Sensitivity-conditioning: Beyond Singular Perturbation for Control Design on Multiple Time Scales

Journal Article

Author(s):
Picallo, Miguel; Bolognani, Saverio; Dörfler, Florian

Publication date:
2023-04

Permanent link:
https://doi.org/10.3929/ethz-b-000579529

Originally published in:
IEEE Transactions on Automatic Control 68(4), https://doi.org/10.1109/tac.2022.3176795

To appear on the IEEE Transactions on Automatic Control
https://doi.org/10.1109/TAC.2022.3176795

Sensitivity-Conditioning: Beyond Singular Perturbation for Control Design on Multiple Time Scales

Miguel Picallo, Saverio Bolognani, Florian Dörfler

***Abstract*— A classical approach to design controllers for interconnected systems is to assume that the different subsystems operate at different time scales, then design simpler controllers within each time scale, and finally certify stability of the interconnected system via singular perturbation analysis. In this work, we propose an alternative approach that also allows to design the controllers of the individual subsystems separately. However, instead of requiring a sufficiently large time-scale separation, our approach consists of adding a feed-forward term to modify the dynamics of faster systems in order to anticipate the dynamics of slower ones. We present several examples in bilevel optimization and cascade control design, where our approach improves the performance of currently available methods.**

***Index Terms*— Bilevel optimization, cascade control, interconnected systems, nonlinear control design, singular perturbation, time-scale separation.**

I. Introduction

Interconnected and nested systems are ubiquitous in control applications, but they may be challenging to analyse and design. If interconnected systems are composed by subsystems operating on multiple time scales [1] and a normal hyperbolicity condition holds [2], then each time scale can be studied independently, substituting dynamics of faster time scales by algebraic equations [3]. Such systems appear in engineering applications like power systems [4], [5], biological systems [6], motion control [7], electrical drives [8], etc. In that context, time-scale separation arguments, like singular perturbation analysis [9], [10], allow to certify when the stability guarantees derived in each separate time scale are preserved in the interconnected system. Standard singular perturbation considers only two time scales [11], although it can be extended to multiple ones [2], [5], [12].

Besides analysis, singular perturbation is also a powerful tool for control design [13], for example as a model reduction technique [14]: complex systems on a single time scale can be artificially separated into subsystems on different time scales, and thus simplify their analysis and controller design. Singular perturbation analysis can then provide additional conditions, for example on the control parameters [5], to ensure that the interconnected system remains stable. Some examples of these applications are hierarchical control architectures, like cascade control [15], or iterative optimization algorithms, like dual ascent [16], interior point methods [17], etc. However, for more than two time scales such singular perturbation conditions may be hard to derive, unless the interconnection present a specific structure [5], [12]. More importantly, since artificial time-scale separation slows down some subsystems with respect to others, it poses a fundamental limit on the convergence rate of the interconnected system.

In this work, we consider interconnected control systems in which the individual subsystems are designed and stabilized (e.g., by means of control) on separate time scales, and we are interested in preserving the overall system stability of the interconnection in a single time scale. Unlike the singular perturbation approach, we propose a single-time scale interconnection that guarantees closed-loop stability without imposing additional conditions on control parameters, nor slowing down any subsystem with respect to others. Additionally, our approach can deal with general interconnection structures, where the dynamics of each subsystem may depend on the states of all other subsystems. Our proposed interconnection can be interpreted as a transient feed-forward term in faster systems, that anticipates the dynamics of slower ones. For that, it uses the sensitivity of the fast system's steady state with respect to the slower system's state. Therefore, we term this approach the sensitivity-conditioning.

This new interconnection is inspired by recently proposed optimization algorithms to solve problems that are usually represented on multiple time scales: the prediction-correction algorithms for time-varying optimization [18], [19], the advance-steps in nonlinear model predictive control [20], [21], and the opponent-learning awareness games [22], [23]. These algorithms use the nonlinear optimization sensitivity [24], [25] to generate feed-forward terms that improve their convergence. Our approach also relates to classic backstepping [11, Ch. 14] in the context of overcoming time-scale separation limitations. However, unlike backstepping, our approach does not require to know a stabilizing state feedback law in closed form. Hence, our approach is implementable in cases where such a feedback law is not available.

Our contributions are the following: First, we divide the problem of designing the interconnection of two subsystems into a design problem of separate time-scales and a conditioning of their interconnection. For the latter, we define

Funding by the Swiss Federal Office of Energy through the project "Renewable Management and Real-Time Control Platform (ReMaP)" (SI/501810-01) and the ETH Foundation is gratefully acknowledged.

The authors are with the Automatic Control Laboratory at ETH Zürich, Switzerland. (emails: {miguelp,bsaverio,dorfler}@ethz.ch)

the sensitivity-conditioning approach, and we show how it corresponds to an additional transient signal to be exchanged between the two subsystems. Second, we prove that the sensitivity-conditioning approach ensures that the interconnected single-time-scale system has the same local (and even global, under further conditions) exponential stability properties as the multiple time scales system where each subsystem evolves on a different time scale. Third, we show how some degrees of freedom in the proposed design method can be used to improve the convergence rate of the interconnected system, and we provide robustness guarantees with respect to model errors. Fourth, we demonstrate the applicability of our approach on two control design problems: cascade control [15] and bilevel optimization [26]. Finally, we show how to extend our approach to multiple time scales.

The rest of this paper is structured as follows: Section II presents the type of systems that we consider and motivates our sensitivity-conditioning approach. Section III introduces the sensitivity-conditioning for two interconnected systems. Sections IV and V show the applications examples. Section VI shows how the sensitivity-conditioning can be extended to multiple-time-scales systems, and discrete-time systems. Finally, Section VII presents some conclusions.

II. Motivation

Consider the interconnection of two systems described by the vector fields $f_i(\cdot)$ on $x_i \in \mathbb{R}^{n_i}$, respectively for $i = 1, 2$:

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, x_2) \\ \dot{x}_2 &= f_2(x_1, x_2). \end{aligned} \tag{1}$$

The study (or design) of such interconnection is challenging in general. One way to tackle these analysis or design problems is to assume that the two subsystems in (1) evolve on separate time scales: x_2 evolves on a faster time τ, where x_1 is constant, while x_1 evolves on a slower time t, where the dynamics of x_2 are replaced by the algebraic equation $f_2(x_1, x_2) = 0$. This two-time-scale interconnection is represented with a differential-algebraic-equation system Σ_1, and a boundary-layer system Σ_2:

$$\Sigma_1 : \dot{x}_1 = f_1(x_1, x_2) \qquad \text{s.t. } f_2(x_1, x_2) = 0 \tag{2a}$$

$$\Sigma_2 : \frac{dx_2}{d\tau} = f_2(x_1, x_2) \qquad \text{s.t. } \frac{dx_1}{d\tau} = 0. \tag{2b}$$

Many interconnected systems become simpler to design and control when the two subsystems are assumed to evolve on two separate time scales as in (2). Classical examples are adaptive control [27], cascade control systems [15], where the state of one system is used as input to the other system, i.e., x_2 to control x_1, or nested iterative numerical algorithms (e.g., in optimization). In the rest of the paper, we assume that the analysis and design of the two-time-scale system (2) are tractable problems, and we provide some examples of how this is done for specific applications in Sections IV and V.

Clearly, any statements on the steady-state behavior and the stability of the two-time-scale system (2) does not automatically hold true for the original single-time-scale system (1). One standard way to ensure that the properties of (2) extend to (1) is to enforce a sufficient time-scale separation between the two subsystems and then employ the tools of singular perturbation analysis [11, Ch. 11]. Under the assumption that $f_2(x_1, \cdot)$ has a finite number of isolated roots $x_2^s(x_1)$, one can define the *standard* singular perturbation conditioned system

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, x_2) \\ \epsilon\dot{x}_2 &= f_2(x_1, x_2). \end{aligned} \tag{3}$$

where $0 < \epsilon \ll 1$ is a design parameter to enforce the desired level of time-scale separation. In the *singular limit* $\epsilon \to 0$, (3) becomes a degenerate system by Tikhonov's Theorem [3] and reduces to (2). Singular perturbation analysis allows to guarantee that if both systems in (2) are asymptotically stable, then the conditioned interconnection (3) is also stable (and has the same equilibria) when ϵ is below a certain threshold $\bar{\epsilon}$ [11, Thm. 11.3,4]. An example of a control design targeting a time-scale separated closed-loop system as in (3) is cascaded control, e.g., in power electronics control systems, where time-scale separation does not exist naturally, but has to be imposed artificially [5]. This type of conditioning comes at a cost: as the second subsystem cannot be made arbitrarily fast in practice, the design choice of $\epsilon \ll 1$ necessarily slows down the first subsystem and thus limits the convergence rate and deteriorates the performance of the entire interconnection (3).

