0 \geq z_0$ or $e_0 < z_0$, because ε_Z and ν_Z remain the same. The validity of rQSSA *is* sensitive to the choice of initial conditions, since ϵ_Z is valid when $e_0 \geq z_0$ and υ_Z is valid when $e_0 < z_0$. This sensitivity is due to the presence of a bifurcation in the singular limit. We will investigate it thoroughly in Section 7.

In conclusion of this section, we have determined that if "experimental" initial conditions are chosen (i.e., $c(0) = w(0) = 0$), then[7]

$$|c - h^-(w)| \leq h^-(0)e^{-k_1 \gamma t/2} + \varepsilon_L \lambda_Z, \qquad \text{if } e_0 \leq z_0, \tag{55a}$$

$$|c - h^-(w)| \leq h^-(0)e^{-k_1 \mu t/2} + \hat{\varepsilon}_L \lambda_Z, \qquad \text{if } z_0 \leq e_0. \tag{55b}$$

Consequently, we see that $c \to h^-(w)$ as $\varepsilon_L \to 0$ and $t \to \infty$. For the unconvinced reader, we invite them to read [8], where we utilize a slightly different approach that involves Lyapunov functions. Although it is unnecessary, it remains to be shown that the exponential terms in (55) decay *quickly* with respect to a slow timescale. To do this, we need to express (55) in terms of a suitable dimensionless slow timescale, T_{slow}, so that exponential terms in (55) assume the form

$$e^{-C_1 T_{slow}/\varepsilon_L}, \quad e^{-C_2 T_{slow}/\hat{\varepsilon}_L}, \tag{56}$$

where "C" denotes a constant that is greater than zero and $\mathcal{O}(1)$. If we can rewrite the exponential terms in (55) in form of (56), then the exponential terms should be $\mathcal{O}(\varepsilon_L)$ and $\mathcal{O}(\hat{\varepsilon}_L)$ once $T_{slow} \sim \varepsilon_L |\ln \varepsilon_L|$ and $\sim \hat{\varepsilon}_L |\ln \hat{\varepsilon}_L|$, respectively. The determination of an appropriate slow timescale is the subject of next section.

6 Fast and slow timescales

Here we estimate the dimensional fast and slow timescales that delimit the transient and QSS regimes of the IAZA reaction mechanism (1). Then, we will proceed to express the bounds (55a)–(55b) in terms of the *dimensionless* fast and slow time scales.

6.1 Case I: $z_0 \leq e_0$

We first consider the case when $z_0 \leq e_0$, and determine appropriate fast and slow timescales. To do this, it is helpful to presuppose fast and slow phase-plane dynamics. Let us consider the fast timescale. Under the a priori assumption that there is a negligible formation of w during the initial transient phase of the reaction, it is reasonable to assume that

$$\frac{\mathrm{d}c}{\mathrm{d}t} \approx k_1 c^2 - k_1(E_T + K_M)c + k_1 z_0 e_0. \tag{57}$$

The above expression (57) is a Riccati equation. Its solution admits a natural fast timescale

$$t_C \equiv \frac{1}{k_1 \sqrt{(E_T + K_M)^2 - 4e_0 z_0}} = -1/k_1 \mu. \tag{58}$$

Now that we have a suitable fast timescale, our goal is to rewrite (55a)–(55b) in terms of the dimensionless time, $\tau \equiv t/t_C$. First, recall that in the regime where $z_0 \leq e_0$, the phase-plane trajectory approaches the curve $c = h^-(w)$ and, once the trajectory intercepts $c - h^-(w)$, the complex concentration decreases monotonically as $w \to z_0$. Since the maximum of $[c - h^+(w)]$ is bounded by $h^-(0) - h^+(0)$ if $(w, c) = (0, 0)$ and $e_0 \geq z_0$, it follows from (55b) that:

$$|c - h^-(w)| \leq h^-(0)e^{-\tau/2} + \hat{\varepsilon}_L \lambda_Z. \tag{59}$$

---

[7]Recall that $h^-(0) \equiv \lambda_Z$ when $z_0 \leq e_0$.

Because (34) defines $\lambda_Z = h^-(0)$ for $e_0 \geq z_0$, dividing (59) by λ_Z yields

$$\frac{1}{\lambda_Z}|c - h^-(w)| \leq e^{-\tau/2} + \hat{\varepsilon}_L, \tag{60}$$

and we obtain an inequality that is entirely expressed in terms of dimensionless variables. It is only natural to scale c and $h^-(w)$ by λ_Z, and the natural scaling of w is $\bar{w} = w/z_0$. Thus, defining $\hat{c} \equiv c/\lambda_Z$, and $\underline{h}^-(\bar{w}) \equiv h^-(\bar{w}z_0)/\lambda_Z$, we can write

$$|\hat{c} - \underline{h}^-(\bar{w})|(\tau) \leq e^{-\tau/2} + \hat{\varepsilon}_L, \tag{61}$$

which provides an upper bound on the corresponding dimensionless system.

We now move on to the slow timescale. When $z_0 \leq e_0$, the slow timescale can be approximated by the ratio of total change in concentration with the maximal speed of the reaction [35]. Thus, we need to compute:

$$\frac{|\Delta w|}{\max |\dot{w}|} = \frac{z_0}{k_2 \lambda_Z} = t_W. \tag{62}$$

The dimensionless slow timescale, T, is defined as: $T \equiv t/t_W$. Rewriting (60) in terms of T yields

$$\frac{1}{\lambda_Z}|c - h^-(w)| \leq e^{-T/2\varepsilon_T} + \hat{\varepsilon}_L, \tag{63}$$

where ε_T is

$$\varepsilon_T = \frac{t_C}{t_W}. \tag{64}$$

The inequality (63) reveals that the exponential term decays rapidly with respect to the slow timescale and, it is relatively straightforward to show that $\varepsilon_T \leq \hat{\varepsilon}_L$. Consequently, $\hat{\varepsilon}_L \to 0$ induces separation in the fast and slow timescales. This suggests that timescale separation (i.e., $\varepsilon_T \ll 1$) is a necessary, but *possibly* insufficient condition for the validity of the QSSA. Finally, since $\varepsilon_T \leq \hat{\varepsilon}_L$, it holds that

$$\frac{1}{\lambda_Z}|c - h^-(w)| \leq e^{-T/2\hat{\varepsilon}_L} + \hat{\varepsilon}_L, \tag{65}$$

with respect to the dimensionless time, T, and thus the phase-plane trajectory is $\mathcal{O}(\hat{\varepsilon}_L)$ from the c-nullcline once $T \sim \hat{\varepsilon}_L |\ln \hat{\varepsilon}_L|$.

6.2   Case II: $e_0 < z_0$

Timescale estimation is part art, part science, and part luck. When $e_0 < z_0$, the fast timescale, t_C, remains equivalent to (58), but calculating an appropriate slow timescale becomes challenging due to the complicated phase–plane geometry described in Section. 3. Therefore, we will not attempt to directly compute a universal slow timescale that accurately measures the duration of the reaction when $e_0 \leq z_0$. Instead, we will define a generic slow timescale. To do this, we know that $w \sim w^*$ once c reaches it maximal value,[8] the time for the reaction to effectively complete at this point is given by Segel's estimate

$$t_{\text{descn.}} \equiv \frac{\Delta z}{k_2 \lambda_Z} = \frac{E_T}{2k_2 \lambda_Z}, \tag{66}$$

since $\Delta z = z_0 - w^* = E_T/2$. The subscript "descn." in (66) is short for descend, since this is roughly the time it takes before the reaction is completed, once c reaches its maximum near λ_Z. The timescale that accounts for the actual duration of the reaction will be $n\, t_{\text{descn.}}$, where n is some number that is greater than or equal to one. The exact value of n is redundant for the generic slow timescale derivation, yet the condition $n \geq 1$ is an important consequence of the phase–plane geometry: when $e_0 < z_0$, it takes an additional amount of time for c to accumulate to its maximum, since c must accumulate in a QSS before it reaches its threshold

---

[8]Recall that c reaches its maximum value when the trajectory intercepts the c-nullcline, $c = h^-(w)$, near its apex.

value (i.e., the phase-plane trajectory must crawl upwards and below the c-nullcline for some time before it can cross it). Thus, the *total* time spent in QSS will be the ascension time, $t_{ascen.}$, plus the descending time, $t_{descn.}$:

$$t_{slow} = t_{ascen.} + t_{descn.} = nt_{descn.}, \quad n = 1 + t_{ascen.}/t_{descn.}. \tag{67}$$

Defining a dimensionless slow timescale, $\bar{T} = t/t_{descn.}$, it is easy to check that

$$|c - h^-(w)| \leq h^-(0)e^{-\kappa\bar{T}/2\varepsilon_L} + \varepsilon_L\lambda_Z, \quad 1 \leq \kappa \equiv 1 + \sqrt{\frac{K_M(K_M + 2E_T)}{(E_T + K_M)^2}}. \tag{68}$$

Thus, in terms of dimensionless variables, we have that

$$\frac{1}{\lambda_Z}|c - h^-(w)| \leq \delta e^{-\kappa\bar{T}/2\varepsilon_L} + \varepsilon_L, \quad \delta \equiv h^-(0)/\lambda_Z \leq 1, \tag{69}$$

and we have convergence with respect to the dimensionless slow timescale, $\bar{T}$, as $\varepsilon_L \to 0$.

7 The quasi-steady-state approximations

In this section, we analyze the sQSSA, tQSSA, and rQSSA associated with the IAZA reaction mechanism (1). When possible, closed-form solutions are obtained from zeroth-order approximations. Our primary aim will be to derive unique small parameters that ensure the validity of either the sQSSA, rQSSA, and tQSSA. Our secondary aim will be to determine an appropriate reduced model that is valid on the slow timescale.

7.1 The sQSSA

The sQSSA corresponds to minimal formation of complex, and the associated critical manifold is $c = 0$. Thus, over the slow timescale, we anticipate

$$w \sim z_0 - z. \tag{70}$$

Our objective will be to determine a dimensionless parameter that ensures the approximation (70) is valid. Because the sQSSA corresponds to negligible formation of complex, the results from previous section imply that $E_T/K_M \ll 1$ places a strict limit the maximal accumulation of c. As a consequence, $\varepsilon_Z \ll 1$ is a natural candidate for the validity of sQSSA.