In this paper, we propose an alternative conditioning of the interconnected system (1) without this drawback. For that, we define the conditioned interconnected system Σ as

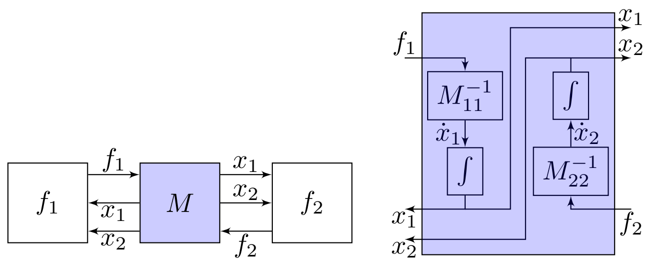
$$\Sigma : \quad M(x_1, x_2) \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{bmatrix}, \tag{4}$$

where $M(\cdot) = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}(\cdot)$ is a general non-singular conditioning matrix, i.e., a generalized time constant, which is a design variable to be chosen. Notice that the singular perturbation conditioned (3) is a special case of (4), with a specific matrix $M = \begin{bmatrix} I & 0 \\ 0 & \epsilon I \end{bmatrix}$, where I is the identity matrix. Nonetheless, a general M can represent a much larger class of interconnections, see Table I for an illustration. For example, it can represent the fully actuated interconnected control system

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, x_2) + u_1 \\ \dot{x}_2 &= f_2(x_1, x_2) + u_2, \end{aligned} \tag{5}$$

where the external control inputs u_i are active only the transient dynamics, i.e., $u_i = 0$ when $f_i(\cdot) = 0$, thus preserving the steady-states of the original system (1) and of the two-time-scale system (2). One example of this general conditioning technique (4) is backstepping in cascade systems [11, Ch. 14], which we will review in Section IV. Other examples appear in the design of nested gradient algorithms for continuous-time optimization, which we will discuss in Section V.

Note that the singular-perturbation conditioned system (3) uses a *transient* control action $u_2 = (\frac{1}{\epsilon} - 1)f_2(x_1, x_2)$ (and $u_1 = 0$) to induce time-scale separation. In this article, we propose an alternative conditioning matrix $M(\cdot)$ in (4) that, in the form (5), corresponds to a derivative-type control action $u_2 = M_{21} f_1(x_1, x_2)$. Loosely, we propose that the x_2-dynamics are additionally driven by $u_2 \approx \frac{d}{dt} x_2^s(x_1(t))$, where $x_2^s(x_1)$ is the steady state of the boundary layer system (2b)



(a) Σ (b) M

Fig. 1: Block diagram of the conditioned interconnected system Σ. Fig. 1b shows an example of a constant block-diagonal conditioning matrix $M(x_1, x_2) = \begin{bmatrix} M_1 & 0 \\ 0 & M_2 \end{bmatrix}$.

parametrized by x_1, i.e., $f_2(x_1, x_2^s(x_1)) = 0$. As a result, under the dynamics (5) we have $\frac{d}{dt}\|x_2(t) - x_2^s(x_1(t))\|_2^2 = 2(x_2(t) - x_2^s(x_1(t)))^T \big(f_2(x_1(t), x_2(t)) + u_2 - \frac{d}{dt}x_2^s(x_1(t))\big)$. Likewise, for (2b) we have $\frac{d}{d\tau}\|x_2(\tau) - x_2^s(x_1)\|_2^2 = 2(x_2(\tau) - x_2^s(x_1))^T f_2(x_1, x_2(\tau))$. In other words, if stability of the instantaneous steady-state $x_2^s(x_1)$ of (2b) can be inferred by means of a quadratic Lyapunov function, so can be the stability of the trajectory $x_2^s(x_1(t))$ of (5). In either case, the stability analysis of the coupled (x_1, x_2) dynamics reduces to that of a cascade system, and no time-scale separation is required.

However, generally $x_2^s(x_1(t))$ and its derivative $\frac{d}{dt}x_2^s(x_1(t))$ are not available in closed form. In what follows, we show how to construct an implementable surrogate for $u_2 \approx \frac{d}{dt}x_2^s(x_1(t))$, analyze the system stability without requiring $x_2^s(x_1(t))$, and extend the argument to an arbitrary number of subsystems.

III. Sensitivity-conditioning for two systems

A. Steady states and sensitivity

For this section we make the following standard simplifying assumption [11, Ch. 11], which we will partially relax later in Section VI:

Assumption 1. *The vector fields $f_i(\cdot)$ are continuously differentiable. For every x_1, $f_2(x_1, \cdot) = 0$ has a single root x_2^s, where the partial derivative $\nabla_{x_2} f_2(x_1, x_2^s)$ is invertible.*

Under Assumption 1, the implicit function theorem [28] guarantees the local existence of a continuously differentiable steady-state map $x_2^s(x_1)$, and gives the sensitivity of this steady state $x_2^s(x_1)$ with respect to x_1 as

$$\nabla_{x_1} x_2^s(x_1) = -\nabla_{x_2} f_2(x_1, x_2^s(x_1))^{-1} \nabla_{x_1} f_2(x_1, x_2^s(x_1)),$$

where $\nabla_{x_1} x_2^s(x_1) \in \mathbb{R}^{n_2 \times n_1}$. Even though this sensitivity is defined only at points where $x_2 = x_2^s(x_1)$, given Assumption 1 its analytic expression is well-defined at any point x_2 in a neighborhood of $x_2^s(x_1)$. This allows to define an extended sensitivity

$$S_{x_1}^{x_2}(x_1, x_2) := -\nabla_{x_2} f_2(x_1, x_2)^{-1} \nabla_{x_1} f_2(x_1, x_2), \qquad (6)$$

which satisfies the restriction $S_{x_1}^{x_2}(x_1, x_2^s(x_1)) = \nabla_{x_1} x_2^s(x_1)$.

The steady state map $x_2^s(x_1)$ allows us to redefine the differential-algebraic-equation system (2a) as a reduced-order system with reduced vector field $f_1^r(\cdot)$, so that the two-time-scale system (2) becomes

$$\Sigma_1 : \dot{x}_1 = f_1^r(x_1) := f_1(x_1, x_2^s(x_1)) \qquad (7a)$$

$$\Sigma_2 : \frac{dx_2}{d\tau} = f_2(x_1, x_2) \text{ s.t. } \frac{dx_1}{d\tau} = 0. \qquad (7b)$$

Then, each steady state x_1^s satisfying $f_1^r(x_1^s) = 0$, defines a steady state $(x_1^s, x_2^s(x_1^s))$ for (7), and by [11, Cor. 4.3] it is a locally exponentially stable steady state of the two-time-scale system (7) if and only if $\nabla_{x_2} f_2(x_1^s, x_2^s(x_1^s))$ and

$$\begin{aligned}&\nabla_{x_1} f_1^r(x_1^s) = \\ &\nabla_{x_1} f_1(x_1^s, x_2^s(x_1^s)) + \nabla_{x_2} f_1(x_1^s, x_2^s(x_1)) \nabla_{x_1} x_2^s(x_1^s)\end{aligned} \qquad (8)$$

have eigenvalues with negative real part, and unstable if any of these matrices has any eigenvalue with positive real part.

Remark 1. *Both the singular perturbed* (3) *and the two time-scale systems* (7) *have the same steady state $(x_1^s, x_2^s(x_1^s))$, but their local exponential stability properties may differ given the value of ϵ, because their Jacobians, and thus their local linearisations, may have different eigenvalues, see Table I.*

For example, consider the linear system $f_1(x_1, x_2) = x_1 - 2x_2$, $f_2(x_1, x_2) = \frac{1}{2}x_1 - \frac{1}{2}x_2$, thus $x_2^s(x_1) = x_1$, $\nabla_{x_1} x_2^s(x_1) = 1$, $f_1^r(x_1) = -x_1$ and $(x_1^s, x_2^s(x_1^s)) = (0, 0)$. Then (7) *is exponentially stable, since $J_2 = \nabla_{x_2} f_2(0, 0) = -\frac{1}{2} < 0$ and $J_1 = \nabla_{x_1} f_1^r(0) = -1 < 0$. The Jacobian of* (3) *in Table I has indeed negative eigenvalues for $\epsilon < \frac{1}{2}$, but positive ones for $\epsilon > \frac{1}{2}$. Thus* (3) *is exponentially stable if and only if $\epsilon < \bar{\epsilon} = \frac{1}{2}$.*

B. Sensitivity-conditioning interconnection

Here we present an alternative interconnection in (4), that can preserve the steady state $(x_1^s, x_2^s(x_1^s))$ of the two time-scale system (7) and its stability, without the need of a sufficiently large time-scale separation via a a singular parameter ϵ. This interconnection uses a sensitivity-conditioning matrix $M = \begin{bmatrix} I & 0 \\ -S_{x_1}^{x_2} & I \end{bmatrix}$, graphically presented in Table I:

$$\begin{bmatrix} I & 0 \\ -S_{x_1}^{x_2}(x_1, x_2) & I \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{bmatrix}. \qquad (9)$$

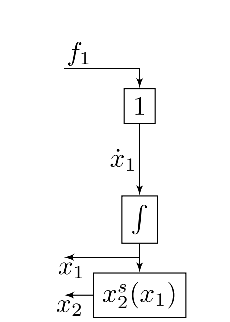
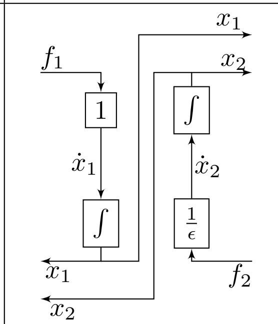
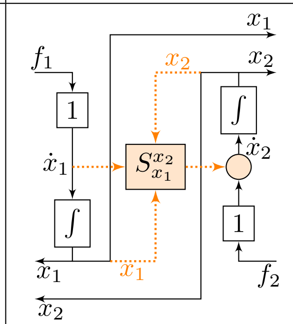
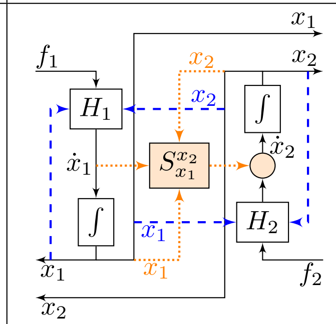
Instead of accelerating the second subsystem as in (3), this conditioning matrix M contains an off-diagonal term that changes the dynamics of the second subsystem to $\dot{x}_2 = f_2(x_1, x_2) + S_{x_1}^{x_2}(x_1, x_2) f_1(x_1, x_2)$, i.e., using the control input $u_2 = S_{x_1}^{x_2}(x_1, x_2) f_1(x_1, x_2)$ in (5). Intuitively, there are now two components in the vector field of $\dot{x}_2$: $f_2(x_1, x_2)$ drives x_2 to the steady state $x_2^s(x_1)$, while the sensitivity-conditioning $S_{x_1}^{x_2}(x_1, x_2) f_1(x_1, x_2)$ can be interpreted as a *feed-forward* term anticipating the change of $x_2^s(x_1)$ due to the dynamics $\dot{x}_1 \neq 0$. This second term affects the transient behavior only and vanishes at steady state.

Given Assumption 1, local existence and uniqueness [11, Thm. 3.1] of a solution $x_i(t)$ for (9) are guaranteed if:

Assumption 2. *The vector field $f_2(x_1, x_2) + S_{x_1}^{x_2}(x_1, x_2) f_1(x_1, x_2)$ is locally Lipschitz continuous.*

For more insight on the benefits of (9), we advance some results, that specialize the more general Theorem 1 (presented

TABLE I: Comparison of interconnections M

Case	Two time scales (7)	Singular perturbed (3)	Sensitivity-cond. (9)	Generalized S-C (16)
System Σ	$\dot{x}_1 = f_1^r(x_1)$ $\frac{dx_2}{d\tau} = f_2(x_1, x_2)$	$M \begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \begin{bmatrix}f_1(x_1,x_2)\\ f_2(x_1,x_2)\end{bmatrix}$		
Matrix M	$\begin{bmatrix}I & 0\\ 0 & 0\end{bmatrix}$	$\begin{bmatrix}I & 0\\ 0 & \epsilon I\end{bmatrix}$	$\begin{bmatrix}I & 0\\ -S_{x_1}^{x_2} & I\end{bmatrix}$	$\begin{bmatrix}H_1^{-1} & 0\\ 0 & H_2^{-1}\end{bmatrix}\begin{bmatrix}I & 0\\ -S_{x_1}^{x_2} & I\end{bmatrix}$
Block diagram of M in Fig. 1				
Jacobian J at $(x_1^s, x_2^s(x_1^s))$	$J_1 = \nabla_{x_1} f_1^r$ $J_2 = \nabla_{x_2} f_2$	$\begin{bmatrix}\nabla_{x_1} f_1 & \nabla_{x_2} f_1\\ \frac{1}{\epsilon}\nabla_{x_1} f_2 & \frac{1}{\epsilon}\nabla_{x_2} f_2\end{bmatrix}$	$\sim \begin{bmatrix}\nabla_{x_1} f_1^r & \nabla_{x_2} f_1\\ 0 & \nabla_{x_2} f_2\end{bmatrix}$	$\sim \begin{bmatrix}H_1 \nabla_{x_1} f_1^r & H_1 \nabla_{x_2} f_1\\ 0 & H_2 \nabla_{x_1} f_2\end{bmatrix}$
Eigenvalues λ & local stability	[11, Cor. 4.3]: Exp. stable if and only if λ of Jacobians J_i have negative real part.	**Remark 1**: different λ, can be unstable even if (7) stable. Stable if $\epsilon < \bar{\epsilon}$.	**Proposition 2 & Cor. 1**: similar to block-diagonal J, same λ as (7), thus same local stability.	**Proposition 4 & Cor. 3**: Preserving stability, λ of J can have lower negative real part than (7) and (9), thus faster convergence.

later in Section VI) to the case of two interconnected systems. The first proposition shows that the singleton $\{x_2^s(x_1)\}$ is a positively invariant set, i.e., once x_2 hits the steady state $x_2^s(x_1)$, it remains at $x_2^s(x_1)$ even if $\dot{x}_1 \neq 0$:

Proposition 1 (Positive invariance)**.** *Consider the dynamics of x_2 in* (9) *initialized at time t_0:*

$$\dot{x}_2 = f_2(x_1, x_2) + S_{x_1}^{x_2}(x_1, x_2)\dot{x}_1 \text{ s.t. } x_2(t_0) = x_2^s(x_1(t_0)). \tag{10}$$

Then, $x_2(t) = x_2^s(x_1(t))$ is the unique solution on the open domain of existence.

Proof. First, note that $x_2(t) = x_2^s(x_1(t))$ satisfies (10):

$$\begin{aligned}\dot{x}_2(t) &= f_2(x_1(t), x_2(t)) + S_{x_1}^{x_2}(x_1(t), x_2(t))\dot{x}_1(t)\\ &= \cancel{f_2(x_1(t), x_2^s(x_1(t)))} + \underbrace{S_{x_1}^{x_2}(x_1(t), x_2^s(x_1(t)))}_{\overset{(6)}{=}\nabla_{x_1} x_2^s(x_1(t))}\dot{x}_1(t)\\ &= \frac{dx_2^s(x_1(t))}{dt}\end{aligned}$$

Local existence and uniqueness of a solution is guaranteed by Assumption 2. Thus, $x_2(t) = x_2^s(x_1(t))$ is the unique solution on this domain of existence, because the derivatives and initial conditions of $x_2(t)$ and $x_2^s(x_1(t))$ coincide for $t \geq t_0$. ■

Moreover, the sensitivity-conditioning (9) allows to preserve the local stability of the two-time-scale system (7):

Proposition 2 (Local stability)**.** *At a steady state $(x_1^s, x_2^s(x_1^s))$, the Jacobian J of* (9) *satisfies:*

$$J \sim \begin{bmatrix}\nabla_{x_1} f_1^r(x_1^s) & \nabla_{x_2} f_1(x_1^s, x_2^s(x_1^s))\\ 0 & \nabla_{x_2} f_2(x_1^s, x_2^s(x_1^s))\end{bmatrix}, \tag{11}$$

where f_1^r is the reduced vector field from (7a)*, and $\sim$ denotes similarity, i.e., related by a similarity transformation.*

Proof. To calculate the Jacobian J of (9) at $(x_1^s, x_2^s(x_1^s))$, we invert M as $M^{-1} = \begin{bmatrix}I & 0\\ -S_{x_1}^{x_2} & I\end{bmatrix}^{-1} = \begin{bmatrix}I & 0\\ S_{x_1}^{x_2} & I\end{bmatrix}$, take derivatives, and evaluate them at steady-state, so that $f_i(x_1^s, x_2^s(x_1^s)) = 0$:

$$J = M^{-1}\begin{bmatrix}\nabla_{x_1} f_1 & \nabla_{x_2} f_1\\ \nabla_{x_1} f_2 & \nabla_{x_2} f_2\end{bmatrix},$$

where for clarity we omit the evaluation point $(x_1^s, x_2^s(x_1^s))$ in the notation. This J is similar to $\tilde{J} := MJM^{-1}$, where

$$J \sim \tilde{J} = \cancel{MM^{-1}}\begin{bmatrix}\nabla_{x_1} f_1 & \nabla_{x_2} f_1\\ \nabla_{x_1} f_2 & \nabla_{x_2} f_2\end{bmatrix} M^{-1} \overset{(6),(8)}{=} \begin{bmatrix}\nabla_{x_1} f_1^r & \nabla_{x_2} f_1\\ 0 & \nabla_{x_2} f_2\end{bmatrix}.$$

■

Corollary 1. *The Jacobians of the two time-scale system* (7) *and the sensitivity-conditioning conditioned system* (9) *have the same eigenvalues, and thus the same local stability properties.*

Remark 2. *The cancellation of one off-diagonal term in* (11) *is due to the sensitivity definition in* (6)*, and will also play a crucial role in the proofs of the results to come. Essentially, the role of the sensitivity-conditioning is to turn a closed-loop into a cascade system from the viewpoint of stability analysis, see also the later Remark 4.*

Propositions 1 and 2 and Corollary 1 establish that invariance and local exponential stability of $x_2^s(x_1)$ are preserved from the two-time-scale system (7) in the single-time-scale sensitivity-conditioning one (9). Furthermore, these results can be extended to contraction regions satisfying the following:

Assumption 3 (Contraction region [29])**.** *There exists $\eta_2 > 0$ and an open ball $\mathcal{B}_{r_2}\big(x_2^s(x_1)\big) = \{x_2 \| \|x_2 - x_2^s(x_1)\|_2 < r_2\}$ centered at $x_2^s(x_1)$, with a positive radius $r_2 > 0$, and a metric defined by a constant symmetric positive definite $P_2 \succ 0$, such that for $x_2 \in \mathcal{B}_{r_2}\big(x_2^s(x_1)\big)$ it holds uniformly for all x_1 that*

$$P_2 \nabla_{x_2} f_2(x_1, x_2) + \nabla_{x_2} f_2(x_1, x_2)^T P_2 \preceq -\eta_2 P_2$$

Under this Assumption 3, $\mathcal{B}_{r_2}\big(x_2^s(x_1)\big)$ is a contraction region for the boundary-layer system (2b) for all x_1, within which the invariant set $x_2^s(x_1)$ (see Proposition 1) is exponentially stable [29, Thm. 2]. Then, such a contraction region is also preserved under the sensitivity-conditioning (9):

Proposition 3 (Stability with a contracting boundary layer)**.** *Under Assumption 3 the following holds:*

1) *The sensitivity-conditioning interconnection* (9) *is well-defined for $x_2 \in \mathcal{B}_{r_2}(x_2^s(x_1))$, i.e., $\nabla_{x_2} f_2(x_1, x_2)$ is uniformly invertible in $\mathcal{B}_{r_2}(x_2^s(x_1))$.*
Furthermore, there exists $r_{2,0} < r_2$ such that if $x_2(0) \in \mathcal{B}_{r_{2,0}}\big(x_2^s(x_1(0))\big)$, then $x_2(t) \in \mathcal{B}_{r_2}\big(x_2^s(x_1(t))\big)$ for all $t > 0$, and $x_2^s(x_1(t))$ is a locally exponentially stable trajectory for x_2 in (9)*, i.e., there exists $\eta, K > 0$ so that $\|x_2(t) - x_2^s(x_1(t))\|_2 \le K \|x_2(0) - x_2^s(x_1(0))\|_2 e^{-\eta t}$.*
2) *Additionally, assume that x_1^s is an asymptotically stable steady state of the reduced-order system* (7a)*, that the ball $\mathcal{B}_{r_1}(x_1^s)$ is in its region of attraction, and that $f_1^r(\cdot)$ is continuously differentiable in the closure of $\mathcal{B}_{r_1}(x_1^s)$.*
Then, there exists $r_{1,0} \le r_1$ and $\tilde{r}_{2,0} \le r_{2,0}$, such that if $x_1(0) \in \mathcal{B}_{r_{1,0}}(x_1^s)$ and $x_2(0) \in \mathcal{B}_{\tilde{r}_{2,0}}(x_2^s(x_1(0)))$, then $x_1(t) \in \mathcal{B}_{r_1}(x_1^s)$, and x_1^s is asymptotically stable under the sensitivity-conditioning interconnection (9)*.*

Proof. We use the following technical result:

Lemma 1. *Consider a system $\dot{x} = f(x)$, with steady-state x^s and a continuous differentiable $f(\cdot)$. If there exist a radius $r > 0$, a symmetric positive definite matrix $P \succ 0$, and a parameter $\eta > 0$, such that $P \nabla_x f(x) + (\nabla_x f(x))^T P^T \preceq -\eta P$ for $x \in \mathcal{B}_r(x^s)$, then in $\mathcal{B}_r(x^s)$ it holds that*

1) *the inverse of the Jacobian $\nabla_x f(x)$ exists and is bounded: $\|\nabla_x f(x)^{-1}\|_2 \le \frac{2\lambda_{\max}(P)}{\eta \lambda_{\min}(P)}$, and*
2) *the vector field $f(x)$ is lower bounded: $\|f(x)\|_P \ge \frac{\eta}{2}\|x - x^s\|_P$.*

Here $\lambda_{\min}(\cdot), \lambda_{\max}(\cdot)$ denote the minimum and maximum eigenvalues.

Proof. See Appendix I. We remark that the second result can be seen as a particular case of [30, Prop. 3]. ■

1) First, note that Lemma 1 assures non-singularity of $\nabla_{x_2} f_2$. Consider the following Krasovskii Lyapunov function $V_2(x_1, x_2) = \|f_2(x_1, x_2)\|_{P_2}^2$ [31, Ch. 5] for x_2 in (9). Since $P_2 \nabla_{x_2} f_2 + \nabla_{x_2} f_2^T P_2^T \preceq -\eta_2 P_2$, under the sensitivity-conditioning dynamics (9) we have:

$$\begin{aligned}
\dot{V}_2 &= f_2^T P_2 \nabla_{x_1} f_2 \dot{x}_1 + f_2^T P_2 \nabla_{x_2} f_2 \dot{x}_2 \\
&\quad + (f_2^T P_2 \nabla_{x_1} f_2 \dot{x}_1 + f_2^T P_2 \nabla_{x_2} f_2 \dot{x}_2)^T \\
&\overset{(9)}{=} f_2^T P_2 \cancel{\big(\nabla_{x_1} f_2 + \nabla_{x_2} f_2 S_{x_1}^{x_2}\big) \dot{x}_1} + f_2^T P_2 \nabla_{x_2} f_2 f_2 \\
&\quad + (f_2^T P_2 \cancel{\big(\nabla_{x_1} f_2 + \nabla_{x_2} f_2 S_{x_1}^{x_2}\big) \dot{x}_1} + f_2^T P_2 \nabla_{x_2} f_2 f_2)^T \\
&\overset{(6)}{=} f_2^T \big(P_2 \nabla_{x_2} f_2 + \nabla_{x_2} f_2^T P_2^T\big) f_2 \le -\eta_2 V_2
\end{aligned} \tag{12}$$

Hence, $\|f_2(x_1(t), x_2(t))\|_{P_2}^2 \le \|f_2(x_1(0), x_2(0))\|_{P_2}^2 e^{-\eta_2 t}$. Since $f_2(\cdot)$ is continuously differentiable, it is locally Lipschitz continuous in $\mathcal{B}_{r_2}(x_2^s(x_1))$ with some constant L_{f_2}, and by involving Lemma 1 we have

$$\begin{aligned}
&\|x_2(t) - x_2^s(x_1(t))\|_2^2 \le \tfrac{4}{\eta_2^2 \lambda_{\min}(P_2)} \|f_2(x_2(t), x_1(t))\|_{P_2}^2 \\
&\le \underbrace{\frac{4L_{f_2}^2 \lambda_{\max}(P_2)}{\eta_2^2 \lambda_{\min}(P_2)}}_{=\frac{r_2^2}{r_{2,0}^2}} e^{-\eta_2 t} \|x_2(0) - x_2^s(x_1(0))\|_2^2,
\end{aligned} \tag{13}$$

where $r_{2,0} := \frac{\eta_2 r_2}{2L_{f_2}} \sqrt{\frac{\lambda_{\min}(P_2)}{\lambda_{\max}(P_2)}} \le r_2$. Hence, if $\|x_2(0) - x_2^s(x_1(0))\|_2 < r_{2,0}$, then $x_2(t) \in \mathcal{B}_{r_2}\big(x_2^s(x_1(t))\big)$ for all $t > 0$, and x_2 converges exponentially to $x_2^s(x_1)$ under (9), despite the varying x_1.

2) Now we analyse the x_1-dynamics subject to the exponential converging input $x_2(t) - x_2^s(x_1(t))$. Since $f_1^r(\cdot)$ is continuously differentiable, it is locally Lipschitz continuous, and $\nabla_{x_1} f_1^r(\cdot)$ is bounded in $\mathcal{B}_{r_1}(x_1^s)$. Hence, by the converse Lyapunov theorem [11, Thm. 4.16], there exists a Lyapunov function $V_1(x_1)$ satisfying:

$$\begin{aligned}
\alpha_1(\|x_1 - x_1^s\|_2) \le V_1(x_1) \le \alpha_2(\|x_1 - x_1^s\|_2) \\
\nabla_x V_1(x_1)^T f_1^r(x_1) \le -\alpha_3(\|x_1 - x_1^s\|_2) \\
\|\nabla_x V_1(x_1)\|_2 \le \alpha_4(\|x_1 - x_1^s\|_2),
\end{aligned} \tag{14}$$

where $\alpha_i(\cdot)$ are $\mathcal{K}$-functions. Since $f_1(\cdot)$ is continuously differentiable, it is locally Lipschitz continuous in $\mathcal{B}_{r_1}(x_1^s) \times \mathcal{B}_{r_2}(x_2^s(x_1))$ with some constant L_{f_1}. Then, the Lyapunov function $V_1(x_1)$ under the sensitivity-conditioning dynamics (9) satisfies:

$$\begin{aligned}
\dot{V}_1 &\le \nabla_{x_1} V_1^T f_1 \le \nabla_{x_1} V_1^T f_1^r + \|\nabla_x V_1(x_1)\|_2 \|f_1 - f_1^r\|_2 \\
&\overset{(14)}{\le} -\alpha_3(\|x_1 - x_1^s\|_2) + L_{f_1} \alpha_4(\|x_1 - x_1^s\|_2) \|x_2 - x_2^s(x_1)\|_2 \\
&\overset{(13)}{\le} -\alpha_3(\alpha_2^{-1}(V_1)) \\
&\qquad + \alpha_4(\alpha_1^{-1}(V_1)) \tfrac{r_2 L_{f_1}}{r_{2,0}} e^{-\eta_2 t} \|x_2(0) - x_2^s(x_1(0))\|_2.
\end{aligned}$$