To verify that the sQSSA is valid when $E_T/K_M \ll 1$, we can show that the mass action equations for c and z constitute a singularly perturbed set of equations when the system is written in terms of appropriate dimensionless variables. To start, the parameter $E_T/K_M \ll 1$ suggests that the slow timescale is given by $\tilde{t} \equiv 1/k_1 E_T$. Introducing the dimensionless variables $\hat{c} \equiv c/\lambda_Z$, $\bar{z} \equiv z/z_0$, $\bar{e} \equiv e/E_T$ and $\tilde{T} \equiv t/\tilde{t}$ yields

$$\alpha\frac{d\bar{z}}{d\tilde{T}} = -\alpha\bar{e}\bar{z} + (1 - \nu_Z)\hat{c} \tag{71a}$$

$$\varepsilon_s\frac{d\hat{c}}{d\tilde{T}} = \alpha\bar{e}\bar{z} - \hat{c}, \tag{71b}$$

where the parameters α and ε_s are given by:

$$\varepsilon_s \equiv E_T/K_M, \quad \alpha \equiv \frac{z_0}{\lambda_Z}\frac{E_T}{K_M} = \frac{z_0}{\lambda_Z}\varepsilon_s. \tag{72}$$

Although α contains the term ε_s, it remains bounded away from 0 as $\varepsilon_s \to 0$, because $z_0/\lambda_Z \to \infty$. In fact, a simple calculation reveals $\alpha \sim \mathcal{O}(1)$ as $\varepsilon_s \to 0$, and the equations comprise a singularly perturbed set of equations, with the singular perturbation parameter defined as ε_s. Note that we have scaled *a posteriori*, and that scaling has been used simply as a means of reaffirming the singular perturbation parameter that regulates the validity of the sQSSA.

Now that we have established that the sQSSA is valid whenever $\varepsilon_s \ll 1$, we need to determine an appropriate reduced model. Since c evolves in a QSS when $\varepsilon_s \ll 1$, the approximation

$$\frac{\mathrm{d}z}{\mathrm{d}t} \sim -\frac{k_2(E_T - z)z}{K_M + 2z} \sim -\frac{k_2}{K_M}(E_T - z)z, \tag{73}$$

holds on over the slow timescale whenever $\varepsilon_Z \ll 1$. Integration of (73) yields

$$z(t) \sim \frac{z_0 E_T}{e_0 e^{k_2 \varepsilon_s t} + z_0}. \tag{74}$$

Furthermore, it follows from (70) that

$$w(t) \sim z_0 \left(1 - \frac{E_T}{e_0 e^{k_2 \varepsilon_s t} + z_0} \right). \tag{75}$$

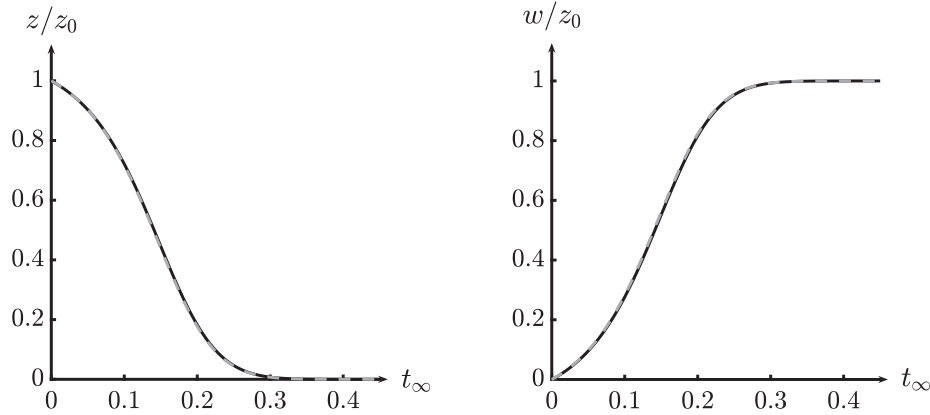The sQSSA approximations, (74) and (75), are easily verified numerically (FIGURE 7).

Figure 7: **The sQSSA, (74) and (75), for the IAZA reaction mechanism (1) is valid when $\mathbf{E_T} \ll \mathbf{K_M}$.** In both panels, the solid black curve is the numerical solution to the mass action equations (8a)–(8c), and the dashed black curve (barely visible) is the numerical solution to (74). The rate constants used to construct the numerical approximation are: $e_0 = 9$, $z0 = 1$, $k_1 = 2$, $k_{-1} = k_2 = 500$. LEFT: In this panel, the solid black curve is the numerical solution to $z(t)$, and the dashed black curve is the solution generated by (74). RIGHT: In this panel, the solid black curve is the numerical solution to $w(t)$, and the dashed black curve is the solution generated by (75). For illustrative purposes, the units of all parameters and concentrations are arbitrary. Time has been mapped to the t_∞ scale: $t_\infty(t) = 1 - 1/\ln[t + \exp(1)]$.

When the sQSSA is valid, the fast timescale is roughly $t_C \approx 1/(k_{-1} + k_2)$, and from the reduced model (74), one can estimate a timescale that provides a rough estimate of duration of the reaction when the sQSSA is valid. The time it takes for z to deplete to a concentration of $X z_0$, where $0 < X \leq 1$, is given by

$$t \sim \frac{1}{k_2 \varepsilon_s} \ln \left[\frac{1}{X} + \left(\frac{1-X}{X} \right) \frac{z_0}{e_0} \right]. \tag{76}$$

Borrowing the familiar definition from linear equations, the natural choice for X is $X = 1/\exp(1)$, where $\exp(1)$ is Napier's constant: $\exp(1) \approx 2.718$. An more accurate estimate would choose $X = 1/5 \exp(1)$ (see FIGURE 8).
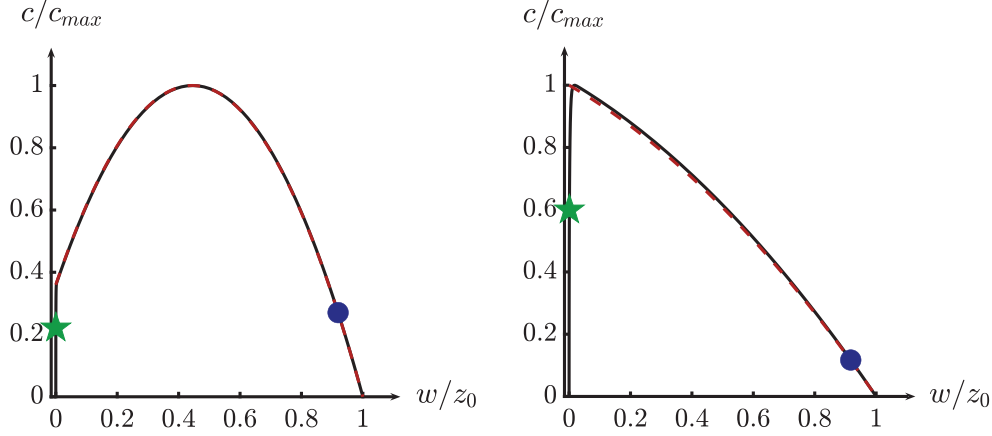
Figure 8: **The fast and slow timescales associated with the validity of the sQSSA in the IAZA reaction mechanism (1) delimit the duration of the transient as QSS regimes when $\mathbf{E_T/K_M} \ll 1$.** In both panels, the solid black curve is the numerical solution to the mass action equations (8a)–(8c), and the dashed red curve (barely visible) is the c-nullcline given by $c = h^-(w)$. The rate constants used to construct the numerical approximation are: $k_1 = 2$, $k_{-1} = k_2 = 500$. The green star demarcates the time point $t = t_C$, and the blue circle demarcates the time point $t = t_s$, where t_s is given by (76) with $X = 1/5 \exp(1)$. LEFT: In this panel, $z_0 = 1$ and $e_0 = 9$, and thus the maximum concentration of c occurs during the QSS phase of the reaction. RIGHT: In this panel, $e_0 = 7$, $z_0 = 3$, and the maximum value of c occurs at the onset of the QSS regime. For illustrative purposes, the units of all parameters and concentrations are arbitrary.

7.2 The tQSSA

Akin to the tQSSA in the single-enzyme, single-substrate, MM reaction mechanism (13), the tQSSA is essentially an umbrella term that implies there is negligible formation of product during the transient phase of the reaction. In other words, when the tQSSA is valid, the approximation $p \approx 0$ holds for $t \lesssim t_C$. The same is true for the IAZA reaction mechanism.

The small parameter responsible for the validity of the tQSSA is ν_Z. In fact, as we established in our discussion on TFPs, setting $k_2 = 0$ results in the formation of two invariant critical manifolds, $c = h^-(w; K_S)$ and $c = h^+(w; K_S)$. The dependency on the term "K_S" signifies that if $k_2 = 0$, then K_M is replaced with $K_S = k_{-1}/k_1$ in $h^\pm(w)$. The corresponding tQSSA is

$$c \sim h^-(w; K_S), \tag{77a}$$

$$\dot{w} \sim k_2 h^-(w; K_S). \tag{77b}$$

Thus, we see that c evolves in a QSS for the majority of the reaction when $k_2 \ll k_{-1}$.