Consider any $\delta \in (0, r_1)$, and define $r_{1,0} := \alpha_2^{-1}(\alpha_1(r_1 - \delta)) < r_1$ and $\tilde{r}_{2,0} := \min\Big(r_{2,0}, \frac{\alpha_3(r_{1,0})}{\alpha_4(r_1 - \delta)\frac{r_2 L_{f_1}}{r_{2,0}}}\Big)$. If $x_1(0) \in \mathcal{B}_{r_{1,0}}(x_1^s)$, then $V_1(x_1(0)) \le \alpha_1(r_1 - \delta)$; and if $x_2(0) \in \mathcal{B}_{\tilde{r}_{2,0}}\big(x_2^s(x_1(0))\big)$, then $\dot{V}_1 \le 0$ whenever $V_1 = \alpha_1(r_1 - \delta)$. Hence, $V_1(x_1(t)) < \alpha_1(r_1)$ and $\|x_1(t) - x_1^s\|_2 < r_1$ for all $t > 0$. Furthermore, for any $\epsilon \in (0, 1)$ it holds that

$$\begin{aligned}
\dot{V}_1 &\le -\alpha_3(\|x_1 - x_1^s\|_2) + L_{f_1} \alpha_4(\|x_1 - x_1^s\|_2) \|x_2 - x_2^s(x_1)\|_2 \\
&\le -\epsilon \alpha_3(\|x_1 - x_1^s\|_2) - (1 - \epsilon) \alpha_3(\|x_1 - x_1^s\|_2) \\
&\qquad + L_{f_1} \alpha_4(r_1) \tfrac{r_2 L_{f_1}}{r_{2,0}} e^{-\eta_2 t} \|x_2(0) - x_2^s(x_1(0))\|_2 \\
&\le -\epsilon \alpha_3(\|x_1 - x_1^s\|_2),
\end{aligned}$$

where the last inequality holds while $\|x_1 - x_1^s\|_2 \ge \alpha_3^{-1}\big(\frac{1}{1-\epsilon} L_{f_1} \alpha_4(r_1) \frac{r_2 L_{f_1}}{r_{2,0}} e^{-\eta_2 t} \|x_2(0) - x_2^s(x_1(0))\|_2\big)$. Hence, x_1^s is asymptotically stable, because it is input-to-state stable

[11, Thm. 4.18] with respect to a vanishing input, see also [11, Lemma 4.7]. ■

Remark 3 (Connection to contraction theory)**.** *The exponential stability of the boundary-layer system* (2b) *for a constant* x_1 *is a standard assumption in the context of singular perturbation analysis [11, Thm. 11.4 and after], and exponential stability implies the existence of a contraction region [29, Reverse Thm. 2]. Proposition 3 establishes that the boundary-layer exponential stability can be preserved in the single-time-scale interconnection* (4) *using the sensitivity-conditioning* (9)*, independently of* $\dot{x}_1$*. Essentially, the sensitivity-conditioning* (9) *turns a system* Σ_2 *that is only contracting under a constant* x_1*, as in* (2b)*, into a partially contracting system* (9) *in* x_2 *[32, Def. 1] under a time-varying* x_1*. Then,* $\|x_2 - x_2^s(x_1)\|$ *becomes an exponentially decaying perturbation for* x_1 *in* (9)*, and thus asymptotic stability of the reduced-order system* (7a) *can be preserved in* (9) *under some additional conditions.*

Remark 4 (Connection to backstepping)**.** *The role of the sensitivity term* $S_{x_1}^{x_2}$ *in* (9) *is to cancel a cross term in the stability analysis of* x_2 *that appears under a time-varying* x_1*, see* (12) *and the proofs of Propositions 2 and 3. In other words, the sensitivity-conditioning* (9) *is turning an interconnected system* (4) *into a cascaded one from the viewpoint of stability analysis, see Remark 2. In a more general setting for Proposition 3, we could assume that the boundary-layer system* (2b) *is asymptotically stable with a general Lyapunov function* $V_2(x_1, x_2)$ *that is positive definite with respect to* $\|x_2 - x_2^s(x_1)\|$*, and satisfies* $\nabla_{x_2} V_2^T f_2 \leq 0$ *uniformly over* x_1*. Then, under the sensitivity-conditioning dynamics* (9) *we would have* $\dot{V}_2 \leq \nabla_{x_2} V_2^T f_2 + (\nabla_{x_2} V_2^T S_{x_2}^{x_1} - \nabla_{x_1} V_2^T)\dot{x}_1 \leq (\nabla_{x_2} V_2^T S_{x_1}^{x_2} - \nabla_{x_1} V_2^T)\dot{x}_1$*, which can be cancelled by choosing an appropriate sensitivity* $S_{x_2}^{x_1}$*. In Proposition 3 we considered* $V_2 = \|f_2\|_{P_2}^2$*. An other option would be* $V_2 = \|x_2 - x_2^s(x_1)\|_{P_2}^2$*, which requires the alternative sensitivity* $\nabla_{x_1} x_2^s(x_1) = -\nabla_{x_2} f_2(x_1, x_2^s(x_1))^{-1} \nabla_{x_1} f_2(x_1, x_2^s(x_1))$ *to cancel the term* $\nabla_{x_2} V_2^T S_{x_2}^{x_2} - \nabla_{x_1} V_2^T = 2(x_2 - x_2^s(x_1))P_2(S_{x_2}^{x_2} - \nabla_{x_1} x_2^s(x_1))$*. This sensitivity* $\nabla_{x_1} x_2^s(x_1)$ *can be interpreted as a backstepping-like approach [11, Ch. 14] to cancel the dynamics of* $\dot{x}_1 \neq 0$ *in* $\dot{V}_2$*, and again turn* (4) *in a cascaded system from the viewpoint of stability analysis. Note that for this sensitivity similar results as in Propositions 2 and 3 can be derived under suitable assumptions. However, implementing the corresponding interconnection* $M = \left[\begin{smallmatrix} I & 0 \\ -\nabla_{x_1} x_2^s(x_1) & I \end{smallmatrix}\right]$ *(or equivalently* $u_2 = \nabla_{x_1} x_2^s(x_1) f_1(x_1, x_2) = \frac{d}{dt} x_2^s(x_1(t))$*) may not be feasible, since it requires a closed-form expression for* $x_2^s(x_1)$ *to evaluate* $\nabla_{x_1} x_2^s(x_1)$*. Such a closed-form expression for* $x_2^s(x_1)$ *may be available in special cases, see cascade control in Section IV, but not in general, see bilevel optimization in Section V. On the other hand, choosing* $V_2 = \|f_2\|_{P_2}^2$ *results in the sensitivity in* (6)*, which does not require to know* $x_2^s(x_1)$*. In this context, the sensitivity-conditioning* (9) *acts as an implementable substitute for such a backstepping-like approach, with the same local properties.*

Corollary 2 (Global exponential stability)**.** *Assume that the vector fields* $f_i(x_1, x_2)$ *are Lipschitz continuous, and there exist* $P_i \succ 0$ *and* $\eta_i > 0$ *such that the following contraction conditions hold globally for all* x_1 *and* x_2*:*

$$\begin{aligned} P_1 \nabla_{x_1} f_1^r(x_1) + (\nabla_{x_1} f_1^r(x_1))^T P_1 &\preceq -\eta_1 P_1 \\ P_2 \nabla_{x_2} f_2(x_1, x_2) + \nabla_{x_2} f_2(x_1, x_2)^T P_2 &\preceq -\eta_2 P_2 \end{aligned} \tag{15}$$

Then $(x_1^s, x_2^s(x_1^s))$ *is a globally exponentially stable steady state of both the two time-scale system* (7) *and the sensitivity-conditioning interconnection* (9)*.*

Proof. If the contraction conditions (15) hold, (7) is globally exponentially stable [29, Thm. 2]. For (9), consider now the Lyapunov function $V(x_1, x_2) = V_1(x_1) + \theta V_2(x_1, x_2)$, where $\theta > 0$, $V_1(x_1) = \|f_1^r(x_1)\|_{P_1}^2$ and $V_2(x_1, x_2) = \|f_2(x_1, x_2)\|_{P_2}^2$. From (12) we have $\dot{V}_2 \leq -\eta_2 V_2$, then

$$\begin{aligned} \dot{V}_1 &= (f_1^r)^T P_1 \nabla_{x_1} f_1^r \dot{x}_1 + ((f_1^r)^T P_1 \nabla_{x_1} f_1^r \dot{x}_1)^T \\ &\overset{(9)}{=} (f_1^r)^T \big(P_1 \nabla_{x_1} f_1^r + (\nabla_{x_1} f_1^r)^T P_1^T\big) f_1^r \\ &\quad + 2(f_1^r)^T P_1 \nabla_{x_1} f_1^r (f_1 - f_1^r) \\ &\overset{(15)}{\leq} -\eta_1 \|f_1^r\|_{P_1}^2 + 2L_{f_1} \|f_1^r\|_{P_1} \|f_1 - f_1^r\|_{P_1} \\ &\leq -\eta_1 \|f_1^r\|_{P_1}^2 + 2L_{f_1}^2 \sqrt{\frac{\lambda_{\max}(P_1)}{\lambda_{\min}(P_2)}} \|f_1^r\|_{P_1} \|x_2 - x_2^s\|_{P_2} \\ &\overset{Lem.\ 1}{\leq} -\eta_1 \|f_1^r\|_{P_1}^2 + \underbrace{\sqrt{\frac{\lambda_{\max}(P_1)}{\lambda_{\min}(P_2)}} \frac{4L_{f_1}^2}{\eta}}_{2\nu} \|f_1^r\|_{P_1} \|f_2\|_{P_2}, \end{aligned}$$