Expression (77b) is an autonomous differential equation that approximates the formation of w on the slow timescale. To obtain a rate equation for z, it helps to change coordinate systems and analyze the mass action equations in the (z, c) phase-plane. The mass action equations for z and c are

$$\frac{\mathrm{d}z}{\mathrm{d}t} = -k_1(E_T - z)z + k_1(K_S + 2z)c, \tag{78a}$$

$$\frac{\mathrm{d}c}{\mathrm{d}t} = k_1(E_T - z)z - k_1(K_M + 2z)c, \tag{78b}$$

and setting $k_2 = 0$ generates a manifold of equilibria given by

$$\mathcal{M}_0^\nu = \left\{ (z, c) \in \mathbb{R}_+^2 : c = \frac{(E_T - z)z}{K_S + 2z} \right\}. \tag{79}$$

20

Linearizing about the critical manifold (79) yields

$$\begin{bmatrix} \dot{z} \\ \dot{c} \end{bmatrix} = \begin{bmatrix} -E_T + 2(z + c) & K_S + 2z \\ E_T - 2(z + c) & -K_S - 2z \end{bmatrix} \Bigg|_{(z,c) \in \mathcal{M}_0^\nu} \begin{bmatrix} z \\ c \end{bmatrix}, \tag{80}$$

and the eigenvector that is tangent to $\mathcal{M}_0^\nu$ is given by

$$\boldsymbol{v} = \begin{bmatrix} 1 \\ \dfrac{(E_T - 2z)K_S - 2z^2}{(K_S + 2z)^2} \end{bmatrix}. \tag{81}$$

Thus, we anticipate

$$\frac{\mathrm{d}c}{\mathrm{d}t} \sim \frac{(E_T - 2z)K_S - 2z^2}{(K_S + 2z)^2} \left(\frac{\mathrm{d}z}{\mathrm{d}t} \right) \tag{82}$$

during the QSS phase of the reaction. Inserting (82) into $\dot{c} + \dot{z} + \dot{w} = 0$ and solving for $\dot{z}$ yields

$$\frac{\mathrm{d}z}{\mathrm{d}t} \sim \frac{-\dot{w}}{1 + \dfrac{(E_T - 2z)K_S - 2z^2}{(K_S + 2z)^2}}. \tag{83}$$

Using (79), we can replace $\dot{w}$ in (83) with

$$\frac{\mathrm{d}w}{\mathrm{d}t} \sim \frac{k_2(E_T - z)z}{K_S + 2z}, \tag{84}$$

which generates an autonomous differential equation for z:

$$\frac{\mathrm{d}z}{\mathrm{d}t} \sim -\frac{k_2 z(E_T - z)(K_S + 2z)}{K_S^2 + (E_T + 2z)K_S + 2z^2}. \tag{85}$$

To supply (85) with boundary condition, we note that $z = z_0 - c - w$. Since "w" is the slow variable when the tQSSA is valid, it holds that $c \sim h^-(0)$ at the onset of the QSS period. It follows from this observation that $z^* = z_0 - h^-(w)$ is the appropriate boundary condition for (85) (see FIGURE 9).

The tQSSA for z, (85), was derived using heuristic "linear" methods. However, similar "linear" methods have been used with a large degree of success in singularly perturbed problems [25, 16], but we note that (85) can be rigorously derived as a first-order approximation to the QSS dynamics by approximating (asymptotically) the solution to the invariance equation. You can read [8, 47] for an example of how to use the invariance equation to generate higher order approximations of the slow manifold when the MM model equations are equipped with two fast variables, s and c, and one slow variable, p.

7.3 The rQSSA

The final reduced model we will consider is the rQSSA. Based on our preliminary analysis, we expect the validity of the rQSSA to hold as $\epsilon_Z \to 0$. This can be accomplished in two ways: we can consider the limit as $E_T \to \infty$, or the limit as $K_M \to 0$. In the singular limit, the vector field contains two submanifolds that intersect and exchange stability at the point $(w, c) = [(z_0 - e_0)/2, E_T/2] \equiv (w_T, c_T)$. Equivalently, the nullclines $c = h^\pm(w)$ "touch" when $K_M = 0$ (see FIGURE 10). The stability of equilibria that comprise the curves $c = e_0 + w$ and $c = z_0 - w$ in the singular limit is straightforward to compute:

$$\frac{\partial \dot{c}}{\partial c} \Bigg|_{c=z_0-w} = -k_1(e_0 - z_0 + 2w), \tag{86a}$$

$$\frac{\partial \dot{c}}{\partial c} \Bigg|_{c=e_0+w} = k_1(e_0 - z_0 + 2w). \tag{86b}$$
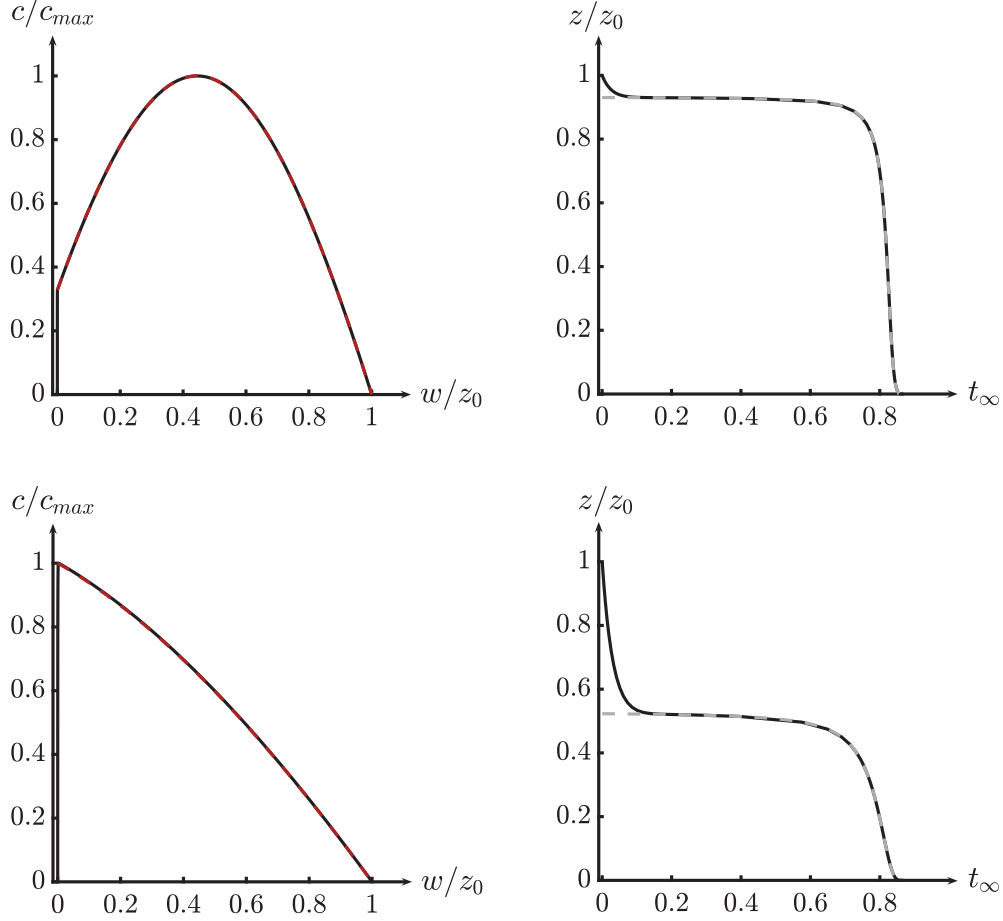
Figure 9: **The tQSSA is valid for the IAZA reaction mechanism (1) when $k_2 \ll k_{-1}$.** In all panels, the solid black curve is the numerical solutions to the mass action equations (8a)–(8c). In the upper and lower right panels, the dashed/dotted curve is the numerical solution to (85) with z^* as the initial condition. In the top panels: $k_1 = 1.0$, $k_{-1} = 5$, $k_2 = 0.01$ and $e_0 = 1$, $z_0 = 9$. In the bottom panels: $k_1 = 1.0$, $k_{-1} = 5$, $k_2 = 0.01$ and $e_0 = 7$, $z_0 = 3$. For illustrative purposes, the units of all parameters and concentrations are arbitrary. In the upper and lower right panels, time has been mapped to the t_∞-scale: $t_\infty(t) = 1 - 1/\ln[t + \exp(1)]$. TOP LEFT: This panel illustrates a typical trajectory in the (w, c) phase-plane when the tQSSA is valid. The dashed red curve is the c-nullcline, and by inspection it is easy to see that (77b) is justified. TOP RIGHT: In this panel, the numerical solutions to z as well as (85) are compared on the t_∞ scale. Clearly, (85) serves as a good approximation to $z(t)$ for $z \leq z^* \approx 8.374$. The lower left and lower right panels give analogous illustrations of the (w, c) phase-plane and temporal dynamics when $e_0 \geq z_0$.

Thus, there is a loss of normal hyperbolicity at the point (w_T, c_T). This particular point is called a *singular point*, and is suggestive of a dynamic transcritical bifurcation. For now, we will assume a generic transcritical bifurcation is present in the singular limit, but we will verify this at the end of the section.

To construct a reduced model that is valid when $0 < \epsilon_Z \ll 1$, we must once again work out the individual cases: $e_0 < z_0$ and $z_0 \leq e_0$. Let us start by assuming $e_0 < z_0$. After the transient, the trajectory closely follows the curve $c = w + e_0$ towards the transcritical point, and then follows the curve $c = z_0 - w$ towards the stable equilibrium point at $(w, c) = (z_0, 0)$. During the ascent of c along the branch $c \sim e_0 + w$, it holds that

$$\frac{\mathrm{d}w}{\mathrm{d}t} \sim k_2(w + e_0), \quad w(0) = 0, \tag{87}$$
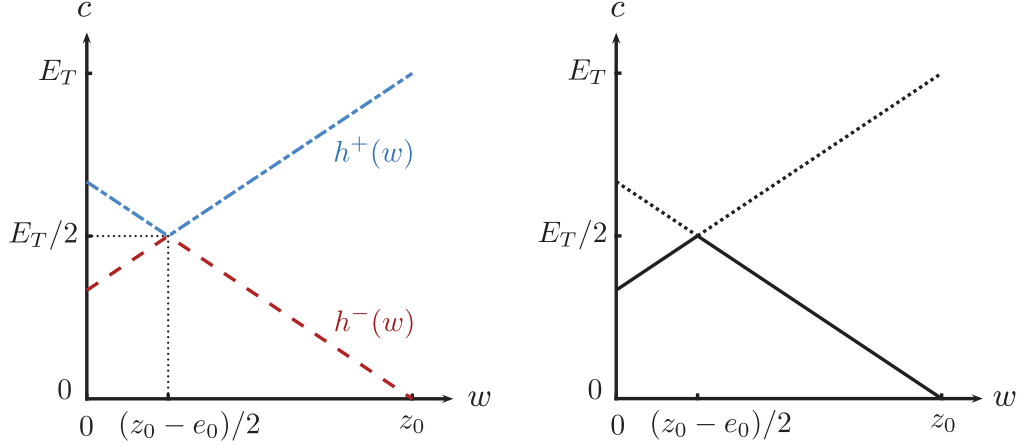
22

Figure 10: **In the singular limit that occurs when $K_M = 0$, a dynamic transcritical bifurcation emerges in the vector field**. LEFT: When $K_M = 0$, the nullclines $c = h^-(w)$ and $c = h^+(w)$ coalesce with the curves $c = z_0 - w$ and $c = w + e_0$. Each curve, $c = z_0 - w$ and $c = w + e_0$, is a critical submanifold of fixed points. These submanifolds intersect and exchange stability at the transcritical bifurcation point $(w, c) = ((z_0 - e_0)/2, E_T/2) \equiv (w_T, c_T)$. RIGHT: Dashed lines correspond to unstable equilibrium points, and solid lines correspond to stable equilibrium points.

and thus

$$w(t) \sim e_0 e^{k_2 t} - e_0. \tag{88}$$

A straightforward calculation reveals the time it takes c to reach its maximum is roughly $t \sim \ln(E_T/2e_0)/k_2 \equiv T^*$. Next, as c descends towards zero along the curve $c = z_0 - w$, we have

$$\frac{\mathrm{d}w}{\mathrm{d}t} \sim k_2(z_0 - w), \quad w(0) = \frac{1}{2}(z_0 - e_0). \tag{89}$$

Thus, we have a piecewise solution

$$w(t) \sim \begin{cases} e_0\left(e^{k_2 t} - 1\right), & t \leq \dfrac{1}{k_2}\ln(E_T/2e_0) \\[2ex] \dfrac{1}{2}(z_0 - e_0)e^{-k_2 t} + z_0\left(1 - e^{-k_2 t}\right), & t \geq \dfrac{1}{k_2}\ln(E_T/2e_0) \end{cases} \tag{90}$$

that is valid during the QSS period of the reaction. It is straightforward to use the Heavyside function, H, to "stitch together" the solutions (90) into:

$$w(t) \sim H(T^* - t)\left[e_0\left(e^{k_2 t} - 1\right)\right] + H(t - T^*)\left[\frac{1}{2}(z_0 - e_0)e^{-k_2(t - T^*)} + z_0\left(1 - e^{-k_2(t - T^*)}\right)\right]. \tag{91}$$

On the other hand, if $z_0 \leq e_0$, then there is no ascending period, and we obtain a smooth approximation

$$w(t) \sim z_0\left(1 - e^{-k_2 t}\right), \tag{92}$$

that is analogous to the rQSSA for product formation in the MM reaction mechanism. We illustrate the validity of the approximations (91) and (92) in FIGURE 11.
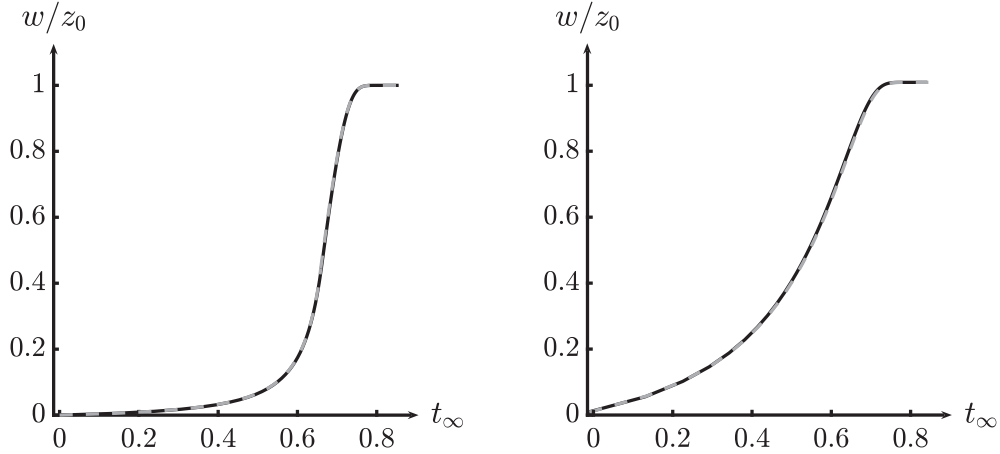
Figure 11: **For the IAZA reaction mechanism (1), the mathematical rQSSA of $w(t)$ differs on the validity of the condition $z_0 \leq e_0$.** LEFT: The solid black curve is the numerical solution to the mass action equations (8a)–(8c) with $k_1 = 10$, $k_2 = 0.1$, $k_{-1} = 0.1$, $e_0 = 2$ and $z_0 = 8$. The dotted curve, which is barely visible, is the rQSSA given by (91). RIGHT: The solid black curve is the numerical solution to the mass action equations (8a)–(8c) with $k_1 = 10$, $k_2 = 0.1$, $k_{-1} = 0.1$, $e_0 = 7$ and $z_0 = 3$. The dotted curve, which is again barely visible, is the rQSSA given by (92). For illustrative purposes, the units of all parameters and concentrations are arbitrary. In both panels, time has been mapped to the t_∞-scale: $t_\infty(t) = 1 - 1/\ln[t + \exp(1)]$.

7.3.1 The transcritical bifurcation: codimension and structural stability

Let "p_T" denote a transcritical bifurcation point for a general singularly perturbed system of the form:

$$\dot{x} = \varepsilon f(x, y; \varepsilon), \tag{93a}$$

$$\dot{y} = g(x, y; \varepsilon). \tag{93b}$$

The conditions

$$g(p_T; 0) = 0, \quad g_y(p_T; 0) = 0, \quad g_x(p_T; 0) = 0, \quad g_{yy}(p_T; 0) \neq 0, \quad \det H_g(p_T; 0) < 0, \quad f(p_T; 0) \neq 0, \tag{94}$$

where "H_g" denotes the Hessian of $g(x, y)$, must hold in order to ensure the bifurcation is non-degenerate [21, 22]. To simplify the calculations, we want to obtain equations (15a)–(15b) into the form (93a)–(93b). This will require scaling the equations. From our analysis of the rQSSA, it seems only natural to define the slow timescale as $1/k_2$, and from this we can define a dimensionless slow time, $\bar{T} = k_2 t$. In addition, the concentrations w and c scale naturally with respect to z_0 and $E_T/2$, respectively. Thus, let us define: $\bar{w} = w/z_0$, $\bar{c} = 2c/E_T$. Expressing (15a)–(15b) in terms of dimensionless variables yields:

$$\varepsilon^* \nu_Z \frac{d\bar{c}}{d\bar{T}} = 2\left(1 + (\bar{w} - 1)\eta - \frac{1}{2}\bar{c}\right)\left(\eta(1 - \bar{w}) - \frac{1}{2}\bar{c}\right) - \varepsilon^* \bar{c}, \qquad \eta \equiv \frac{z_0}{E_T}, \tag{95a}$$

$$\frac{d\bar{w}}{d\bar{T}} = \frac{1}{2}(1 + \sigma)\bar{c}, \qquad \sigma \equiv e_0/z_0. \tag{95b}$$

From (95), we observe the presence of two singular perturbation parameters: $\varepsilon^* \equiv E_T/K_M$, and ν_Z. Luckily, we can change ε^* independently of ν_Z by varying k_1. In other words, if we want ε^* to vanish while holding ν_Z constant, we can fix k_{-1} and k_2 and then take $k_1 \to \infty$. The same structure is present in the mass action equations that model (13) [see 8, for details].

Independence aside, it will still be nice to understand *why* the term ν_Z appears as a singular perturbation parameter. The reason is due to the fact that the term "$k_{-1}c$" that appears in (15a) is an *imperfect* term.

24

To see why, consider the normal form for a typical *static* transcritical bifurcation,

$$\dot{x} = px - x^2, \tag{96}$$

where "p" is a parameter. While the normal form (96) is expressed in terms of only one parameter, p, the bifurcation is *not* of codimension one. The result is that the vector field "$px - x^2$" is *structurally* unstable, meaning its topology can be destroyed with the addition of an arbitrarily small perturbation. This is easy to see, since adding a term like "$0 < \varepsilon \ll 1$" to (96) would remove the bifurcation entirely.

Armed with the insight that that term $k_{-1}c$ is an imperfection, let us revisit the scaled equations (95a)–(95b), and re-write them as

$$\varepsilon_B \frac{d\bar{c}}{d\bar{T}} = 2\left(1 + (\bar{w} - 1)\eta - \frac{1}{2}\bar{c}\right)\left(\eta(1 - \bar{w}) - \frac{1}{2}\bar{c}\right) - \left(\varepsilon_B + \varepsilon_I\right)\bar{c}, \tag{97a}$$

$$\frac{d\bar{w}}{d\bar{T}} = \frac{1}{2}(1 + \sigma)\bar{c}, \tag{97b}$$

where the terms ε_B and ε_I are given by:

$$\varepsilon_B \equiv \frac{k_2}{k_1 E_T}, \qquad \varepsilon_I \equiv \frac{k_{-1}}{k_1 E_T}. \tag{98}$$

It is clear that the transcritical bifurcation occurs at the point in parameter space where both ε_B and ε_I vanish. Combining the terms in a sum implies $K_M/E_T \equiv \varepsilon^* = 0$ for the vector field to contain the transcritical bifurcation. For this reason, we designate ε^* as the parameter that ensures the validity of the rQSSA, even though the imperfection is "absorbed" into the parameter.