Hence, $\dot{V} \overset{(12)}{\leq} - \begin{bmatrix} \|f_2\|_{P_2} \\ \|f_1^r\|_{P_1} \end{bmatrix}^T \begin{bmatrix} \theta\eta_2 & -\nu \\ -\nu & \eta_1 \end{bmatrix} \begin{bmatrix} \|f_2\|_{P_2} \\ \|f_1^r\|_{P_1} \end{bmatrix} \leq -\zeta V$, for some $\zeta > 0$, by choosing $\theta > \frac{\nu^2}{\eta_1 \eta_2}$. ■

C. Accelerated sensitivity-conditioning

The design of the conditioning matrix $M(x_1, x_2)$ in (4) offers to generalize the sensitivity-conditioning (9) to introduce additional degrees of freedom and achieve a better performance of the interconnection, e.g., a faster convergence. Consider two uniformly positive definite matrices $H_1(x_1, x_2) \succ 0, H_2(x_1, x_2) \succ 0$, and a generalized sensitivity-conditioning:

$$H(x_1, x_2)^{-1} \begin{bmatrix} I & 0 \\ -S_{x_1}^{x_2}(x_1, x_2) & I \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{bmatrix}, \tag{16}$$

where $H(x_1, x_2) = \left[\begin{smallmatrix} H_1(x_1, x_2) & 0 \\ 0 & H_2(x_1, x_2) \end{smallmatrix}\right]$.

Proposition 4 (Extension of Propositions 1, 2, and Corollary 2)**.** *The generalized sensitivity-conditioning* (16) *satisfies:*

1) *The singleton* $\{x_2^s(x_1)\}$ *is a positively invariant set under the sensitivity-conditioning dynamics* (16)*.*
2) *At steady state* $(x_1^s, x_2^s(x_1^s))$*, the Jacobian of* (16) *satisfies*
$$J \sim \begin{bmatrix} H_1 \nabla_{x_1} f_1^r & H_1 \nabla_{x_2} f_1 \\ 0 & H_2 \nabla_{x_2} f_2 \end{bmatrix}$$
3) *If the vector fields* $f_i(x_1, x_2)$ *are Lipschitz continuous, and there exists* $P_i \succ 0$ *and* $\eta_i > 0$ *such that:*
$$\begin{aligned} P_1 H_1 \nabla_{x_1} f_1^r + (\nabla_{x_1} f_1^r)^T H_1^T P_1 &\preceq -\eta_1 P_1 \\ P_2 H_2 \nabla_{x_2} f_2 + \nabla_{x_2} f_2^T H_2^T P_2 &\preceq -\eta_2 P_2, \end{aligned}$$
then $(x_1^s, x_2^s(x_1^s))$ *is a globally exponentially stable steady state of the generalized sensitivity-conditioning* (16)*.*

For clarity we omit the evaluations at $(x_1^s, x_2^s(x_1^s))$*.*

Proof. The proof follows analogous steps as the ones for Proposition 1, 2, and Corollary 2. ■

Corollary 3 (Accelerated sensitivity-conditioning)**.** *If the two-time-scale system* (7) *is locally exponentially stable, i.e.,* $J_1 = \nabla_{x_1} f_1^r(x_1^s)$ *and* $J_2 = \nabla_{x_2} f_2(x_1^s, x_2^s(x_1^s))$ *have eigenvalues with strictly negative real part, the generalized sensitivity-conditioning system* (16) *is locally exponentially stable if using positive scalars* $h_i > 0$ *and* $H_i(x_1, x_2) = h_i I$*. Moreover, the exponential convergence rate is improved for* $h_i > 1$*.*

See Table I for a comparison of the sensitivity-conditioning approach (9) and (16), the two-time-scale system (7) and the singular pertubated (3), summarizing these results.

To conclude, the sensitivity-conditioning (9) allows to preserve the stability of the two-time-scale system (7) in a single time-scale. This way, the need of artificially slowing down one subsystem (and, consequently, their interconnection) through a singular perturbation (3) is removed. However, a disadvantage of the sensitivity-conditioning is that it could produce large inputs u for Σ_2 in (5), even larger if using the generalization in (16), which changes the sensitivity-conditioning term to $S_{x_1}^{x_2}(x_1, x_2)H_1(x_1, x_2)f_1(x_1, x_2)$. This sensitivity-conditioning term could even become unrealizable in systems with control saturation in u. On the other hand, if the two-time-scale system (7) is lcoally stable, the generalization (16) can be chosen as $H_1(x_1, x_2) = \epsilon$ with a sufficiently small ϵ, see Corollary 3. Then, Σ_1 can be slowed down as with a singular perturbation term, and the sensitivity-conditioning term can be made realizable. This interpretation suggests that singular perturbation (3) and sensitivity-conditioning (9) are not mutually exclusive, but can be combined. Interestingly, Corollary 3 also allows to choose arbitrary time scales, e.g., Σ_1 faster than Σ_2 by choosing $h_1 \gg h_2$, and still preserve the stability of (7).

D. Robust sensitivity-conditioning

The sensitivity-conditioning system (9) requires a precise knowledge of the vector fields, essentially the model of the system, to evaluate the matrix $S_{x_1}^{x_2}(x_1, x_2)$. Here we analyse the implications of model errors: assume that instead of $S_{x_1}^{x_2}(x_1, x_2)$, only an approximation $\hat{S}_{x_1}^{x_2}(x_1, x_2)$ is available, and consider the approximated sensitivity-conditioning

$$\begin{bmatrix} I & 0 \\ -\hat{S}_{x_1}^{x_2}(x_1, x_2) & I \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{bmatrix}, \tag{17}$$

which also preserves the steady-state $(x_1^s, x_2^s(x_1^s))$.

Under Assumptions 1 and 2, we can derive a strong robustness certificate in the form of input-to-state stability [11, Def. 4.7]:

Proposition 5 (Input-to-state stability)**.** *Consider the same conditions as in Corollary 2, and define the error*

$$\xi := (S_{x_1}^{x_2}(x_1, x_2) - \hat{S}_{x_1}^{x_2}(x_1, x_2))f_1(x_1, x_2).$$

Then, the approximated sensitivity-conditioning system (17) *is input-to-state stable with respect to* ξ*.*

Proof. Since $(x_1, x_2^s(x_1))$ under (9) is exponentially stable under the conditions of Corollary 2, it is input-to-state stable in (17) with respect to ξ [11, Lemma 4.6]. ■

If the contraction conditions (15) in Corollary 2 do not hold globally, a local result along the line of Proposition 3 can be derived based on the local exponential stability of $x_2^s(x_1)$.

IV. Example I: Cascade Control

Consider a standard cascade control architecture [15], see Fig. 2, with a fast-inner closep-loop system Σ_2, and a slow-outer closed-loop system Σ_1, both with plants P_i and controllers C_i. More concretely, consider an example with two linear scalar first-order systems:

$$\begin{aligned} P_1 &: \dot{x}_1 = a_1 x_1 + b_1 u_1, \; u_1 = x_2 \\ P_2 &: \dot{x}_2 = a_2 x_2 + b_2 u_2, \end{aligned}$$

where all parameters are real-valued. PI (proportional-integral) controllers are typically used for C_1, C_2:

$$\begin{aligned} C_1 &: x_2^r = u_1^r = -\tfrac{1}{b_1}\big(a_1 x_1 + K_{P,1}(x_1 - x_1^r) + K_{I,1}\zeta_1\big) \\ &\quad \dot{\zeta}_1 = (x_1 - x_1^r) \\ C_2 &: u_2 = -\tfrac{1}{b_2}\big(a_2 x_2 + K_{P,2}(x_2 - x_2^r) + K_{I,2}\zeta_2\big) \\ &\quad \dot{\zeta}_2 = (x_2 - x_2^r). \end{aligned} \tag{18}$$

where ζ_i are the integral error states, $K_{P,i}, K_{I,i}$ are control gains to be determined, and the terms $a_i x_i$ are feed-forward terms to cancel the system dynamics. If the systems Σ_i had a time-scale separation as (7), the resulting interconnected system, with states x_i, ζ_i for each Σ_i, can be expressed as:

$$\begin{aligned} \Sigma_1 : \dot{x}_1 &= -K_{P,1}(x_1 - x_1^r) - K_{I,1}\zeta_1 \\ \dot{\zeta}_1 &= (x_1 - x_1^r) \\ \Sigma_2 : \tfrac{dx_2}{d\tau} &= -K_{P,2}(x_2 - x_2^r) - K_{I,2}\zeta_2 \\ \tfrac{d\zeta_2}{d\tau} &= (x_2 - x_2^r) \\ x_2^r &= -\tfrac{1}{b_1}\big(a_1 x_1 + K_{P,1}(x_1 - x_1^r) + K_{I,1}\zeta_1\big) \end{aligned} \tag{19}$$

which admits the globally asymptotically stable steady state $x_1^s = x_1^r, x_2^s = x_2^r = u_1^r = \frac{-a_1 x_1^r}{b_1}, \zeta_1^s = 0, \zeta^s = 0$ for positive gains $K_{P,i} > 0, K_{I,i} > 0$. See Fig. 3 for a block diagram representation of this control architecture.

Remark 5. *The feed-forward control inputs in* (18) *can also be implemented using the references* $x_1^r, x_2^r (= u_1^r)$ *instead of the states* x_1, x_2*, to compensate the plant dynamics. Then, the conditions for asymptotic stability of* (19) *are* $K_{P,i} > a_i$ *and* $K_{I,i} > 0$*. These controllers* (18) *may also not include any feed-forward compensation at all. Then, the conditions for asymptotic stability of* (19) *are* $a_i - b_i K_{P,i} < 0, K_{I,i} > 0$*. In either case, all our subsequent results hold with minor adjustments.*

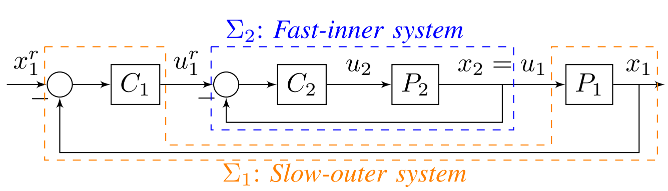


Fig. 2: Block diagram of a cascade control.

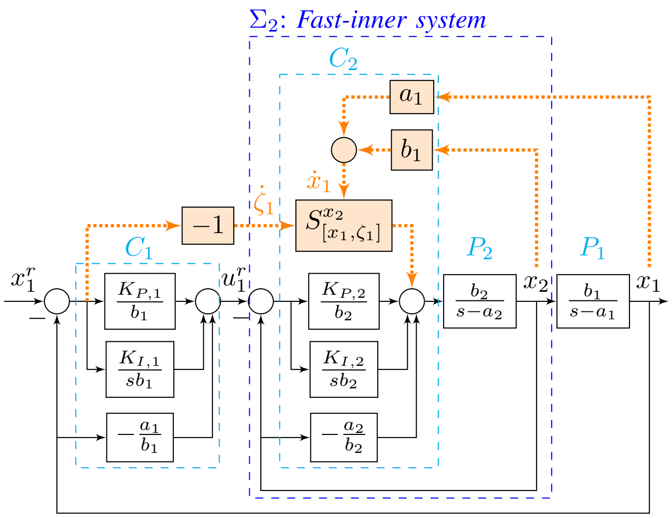


Fig. 3: Block diagram of cascade PI control (19), with plants P_i and controllers C_i in Laplace domain. The additional sensitivity-conditioning elements in (20) are depicted in orange blocks and dotted thicker arrows.

To preserve the stability of the time-scale separated cascaded system with controllers (19) in a single time scale we apply the sensitivity-conditioning (9):

$$\begin{aligned} \dot{x}_1 &= a_1 x_1 + b_1 x_2 \\ \dot{\zeta}_1 &= (x_1 - x_1^r) \\ x_2^r &= u_1^r = -\tfrac{1}{b_1}\left(a_1 x_1 + K_{P,1}(x_1 - x_1^r) + K_{I,1}\zeta_1\right) \\ \dot{x}_2 &= -K_{P,2}(x_2 - x_2^r) - K_{I,2}\zeta_2 + S^{x_2}_{[x_1,\zeta_1]} \begin{bmatrix} \dot{x}_1 \\ \dot{\zeta}_1 \end{bmatrix} \\ \dot{\zeta}_2 &= (x_2 - x_2^r) + S^{\zeta_2}_{[x_1,\zeta_1]} \begin{bmatrix} \dot{x}_1 \\ \dot{\zeta}_1 \end{bmatrix}, \end{aligned} \tag{20}$$

where

$$\begin{aligned} \begin{bmatrix} S^{x_2}_{[x_1,\zeta_1]} \\ S^{\zeta_2}_{[x_1,\zeta_1]} \end{bmatrix} &= -\begin{bmatrix} -K_{P,2} & -K_{I,2} \\ 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} K_{P,2} \\ -1 \end{bmatrix} \begin{bmatrix} -\frac{a_1 + K_{P,1}}{b_1} \\ -\frac{K_{I,1}}{b_1} \end{bmatrix}^T \\ &= \begin{bmatrix} -\frac{a_1+K_{P,1}}{b_1} & -\frac{K_{I,1}}{b_1} \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

Note that $S^{\zeta_2}_{[x_1,\zeta_1]} = [0\ 0]$ is due to $\zeta_2^s = 0$ for all x_1, ζ_1, for a stable inner system Σ_2. In summary, the control structure can be graphically represented as in Fig. 3, with the sensitivity-conditioning elements acting as a derivative-type control.

In compact matrix form, the closed-loop system reads as

$$\begin{aligned} \dot{x} &= TAx + Bx_1^r, \ x = [x_1{}^T, \zeta_1{}^T, x_2{}^T, \zeta_2{}^T]^T \\ A &= \begin{bmatrix} a_1 & 0 & b_1 & 0 \\ 1 & 0 & 0 & 0 \\ -K_{P,2}\frac{a_1+K_{P,1}}{b_1} & -K_{P,2}\frac{K_{I,1}}{b_1} & -K_{P,2} & -K_{I,2} \\ \frac{a_1+K_{P,1}}{b_1} & \frac{K_{I,1}}{b_1} & 1 & 0 \end{bmatrix} \\ T &= \begin{bmatrix} I & 0 \\ S^{[x_2,\zeta_2]}_{[x_1,\zeta_1]} & I \end{bmatrix}, \ T^{-1} = \begin{bmatrix} I & 0 \\ -S^{[x_2,\zeta_2]}_{[x_1,\zeta_1]} & I \end{bmatrix} \\ B &= \begin{bmatrix} 0 & -1 & K_{P,2}\frac{K_{P,1}}{b_1} & -\frac{K_{P,1}}{b_1} \end{bmatrix}^T. \end{aligned}$$

By means of the similarity transformation used in the proof of Proposition 2, we obtain

$$TA \sim T^{-1}(TA)T = \begin{bmatrix} -K_{P,1} & -K_{I,1} & \star & \star \\ 1 & 0 & \star & \star \\ 0 & 0 & -K_{P,2} & -K_{I,2} \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

where $\star$ are irrelevant terms for the following considerations. This block-companion form of TA confirms that the only conditions required for stability of (20) are $K_{P,i} > 0$ and $K_{I,i} > 0$, as for (19). This is aligned with Proposition 2: stability of (19) is preserved in (20) using the sensitivity-conditioning (9). Note that the system (20) without the sensitivity-conditioning term would have system matrix A instead of TA: $\dot{x} = Ax + Bx_1^r$. Then, for example with $a_i = 0, b_i = 1, K_{P,i} = 1 > 0, K_{I,i} = 1 > 0$, A has positive eigenvalues despite having positive control parameters, so stability is lost without the sensitivity-conditioning term.

Remark 6. *In cascade control the steady-state closed form $x_2^s(x_1)$ in* (7) *is typically known by design. Hence, we could use an alternative sensitivity based on the Lyapunov function $\|x_2 - x_2^s(x_1)\|^2_{P_2}$, see Remark 4. Therefore, in this cascade control example* (19)*, the sensitivity-conditioning approach* (9) *is equivalent to backstepping [11, Ch. 14] for nonlinear control design, up to an extra proportional control term.*

A. Numerical simulation: DC/AC-converter + RLC filter

Designing a DC/AC-converter connected to a RLC-filter [33] is a standard control problem in power electronics. Here we present a simplified version, where the DC/AC-converter modulates the DC voltage v_{dc} into the three-phase AC voltage v_m. Using an averaged converter with stiff DC voltage, v_m is a fully controllable voltage source. This modulated voltage v_m is then used to control the three-phase current i through the resistance R and inductance L, which in turn is used to control the output voltage v at the capacitor C to follow a reference v^r, see Fig. 4.