With the understanding of the bifurcation parameter, as well as the imperfection, we are finally in a position to verify the non-degeneracy conditions. Let us define $g(\bar{w}, \bar{c}; \varepsilon_B)$ and $f(\bar{w}, \bar{c})$ as

$$g(\bar{w}, \bar{c}; \varepsilon_B) \equiv 2\left(1 + (\bar{w} - 1)\eta - \frac{1}{2}\bar{c}\right)\left(\eta(1 - \bar{w}) - \frac{1}{2}\bar{c}\right) - \varepsilon_B \bar{c}, \tag{99a}$$

$$f(\bar{w}, \bar{c}) \equiv \frac{1}{2}(1 + \sigma)\bar{c}, \tag{99b}$$

where we have set the imperfect term to zero. The transcritical bifurcation point is given by

$$p_T \equiv (\bar{w}, \bar{c}) = (1 - 1/2\eta, 1) \equiv (\bar{w}_T, \bar{c}_T). \tag{100}$$

It is straight forward to calculate:

$$f(p_T; 0) = \frac{1}{2}(1 + \sigma) \neq 0, \tag{101a}$$

$$g(p_T; 0) = 0, \tag{101b}$$

$$g_{\bar{w}} = 4(1 - \bar{w})\eta^2 - 2\eta, \quad g(\bar{w}_T; 0) = 0, \tag{101c}$$

$$g_{\bar{c}} = (\bar{c} - 1) - \varepsilon_B, \quad g(\bar{c}_T; 0) = 0, \tag{101d}$$

$$g_{\bar{c}\bar{c}} = 1 \neq 0, \tag{101e}$$

$$g_{\bar{w}\bar{w}} = -4\eta^2, \tag{101f}$$

$$g_{\bar{c}\bar{w}} = g_{\bar{w}\bar{c}} = 0. \tag{101g}$$

Finally, the Hessian of $g(\bar{w}, \bar{c}; \varepsilon_B)$ is formulated from (101e)–(101g)

$$H_g = \begin{bmatrix} g_{\bar{c}\bar{c}} & g_{\bar{c}\bar{w}} \\ g_{\bar{w}\bar{c}} & g_{\bar{w}\bar{w}} \end{bmatrix}, \tag{102}$$

and it is clear that $\det H_g = -4\eta^2 < 0$. Thus, in the singular limit, a generic transcritical bifurcation occurs at the point $(w, c) = ((z_0 - e_0)/2, E_T/2)$.

7.3.2 Finding French ducks where you least expect them: Following the repulsive SIM along a canard

As we mentioned in the last subsection 7.3.1, the term k_{-1} should really be viewed as an imperfection. Perturbing the vector field with $k_{-1} = 0 < k_2 \ll k_1$ so that $k_2/k_1 \ll E_T$ results in the formation of an interesting solution: $c = z_0 - w$. The unique feature of this particular solution is rather subtle, and we must remind ourselves that Fenichel's theorem applies to both attracting and repulsing critical manifolds. In the singular limit, we have attracting critical submanifolds, as well as repulsive critical submanifolds. The attracting critical submanifolds correspond to the lower branches of the curves $c = z_0 - w$ and $c = e_0 + w$, while the repulsive critical submanifolds correspond to the upper branches of these curves. Fenichel theory implies that once the vector field is perturbed, there will exist both attractive and repulsive SIMs. When the imperfect term k_{-1} is zero, perturbation of the vector field results in the formation an invariant manifold: $c = z_0 - w$. A canard solution is a solution that closely follows the repulsive branch of the SIM for some time, before passing near the bifurcation point and proceeding to follow the attracting SIM. This is very counter-intuitive, since repulsive SIMs should repel nearby trajectories, but this happens because the canard is comprised of the intersection of the repulsive and attractive SIMs (see FIGURE 12). The existence of the
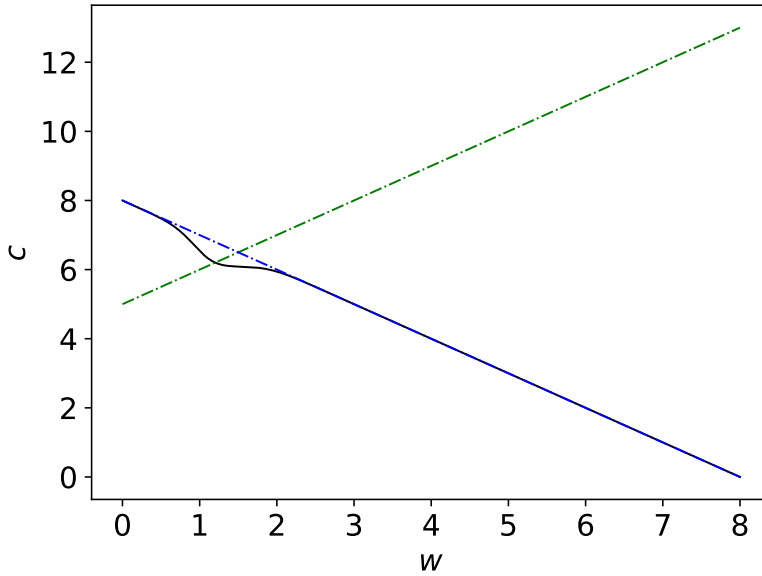

Figure 12: **Canard solutions $c = z_0 - w$ appears when $k_{-1} = 0 < k_2 \ll k_1$ in the IAZA reaction mechanism (1).** The canard solution is given by the thin black curve, and the numerically integrated solution of the mass action equations (3a)–(3d). The dashed/dotted blue line is the invariant manifold $c = z_0 - w$, and the dashed/dotted green line is the curve $c = e_0 + w$. The canard solution follows the unstable branch of the invariant manifold for some time before passing closely near the bifurcation point (the intersection of the dashed/dotted green and blue lines) and proceeding to follow the stable branch towards the equilibrium point. Parameter values used in the simulation are as follows: $k_{-1} = 0, k_2 = 3, k1 = 100, z_0 = 8, e_0 = 5$. The initial condition used to construct the canard trajectory is: $(z, c, w)(0) = (0, 7.9999, 0)$. All units are arbitrary.

canard can be proven using normal form theory, and we refer the reader to the Appendix for a detailed calculation of the normal form coefficient. We also note that canards are present in vector field of the mass action equations that model the MM reaction mechanism (13). We refer the reader to [8] for a detailed overview of the phase-plane geometry.

8 Comparison between the IZACA and MM reactions mechanisms

In this section we compare the results of our analysis with those of the MM reaction mechanism (13). In order to draw a useful comparison, we will define a metric that measures the applicability of the general QSSA to the IAZA reaction mechanism (1) in the subsection that follows.

8.1 The IAZA reaction mechanism: long-time bounds

Here we will compare the results of our analysis with those of the MM reaction mechanism (13). A good indicator of the propensity for the QSSA is the magnitude of

$$\left(\frac{E_T}{K_M + E_T}\right)\left(\frac{k_2}{k_{-1} + k_2}\right)\left(\frac{K_M}{\sqrt{K_M(K_M + 2E_T)}}\right) \equiv \varepsilon_L. \tag{103}$$

Since ε_L is the small parameter associated with the validity of the QSSA, it is useful to bound it as this gives us an indication of how "poor" the approximation can be. This is straightforward to do using elementary calculus. First, let us take the term $\nu_Z \leq 1$. Thus we can bound (103)

$$\left(\frac{E_T}{K_M + E_T}\right)\left(\frac{K_M}{\sqrt{K_M(K_M + 2E_T)}}\right) \equiv L(K_M, E_T), \tag{104}$$

which is a multivariate function of *two* variables. The critical points are found by setting the components of the gradient to zero:

$$\partial_{K_M} L = -\frac{E_T(K_M^2 + E_T K_M - E_T^2)}{(E_T + K_M)^2(K_M + 2E_T)\sqrt{K_M(K_M + 2E_T)}} = 0, \tag{105a}$$

$$\partial_{E_T} L = \frac{K_M(K_M^2 + E_T K_M - E_T^2)}{(E_T + K_M)^2(K_M + 2E_T)\sqrt{K_M(K_M + 2E_T)}} = 0. \tag{105b}$$

Setting the term "$K_M^2 + E_T K_M - E_T^2$" in the numerator to zero and solving for K_M yields

$$\max_{K_M, E_T \in \mathbb{R}_0^+} L(K_M, E_T) < 1/3, \tag{106}$$

and thus we have that ε_L is bounded *strictly* by $1/3$.

Before we move on and discuss the classic MM reaction mechanism (13), we should note that (103) was obtained with respect to initial conditions that belong to Λ^*. However, given our set of conservation laws, an initial condition such as $[w(0), c(0)] = (0, E_T/2)$ is certainly allowable. We simply chose Λ^* because we were interested the standard "experimental initial condition" of $(w(0), c(0)) = (w, 0)$, and we proceeded to compute the lim sup accordingly. But nothing prevents us from considering other conditions. If we consider a condition such as $[w(0), c(0)] = (0, E_T/2)$, then it helps to redraw the forward invariant set so that it includes the point $(z, w, c) = (0, w^*, E_T/2)$. An obvious choice is Δ, which we illustrate in FIGURE 13. Recalculating the limit supremum using Δ, it holds that

$$0 \leq \frac{dw}{dt} \leq k_2 \frac{E_T}{2} \qquad \forall (w, c) \in \Delta. \tag{107}$$

The derivative term, $\max|dh^-(w)/dw|$, remains the same:

$$\max\left|\frac{dh^-(w)}{dw}\right| \leq \frac{E_T}{K_M + E_T}. \tag{108}$$

The remaining term, $\max(c - h^+(w))$, can be recalculated as follows:

$$\max k_1(c - h^+(w)) = k_1\left(\frac{E_T}{2} - h^+(w^*)\right) = -\frac{k_1}{2}\left(K_M + \sqrt{K_M(K_M + 2E_T)}\right). \tag{109}$$
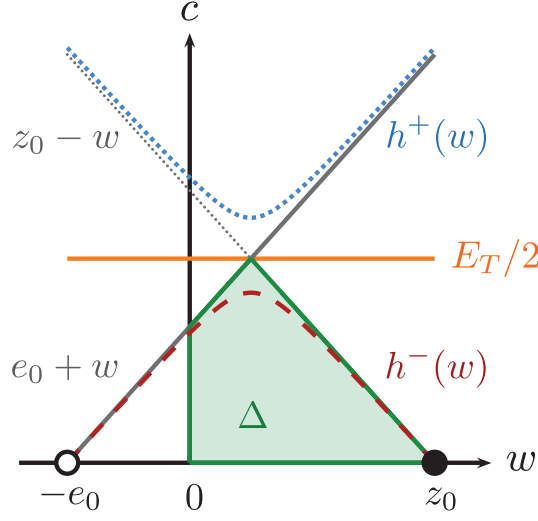
Figure 13: **We can redraw the forward invariant set to include "atypical" initial conditions for the IAZA reaction mechanism (1).** The set Δ is outlined by the thick black lines. Any trajectory that starts on or in the Δ stays on or in it for all $t > 0$. Note this set contains more initial conditions than Λ^*. Please compare with FIGURE 3.

Piecing together (107)–(109), we have:

$$\limsup_{t \to \infty}(c - h^-(w))^2 \leq \frac{E_T^2}{4}\left(\frac{E_T}{K_M + E_T}\right)^2\left(\frac{k_2}{k_2 + k_{-1}}\right)^2\left(\frac{2K_M}{K_M + \sqrt{K_M(K_M + 2E_T)}}\right)^2. \tag{110}$$

For any bounded sequence, x_n with $n \in \mathbb{N}$, it holds that $\limsup \sqrt{x_n} = \sqrt{\limsup x_n}$. Therefore, we have that

$$\limsup_{t \to \infty}|c - h^-(w)| \leq \frac{E_T}{2}\left(\frac{E_T}{K_M + E_T}\right)\left(\frac{k_2}{k_2 + k_{-1}}\right)\left(\frac{2K_M}{K_M + \sqrt{K_M(K_M + 2E_T)}}\right). \tag{111}$$

To extract a dimensionless parameter we scale c with respect to its maximum value of $E_T/2$. This implies our dimensionless small parameter is:

$$\left(\frac{E_T}{K_M + E_T}\right)\left(\frac{k_2}{k_2 + k_{-1}}\right)\left(\frac{2K_M}{K_M + \sqrt{K_M(K_M + 2E_T)}}\right). \tag{112}$$

The only term in (112) that differs from (47) is the last term. To investigate changes in asymptotic behavior, we define

$$\vartheta(E_T, K_M) \equiv \frac{2K_M}{K_M + \sqrt{K_M(K_M + 2E_T)}}, \tag{113a}$$

$$\rho(E_T, K_M) \equiv \frac{K_M}{\sqrt{K_M(K_M + 2E_T)}}, \tag{113b}$$

$$\omega(E_T, K_M) \equiv \sqrt{K_M/E_T}, \tag{113c}$$

and note the following:

$$\lim_{K_M \to 0}\frac{\vartheta}{\omega} = 1, \quad \lim_{K_M \to 0}\frac{\omega}{\vartheta} = 1, \implies \vartheta \sim \mathcal{O}(\omega), \tag{114a}$$

$$\lim_{K_M \to 0}\frac{\rho}{\omega} = \frac{1}{2}, \quad \lim_{K_M \to 0}\frac{\omega}{\rho} = 2, \implies \rho \sim \mathcal{O}(\omega). \tag{114b}$$

Thus, in either case, the asymptotic convergence rate of the $\limsup$ as $K_M \to 0$ is bounded by $\mathcal{O}(\sqrt{\varepsilon^*})$, even though the small parameter established in (7.3.1) is ε^*. Theorists have shown that this is common in vector fields that contain transcritical bifurcation points and, excluding canards, it is well-established that trajectories of fast/slow systems of the form (93a)–(93b) remain at a distance of $\mathcal{O}(\sqrt{\varepsilon})$ from the transcritical point (see FIGURE 14, as well as [2]). However, this does *not* imply that the approximation $[c - h^-(w)]$ will
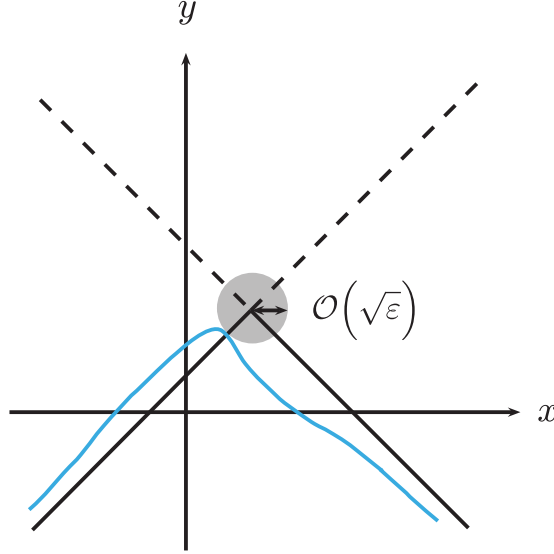

Figure 14: **Phase-plane trajectories of the IAZA reaction mechanism (1) are bounded away from the transcritical bifurcation point by a distance that is $\mathcal{O}(\sqrt{\varepsilon})$.** In this illustration, a transcritical bifurcation occurs at the the center of the gray circle. For general fast/slow systems that contain a transcritical bifurcation point, one can prove, using normal form theory, that certain trajectories (light blue curve in this figure) are bounded away from the transcritical point by a distance that is $\mathcal{O}(\sqrt{\varepsilon})$.

necessarily be $\mathcal{O}(\sqrt{\varepsilon^*})$ near the transcritical point. This is because the vector field can be perturbed in such a way that $k_2 \ll k_{-1} \ll k_1$. Therefore, the tQSSA can still be very accurate, even in the neighborhood of the transcritical bifurcation point. The problem arises when we use the approximation (92). As $z_0 \to e_0$, this approximation will lose asymptotic accuracy near the bifurcation point. This is a consequence of Fenichel theory, which breaks down at bifurcation points. While Fenichel theory indicates that the asymptotic validity is $\mathcal{O}(\varepsilon^*)$ over normally hyperbolic subsets, the same may not hold when normal hyperbolicity is lost. The transcritical bifurcation point is a particular example of this. Of course, away from the transcritical point Fenichel theory holds, and the error will be $\mathcal{O}(\varepsilon^*)$.

What about an upper bound on (112)? Once again using straightforward calculus, it is easy to show that

$$\left(\frac{E_T}{K_M + E_T} \right) \left(\frac{k_2}{k_2 + k_{-1}} \right) \left(\frac{2K_M}{K_M + \sqrt{K_M(K_M + 2E_T)}} \right) \leq \sqrt{2} - 1. \tag{115}$$

In contrast to (106), the inequality (112) is not strict, but it is larger than $1/3$. Of course, once the trajectory enters Λ^* it is subject to the same limit supremum defined by (106). The important point to note is that upper bound on the asymptotic rate of convergence remains unchanged, as both functions, ϑ and ρ, are $\mathcal{O}(\sqrt{\varepsilon^*})$.

Finally, what does happen when $z_0 \ll e_0$ and the transcritical bifurcation moves to the second quadrant? Fenichel's first theorem holds in neighborhoods that do not contain the transcritical point. Thus, we expect convergence to be of $\mathcal{O}(\varepsilon)$. To check this, we examine the $\limsup$ when $z_0 < e_0$. In the case the transcritical point is "translated out" of our forward invariant set, it follows from (40) that

$$\frac{1}{\lambda_Z} \limsup_{t \to \infty} |c - h^-(w)| \leq \left(\frac{E_T}{K_M + E_T} \right) \left(\frac{k_2}{k_2 + k_{-1}} \right) \left(\frac{K_M}{E_T + K_M - 2\lambda_Z} \right), \tag{116}$$

29

where $\lambda_Z \equiv h^-(0)$ when $z_0 < e_0$. The last term on the right hand side of (116) is $\mathcal{O}(\varepsilon^*)$ when $z_0 \ll e_0$.

8.2 Revisiting the validity of the tQSSA for MM reaction mechanism

The equations that govern the MM reaction mechanism (13) have a structure that is similar to those of the IAZA reaction mechanism (1). There are, however, qualitative and quantitative differences. Qualitatively, the IAZA reaction mechanism is an autocatalytic enzyme catalyzed reaction following a MM mechanism of enzyme action. The MM reaction mechanism is not autocatalytic. Quantitatively, the mass action equations that govern the formation of complex concentration, c_S,[9] and generation of product, p, for the MM reaction mechanism (13) are

$$\dot{c}_S = k_1(e_0 - c_S)(s_0 - c_S - p) - k_1 K_M c_S \tag{117a}$$

$$\dot{p} = k_2 c_S, \tag{117b}$$

where e_0 denotes the initial enzyme concentration, and s_0 denotes the initial substrate concentration. Equation (117a) factors as

$$\dot{c}_S = k_1(c_S - g^-(p))(c_S - g^+(p)), \tag{118}$$

and the nullclines, $c = g^\pm(p)$, are given by

$$g^\pm(p) \equiv \frac{1}{2}(e_0 + K_M + s_0 - p) \pm \frac{1}{2}\sqrt{(e_0 + K_M + s_0 - p)^2 - 4e_0(s_0 - p)}. \tag{119}$$

Under the QSSA, the assumption that

$$c_S \sim g^-(p), \tag{120a}$$

$$\dot{p} \sim k_2 g^-(p). \tag{120b}$$

When (120) is valid, the fast and slow timescales that delimit the respective time to maximal complex concentration and formation of product are

$$t_{c_S} = \frac{1}{e_0 + s_0 + K_M - 2\lambda_S}, \qquad t_P = \frac{s_0}{k_2 \lambda_S}, \qquad \lambda_S = g^-(0). \tag{121}$$

Traditionally, the "propensity" of (120) is measured by timescale separation. Tzafriri [42] was the first to obtain the upper bound

$$\frac{t_{c_S}}{t_P} \leq 1/4. \tag{122}$$

There is an obvious caveat with (122). Timescale separation is *necessary*, but possibly not sufficient for the validity of (120). The possible insufficiency of (122) is due to the fact that the timescales were derived under the assumed validity of (120). Segel and Slemrod [36] understood this, as well as Tzafriri [42], and the estimates of Tzafiri should be seen as significant "first step" towards quantifying the validity of (120). However, we are now in a position to introduce a more informed metric. In a recent work, Eilertsen and Schnell [8] found that

$$\limsup_{t \to \infty} |c_S - g^-(p)| \leq \lambda_S \left(\frac{e_0}{K_M + e_0}\right)\left(\frac{k_2}{k_2 + k_{-1}}\right)\left(\frac{2K_M}{K_M + \sqrt{K_M(K_M + 4e_0)}}\right), \quad e_0 < s_0, \tag{123a}$$

$$\limsup_{t \to \infty} |c_S - g^-(p)| \leq \lambda_S \left(\frac{e_0}{K_M + e_0}\right)\left(\frac{k_2}{k_2 + k_{-1}}\right)\left(\frac{K_M}{e_0 + K_M - \lambda_S}\right), \quad s_0 < e_0. \tag{123b}$$

Since it holds that

$$\frac{2K_M}{K_M + \sqrt{K_M(K_M + 4e_0)}} \leq \frac{K_M}{e_0 + K_M - \lambda_S}, \tag{124}$$

---

[9]The subscript "S" has been added to indicate that c_S is referring to standard complex concentration of the MM reaction mechanism.

it is easy to show that

$$\limsup_{t \to \infty} |c_S - g^-(p)| \leq \lambda_S/3. \tag{125}$$

The propensity of the IAZA reaction mechanism (1) to evolve in a QSS is roughly the same as the propensity of the MM reaction mechanism (13). In both cases the limiting factor of the approximation is due to the appearance of a transcritical bifurcation. Near this point, the QSS approximation is $\mathcal{O}\sqrt{K_M/e_0}$ for the MM reaction mechanism (13), and $\mathcal{O}\sqrt{K_M/E_T}$ for the IAZA reaction mechanism (1). However, it is important to note that the upper bounds are of order unity, which means that there are parameter ranges for which the tQSSA can and will fail. We emphasize this statement because the bound (122) is often–and incorrectly–interpreted to imply that the tQSSA for the MM reaction mechanism (13) is in some sense *always* valid.

9 Discussion

In summary, we have successfully quantified the long-time validity of the sQSSA, rQSSA, and tQSSA for the IAZA reaction mechanism (1). We have also derived approximations, which can be used to model IAZA reaction mechanism (1) or estimate the reaction parameters from time course experiments.

Our novel method of analysis has revealed a rich dynamical structure that includes a dynamic transcritical bifurcation, and imperfection, as well as a canard solution. Specifically, we found that the asymptotic validity of the sQSSA is $\mathcal{O}(E_T/K_M)$, and the asymptotic of the tQSSA is of order $\mathcal{O}(k_{-1}/k_2)$. The asymptotic validity of the rQSSA is dependent on the location of the transcritical bifurcation point. Away from the transcritical point, the asymptotic validity of the rQSSA is of order $\mathcal{O}(K_M/E_T)$. On the other hand, in neighborhoods close to the bifurcation point, the validity is bounded asymptotically by $\mathcal{O}(\sqrt{K_M/E_T})$. These estimates are the analogous the estimates computed by Eilertsen and Schnell [8] for the MM reaction mechanism. However, our estimates differ from those obtained through the scaling approach developed by Segel [35] for the sQSSA, and later by Tzafriri [42] for the tQSSA of the MM reaction mechanism (13).

At this point, it is natural to ask whether or not there is a meaningful comparison between the small parameter obtained from the ratio of timescale method developed first by Segel [35] and later by Tzafriri [42], and the small parameters obtained in this work, as well as in our previous work [8]? This answer is yes, although it is a bit more subtle than one might think. First, the fast timescale, t_C, can be written in the following form

$$t_C = \frac{1}{k_1(h^+(w) - h^-(w))}, \tag{126}$$

and thus we have that

$$\max t_C = \frac{1}{k_1 \min(h^+(w) - h^-(w))}. \tag{127}$$

Now consider the slow timescale, t_W, given again by (62)

$$t_W = \frac{z_0}{k_2 \lambda_Z}. \tag{128}$$

Minimizing t_W can be done heuristically. Since $z_0/\lambda_Z \geq 1$, the minimal of t_W is $1/k_2$. Taking the ratio of these optimal scales yields:

$$\frac{\max t_C}{\min t_W} = \left(\frac{k_2}{k_{-1} + k_2}\right)\left(\frac{K_M}{\sqrt{K_M(K_M + 2E_T)}}\right). \tag{129}$$

The last term in the small parameter, ε_Z, stems from the geometry of the c-nullcline, and regulates the validity of the sQSSA. A similar calculation can be done for the MM reaction mechanism (13), and the bounds given in [8] can be tightened.

In either the case of the IAZA or MM reaction mechanism, the conditions for the rQSSA, sQSSA and tQSSA can be recovered from timescale separation calculations. Our work illustrates that if the goal is to

extract dimensionless variables that regulate the validity of a *specific* QSSA, then it appears to be far more prudent to consider the ratio

$$\varepsilon_G \equiv \frac{\text{Maximum fast timescale}}{\text{Minimum slow timescale}}, \tag{130}$$

which gives a more global description of the the legitimacy of the QSSA away from initial conditions, and this serves as the main takeaway for applied scientists. The global description is consequence of the fact that if the ratio (130) is small, then it is small for *any* initial condition.

Ultimately, the ratio (130) is analogous to a stiffness ratio, r_S,

$$r_S = \frac{\max_i |\lambda_i|}{\min_i |\lambda_i|}, \tag{131}$$

where λ_i are the eigenvalues of a linearized operator. Côme [5] was one of the first recognize the significance of the stiffness ratio in the application of the QSSA to enzyme-catalyzed reactions, and Palsson [27] even developed a methodology for obtaining conditions that the ensure the validity of the QSSA based on the stiffness ratio. The difference between our work and that of Palsson [27] and Côme [5] is that we are essentially computing the "worst-case-scenario" stiffness ratio over the entire critical manifold as opposed to computing the stiffness ratio by linearizing about the equilibrium point of the non-singular (perturbed) problem. The motivation for this follows from Fenichel theory, in which case one has an infinite number of fixed points in the initial layer problem, and so it is more consistent to linearize the system about each fixed point that belongs to the critical manifold, compute a complete set of stiffness ratios, and take the smallest of these ratios as the singular perturbation parameter.

One of the most important contributions this manuscript lies on the nature of our scaling approach. We have outlined a procedure that allows for the asymptotic quantification of the QSSA that avoids the direct need for scaling, non-dimensionalization, or a priori timescale estimation: our method relies entirely on the exploitation of phase-plane geometry. By using a combination of Tikhonov-Fenichel parameters, Grönwall's lemma, and phase-plane geometry, we have recovered upper bounds that quantify the asymptotic validity of the various QSS approximations. Although the case study we addressed in this work is two-dimensional, our approach is powerful, because it can be easily extend to coupled reactions of both the auxiliary and sequential variety. Perhaps most importantly, the methodology we have developed can be implemented to resolve the validity of the QSSA in the reversible forms of the MM reaction mechanism (13)

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E + P, \tag{132a}$$

and the IAZA reaction mechanism (1)

$$Z + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} 2E + W. \tag{133a}$$

This is significant for two reasons. First, the above reaction mechanism are thermodynamically more realistic than their irreversible counterparts. Second, the QSSA for reversible reactions have traditionally been much more difficult to analyze and validate. However, the geometric method we have developed here, combined with the work of Goeke et al. [14, 15] easily extends to the resolution of this, and other more complex nonlinear problems.

Appendix

Here we prove the existence of the canard, $c = z_0 - w$, when $k_{-1} = 0$ and $K = k_2/k_1 \ll E_T$. The dimensionless system is given by:

$$\varepsilon_B \frac{\mathrm{d}\bar{c}}{\mathrm{d}\bar{T}} = 2\left(1 + (\bar{w} - 1)\eta - \frac{1}{2}\bar{c}\right)\left(\eta(1 - \bar{w}) - \frac{1}{2}\bar{c}\right) - \varepsilon_B \bar{c}, \tag{134a}$$

$$\frac{\mathrm{d}\bar{w}}{\mathrm{d}\bar{T}} = \frac{1}{2}(1 + \sigma)\bar{c}. \tag{134b}$$

To begin, we will make the following transformation:

$$\bar{c} \mapsto -\bar{c} \quad \& \quad \bar{T} \mapsto -\bar{T}. \tag{135}$$

Applying the transformation (135) yields:

$$\varepsilon_B \frac{\mathrm{d}\bar{c}}{\mathrm{d}\bar{T}} = g(\bar{c}, \bar{w}, \varepsilon_B) \text{ and } \frac{\mathrm{d}\bar{w}}{\mathrm{d}\bar{T}} = f(\bar{c}, \bar{w}, \varepsilon_B), \tag{136a}$$

where

$$g(\bar{c}, \bar{w}, \varepsilon_B) \equiv 2\left(1 + (\bar{w} - 1)\eta + \frac{1}{2}\bar{c}\right)\left(\eta(1 - \bar{w}) + \frac{1}{2}\bar{c}\right) + \varepsilon_B \bar{c}, \tag{137a}$$

$$f(\bar{c}, \bar{w}, \varepsilon_B) = \frac{1}{2}\eta^{-1}\bar{c}. \tag{137b}$$

Furthermore, the translated bifurcation point occurs at

$$p_T \equiv (\bar{w}, \bar{c}) = (1 - 1/2\eta, -1) \equiv (\bar{w}_T, \bar{c}_T). \tag{138}$$

A canard solution will exist if the normal form coefficient, λ_N,

$$\lambda_N = \frac{1}{|f_0|\sqrt{k_{\bar{c}\bar{w}}^2 - k_{\bar{w}\bar{w}}k_{\bar{c}\bar{c}}}}(k_{\varepsilon_B}k_{\bar{c}\bar{c}} + f_0 k_{\bar{c}\bar{w}}) \tag{139}$$

is equal to one. The terms that comprise (139) are as follows:

$$k_{\bar{c}\bar{c}} = \frac{1}{2}\frac{\partial^2}{\partial\bar{c}^2}g(\bar{w}_T, \bar{c}_T, 0), \quad k_{\bar{c}\bar{w}} = \frac{1}{2}\frac{\partial^2}{\partial\bar{c}\partial\bar{w}}g(\bar{w}_T, \bar{c}_T, 0) \tag{140a}$$

$$k_{\bar{w}\bar{w}} = \frac{1}{2}\frac{\partial^2}{\partial\bar{w}^2}g(\bar{w}_T, \bar{c}_T, 0), \quad k_{\varepsilon_B} = \frac{\partial}{\partial\varepsilon_B}g(\bar{w}_T, \bar{c}_T, 0), \tag{140b}$$

and $f_0 = f(\bar{w}_T, \bar{c}_T, 0) = \eta^{-1}/2$. The coefficients $k_{\bar{c}\bar{w}}$, $k_{\bar{c}\bar{c}}$ and k_{ε_B} are straightforward to calculate:

$$k_{\bar{c}\bar{w}} = 0, \quad k_{\bar{c}\bar{c}} = \frac{1}{2}, \quad k_{\varepsilon_B} = 1, \tag{141}$$

and the term $k_{\bar{w}\bar{w}}k_{\bar{c}\bar{c}}$ is given by

$$k_{\bar{w}\bar{w}}k_{\bar{c}\bar{c}} = -4\eta^2. \tag{142}$$

Thus, λ_N is given by:

$$\lambda_N = \frac{2}{\eta^{-1}\sqrt{4\eta^2}} = \frac{2}{2\eta^{-1}\eta} = 1, \tag{143}$$

which confirms that a canard solution is present in the vector field when $0 < \varepsilon_B \ll 1$ and $\varepsilon_I = 0$. By extension, a canard solution is also present in the vector field of the classic MM reaction mechanism (13) when $k_{-1} = 0$, given by $c = s_0 - p$ [see 8].

Acknowledgement

References

[1] J. Al-Janabi, J. A. Hartsuck, and J. Tang. Kinetics and mechanism of pepsinogen activation. *J. Biol. Chem.*, 247:4628–4632, 1972.

[2] N. Berglund and B. Gentz. *Noise-induced phenomena in slow-fast dynamical systems.* Springer-Verlag London, Ltd., London, 2006.

[3] A. M. Bersani, E. Bersani, G. Dell'Acqua, and M. G. Pedersen. New trends and perspectives in nonlinear intracellular dynamics: one century from michaelis–menten paper. *Continuum Mech. Therm.*, 27:659–684, 2015.

[4] J. A. M. Borghans, R. J. De Boer, and L. A. Segel. Extending the quasi-steady state approximation by changing variables. *Bull. Math. Biol.*, 58:43–63, 1996.

[5] G. Côme. Mechanistic modelling of homogeneous reactors: A numerical method. *Comput. Chem. Eng.*, 3:603–609, 1979.

[6] B. Choi, G. A. Rempala, and J. K. Kim. Beyond the Michaelis–Menten equation: Accurate and efficient estimation of enzyme kinetic parameters. *Sci. Rep.*, 7:17018, 2017.

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

[8] J. Eilertsen and S. Schnell. The quasi-steady-state approximations revisited: Timescales, small parameters, singularities, and normal forms in enzyme kinetics. *Math. Biosci*, 325:108339, 2020.

[9] J. Eldering. *Normally hyperbolic invariant manifolds: The noncompact case*, volume 2 of *Atlantis Studies in Dynamical Systems.* Atlantis Press, Paris, 2013.

[10] N. Fenichel. Geometric singular perturbation theory for ordinary differential equations. *J. Differ. Equations*, 31:53–98, 1979.

[11] M. E. Fuentes, R. Varón, M. García-Moreno, and E. Valero. Kinetics of intra-and intermolecular zymogen activation with formation of an enzyme–zymogen complex. *FEBS J.*, 272:85–96, 2005.

[12] M.-E. Fuentes, R. Varón, M. García-Moreno, and E. Valero. Kinetics of autocatalytic zymogen activation measured by a coupled reaction: pepsinogen autoactivation. *Biol. Chem.*, 386:689–698, 2005.

[13] M. García-Moreno, B. H. Havsteen, R. Varón, and H. Rix-Matzen. Evaluation of the kinetic parameters of the activation of trypsinogen by trypsin. *Biochim. Biophys. Acta*, 1080:143–147, 1991.

[14] A. Goeke, C. Schilli, S. Walcher, and E. Zerz. Computing quasi-steady state reductions. *J. Math. Chem.*, 50:1495–1513, 2012.

[15] A. Goeke, S. Walcher, and E. Zerz. Classical quasi-steady state reduction – A mathematical characterization. *Physica D*, 345:11–26, 2017.

[16] D. Goluskin. Who ate whom: population dynamics with age-structured predation. *Proceedings: WHOI GFD 2010 program of study: Swimming and swirling in turbulence*, 2010.

[17] F. G. Heineken, H. M. Tsuchiya, and R. Aris. On the mathematical status of the pseudo-steady hypothesis of biochemical kinetics. *Math. Biosci.*, 1:95–113, 1967.

[18] C. K. R. T. Jones. Geometric singular perturbation theory, 1995.

[19] B. Kassell and J. Kay. Zymogens of proteolytic enzymes. *Science*, 180:1022–1027, 1973.

[20] A. R. Khan and M. N. G. James. Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci.*, 7:815–836, 1998.

[21] M. Krupa and P. Szmolyan. Extending slow manifolds near transcritical and pitchfork singularities. *Nonlinearity*, 14(6):1473–1491, 2001.

[22] C. Kuehn. *Multiple time scale dynamics*, volume 191 of *Applied Mathematical Sciences*. Springer, 2015.

[23] K. J. Laidler. Theory of the transient phase in kinetics, with special reference to enzyme systems. *Can. J. Chem.*, 33:1614–1624, 1955.

[24] J. H. Liu and Z. X. Wang. Kinetic analysis of ligand-induced autocatalytic reactions. *Biochem. J.*, 379 (3):697–702, 2004.

[25] U. Maas and S. Pope. Simplifying chemical kinetics: Intrinsic low-dimensional manifolds in composition space. *Combust. Flame*, 88:239–264, 1992.

[26] J. D. Meiss. *Differential dynamical systems*, volume 22 of *Mathematical Modeling and Computation*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, revised edition, 2017.

[27] B. O. Palsson. On the dynamics of the irreversible michaelis–menten reaction mechanism. *Chem. Eng. Sci.*, 42:447–458, 1987.

[28] B. O. Palsson and E. N. Lightfoot. Mathematical modelling of dynamics and control in metabolic networks. I. On Michaelis-Menten kinetics. *J. Theor. Biol.*, 111:273–302, 1984.

[29] M. G. Pedersen, A. M. Bersani, E. Bersani, and G. Cortese. The total quasi-steady-state approximation for complex enzyme reactions. *Math. Comput. Simulat.*, 79:1010–1019, 2008.

[30] D. L. Purich. *Enzyme kinetics: Catalysis and Control*. Elsevier, 2010.

[31] J. G. Reich and E. E. Sel'kov. Mathematical analysis of metabolic networks. *FEBS Lett.*, 40:S112–S118, 1974.

[32] C. G. Sanny, J. A. Hartsuck, and J. Tang. Conversion of pepsinogen to pepsin. Further evidence for intramolecular and pepsin-catalyzed activation. *J. Biol. Chem.*, 250:2635–2639, 1975.

[33] M. Schauer and R. Heinrich. Analysis of the quasi-steady-state approximation for an enzymatic one-substrate reaction. *J. Theor. Biol.*, 79:425–442, 1979.

[34] L. A. Segel. *Modeling dynamic phenomena in molecular and cellular biology*. Cambridge University Press, 1984.

[35] L. A. Segel. On the validity of the steady state assumption of enzyme kinetics. *Bull. Math. Biol.*, 50: 579–593, 1988.

[36] L. A. Segel and M. Slemrod. The Quasi-Steady-State Assumption: A case study in perturbation. *SIAM Rev.*, 31:446–477, 1989.

[37] W. Stroberg and S. Schnell. On the estimation errors of K_M and v from time-course experiments using the Michaelis—Menten equation. *Biophys. Chem.*, 219:17–27, 2016.

[38] G. Tans, J. Rosing, and J. Griffin. Sulfatide-dependent autoactivation of human blood coagulation factor xii (hageman factor). *J. Biol. Chem.*, 258:8215–22, 1983.

[39] G. Tans, J. Rosing, M. Berrettini, B. Lammie, and J. Griffin. Autoactivation of human plasma prekallikrein. *J. Biol. Chem.*, 262:11308–14, 1987.

[40] E. C. Thrower, A. P. E. Diaz de Villalvilla, T. R. Kolodecik, and F. S. Gorelick. Zymogen activation in a reconstituted pancreatic acinar cell system. *Am. J. Physiol. Gastrointest. Liver Physiol.*, 290: G894–G902, 2006.

[41] A. Tikhonov. Systems of differential equations containing small parameters in their derivatives. *Mat. Sb. (N.S.)*, 31:575–586, 1952.

[42] A. R. Tzafriri. Michaelis-Menten kinetics at high enzyme concentrations. *Bull. Math. Biol.*, 65:1111–1129, 2003.

[43] R. Varón, M. E. Fuentes, M. García-Moreno, F. García-Sevilla, E. Arias, E. Valero, and E. Arribas. Contribution of the intra- and intermolecular routes in autocatalytic zymogen activation: Application to pepsinogen activation. *Acta Biochim. Pol.*, 53:407–420, 2006.

[44] R. Varón, M. García-Moreno, D. Valera-Ruiperez, F. Garcia-Molina, F. Garcia-Canovas, R. Ladron-de Guevara, J. Masia-Perez, and B. Havsteen. Kinetic analysis of a general model of activation of aspartic proteinase zymogens. *J. Theor. Biol.*, 242:743–754, 2006.

[45] R. Varón, M. A. Minaya-Pacheco, F. García-Molina, E. Arribas, E. Arias, J. Masiá, and F. García-Sevilla. Competitive and uncompetitive inhibitors simultaneously acting on an autocatalytic zymogen activation reaction. *J. Enzyme Inhib. Med. Chem.*, 21:635–645, 2006.

[46] W.-N. Wang, X.-M. Pan, and Z.-X. Wang. Kinetic analysis of zymogen autoactivation in the presence of a reversible inhibitor. *Eur. J. Biochem.*, 271(23-24):4638–4645, 2004.

[47] M. Wechselberger. *Geometric Singular Perturbation Theory Beyond the Standard Forms.* Springer, 2020.

[48] J.-W. Wu, Y. Wu, and Z.-X. Wang. Kinetic analysis of a simplified scheme of autocatalytic zymogen activation. *Eur. J. Biochem.*, 268:1547–1553, 2001.