Let the electrical signals be represented in rectangular coordinates, using the real and imaginary parts, so $i, v, v_m \in \mathbb{R}^2$, and define the frequency ω and the rotation matrix $\mathcal{J} = \left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right]$. Then, according to Kirchoff's laws, the electrical signals dynamics in the rotation frame coordinates [5] are:

$$C\frac{dv}{dt} = i - \mathcal{J}\omega C v, \ L\frac{di}{dt} = v_m - (R + \mathcal{J}\omega L)i - v,$$

Following the cascaded PI example (19), the reference i^r and the controller v_m are chosen as PI controllers with every

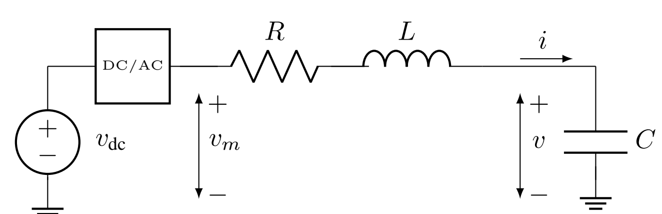


Fig. 4: DC/AC-converter with RLC circuit

TABLE II: Simulation parameters

RLC-filter	$R = 1m\Omega, L = 1mH, C = 300\mu F$
Frequency	$f = 50\frac{1}{\text{sec}},\ \omega = 2\pi 50\frac{\text{rad}}{\text{sec}}$
Outer Controller C_1	$k_{P,v} = 30\frac{A}{VF}, k_{I,v} = 0.3\frac{A}{VF}\frac{\text{rad}}{\text{sec}}$
Reference	Real and imaginary parts: $v^r_{\Re} = 120V,\ v^r_{\Im} = 0V$ Magnitude: $\lvert v^r\rvert = 120V = 1$ p.u. (per unit)
Black start	$v(0) = 0V = 0$ p.u., $i(0) = 0A = 0$ p.u.

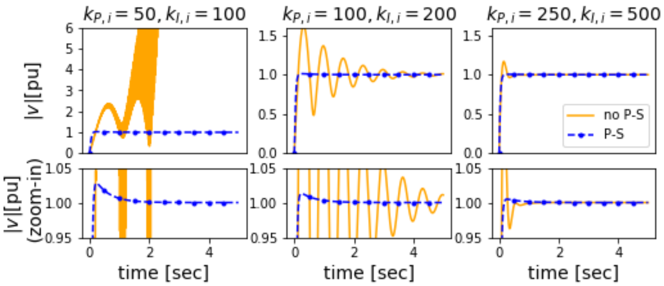


Fig. 5: Simulation of a black-start of the electrical system (21) with increasing values for the control parameters $k_{P,i}, k_{I,i}$, with and without sensitivity-conditioning (P-S).

$K_{(\cdot,\cdot)} = k_{(\cdot,\cdot)} I_2$, where I_2 is identity of dimension 2:

$$\begin{aligned} \tfrac{dv}{dt} &= C^{-1} i - \mathcal{J}\omega v \\ \tfrac{d\zeta_v}{dt} &= (v - v^r) \\ i^r &= \mathcal{J}\omega C v + C(-K_{P,v}(v - v^r) - K_{I,v}\zeta_v) \\ v_m &= (R + \mathcal{J}\omega L) i + v + L(-K_{P,i}(i - i^r) - K_{I,i}\zeta_i) \\ \tfrac{di}{dt} &= -K_{P,i}(i - i^r) - K_{I,i}\zeta_i \\ \tfrac{d\zeta_i}{dt} &= (i - i^r) \end{aligned} \tag{21}$$

Fig. 5 shows simulation results for (21) with the parameters in Table II. We simulate a black start, i.e., the system starts at time $t = 0$ with zero values, and needs to track a given constant reference, which can be interpreted as a step input. Since system (21) does not have an intrinsic time-scale separation, we use sufficiently large gains $k_{P,i}, k_{I,i}$ for the inner controller C_2 to artificially enforce a wider time-scale separation, i.e., lower ϵ in (3):

- $k_{P,i} = 50\frac{V}{AH}, k_{I,i} = 100\frac{V}{AH}\frac{\text{rad}}{\text{sec}}$: the system without sensitivity-conditioning is unstable even though the control parameters are all positive. On the contrary, the sensitivity-conditioning turns it into a stable system, with the magnitude $|v|$ stabilizing quickly at 1 p.u. and the frequency f at $50\frac{1}{\text{sec}}$.
- $k_{P,i} = 100\frac{V}{AH}, k_{I,i} = 200\frac{V}{AH}\frac{\text{rad}}{\text{sec}}$: the system without sensitivity-conditioning becomes stable, but it is still an unacceptable controller due to the high overshoot and relatively large settling time. Again the sensitivity-conditioning turns it into a well-performing controller.
- $k_{P,i} = 250\frac{V}{AH}, k_{I,i} = 500\frac{V}{AH}\frac{\text{rad}}{\text{sec}}$: the system without sensitivity-conditioning becomes stable with acceptable control performance. However, the sensitivity-conditioning approach performs much better with negligible overshoots.

Moreover, note in the zoom-in in Fig. 5 that the overshoot decreases as $k_{P,i}, k_{I,i}$ increase, so the system with sensitivity-conditioning also benefits from having a faster controller C_2. Such faster C_2 can further increase the convergence rate.

V. Example II: Bilevel Optimization

In this section, we show an application of the sensitivity-conditioning (9) to bilevel optimization [26]. As opposed to the cascade control example in Section IV, in this case the steady-state map $x_2^s(x_1)$ is not available in closed form, thus backstepping is not applicable, see Remarks 4 and 6. Yet, the sensitivity-conditioning (9) can still be used to preserve the stability of the two-time-scale system (7) in a single one.

Consider a general unconstrained bilevel problem [26], [34]:

$$\begin{aligned} &\min_{x_1, x_2^*} F_1(x_1, x_2^*) \\ &\text{s.t. } x_2^* \in \arg\min_{x_2} F_2(x_1, x_2) \end{aligned} \tag{22}$$

where $F_1(\cdot), F_2(\cdot)$ are the upper- and lower-level objective functions, respectively.

Assumption 4 (Adaptation of Assumptions 1 and 2)**.**

- *The functions $F_1(\cdot)$ and $F_2(\cdot)$ are twice and thrice continuously differentiable, respectively, with Lipschitz continuous partial derivatives.*
- *For every x_1, the lower-level problem $\arg\min_{x_2} F_2(x_1, x_2)$ has at most a single solution x_2^*, where the second-order partial derivative $\nabla^2_{x_2 x_2} f_2(x_1, x_2^*)$ is invertible.*

The single solution assumption is often used as simplification in bilevel problems [26]. It ensures that (22) is well-posed, and it allows to simplify the constraint to $x_2^* = \arg\min_{x_2} F_2(x_1, x_2)$. Moreover, the invertibility assumption allows to define the sensitivity of x_2^*, a known concept in bilevel optimization [35], similar to the one introduced in (6): Consider a point (x_1, x_2^*) satisfying the first-order optimality conditions of the lower-level problem, i.e., $\nabla_{x_2} F_2(x_1, x_2^*) = 0$. Since $\nabla^2_{x_2 x_2} F_2(x_1, x_2^*)$ is invertible, the implicit function theorem [28] guarantees the local existence of the map $x_2^*(x_1)$, and gives an expression for its derivative:

$$\nabla_{x_1} x_2^*(x_1) = -\left(\nabla^2_{x_2 x_2} F_2(x_1, x_2^*(x_1))\right)^{-1} \nabla^2_{x_2 x_1} F_2(x_1, x_2^*(x_1))$$

Additionally, $x_2^*(x_1)$ can be used to locally define a reduced objective F_1: $F_1^r(x_1) := F_1(x_1, x_2^*(x_1))$, as in (7), and use $\nabla_{x_1} x_2^*(x_1)$ to give an expression for the *total derivative*,

$$\begin{aligned} &D_{x_1} F_1(x_1, x_2^*(x_1)) := \nabla_{x_1} F_1^r(x_1) \\ &= \nabla_{x_1} F_1(x_1, x_2^*(x_1)) + \nabla_{x_1} x_2^*(x_1)^T \nabla_{x_2} F_2(x_1, x_2^*(x_1)) \end{aligned} \tag{23}$$

defined for points where $x_2 = x_2^*(x_1)$. Under Assumption 4, $\nabla^2_{x_2 x_2} f_2(x_1, x_2)$ is invertible for x_2 in a neighborhood of $x_2^s(x_1)$. Hence, $\nabla_{x_1} x_2^*(x_1)$ and $\nabla_{x_1} F_1^r(x_1)$ can be extended to these x_2, as the extended sensitivity in (6):

$$\begin{aligned} D_{x_1} F_1(x_1, x_2) &:= \nabla_{x_1} F_1(x_1, x_2) + S_{x_1}^{x_2}(x_1, x_2)^T \nabla_{x_2} F_1(x_1, x_2) \\ S_{x_1}^{x_2}(x_1, x_2) &:= -(\nabla^2_{x_2 x_2} F_2(x_1, x_2))^{-1} \nabla^2_{x_2 x_1} F_2(x_1, x_2), \end{aligned} \tag{24}$$

satisfying the restrictions $D_{x_1} F_1(x_1, x_2)_{|(x_1, x_2^*(x_1))} = \nabla_{x_1} F_1^r(x_1)$, $S_{x_1}^{x_2}(x_1, x_2)_{|(x_1, x_2^*(x_1))} = \nabla_{x_1} x_2^*(x_1)$.

A. Bilevel local solutions

Understanding the properties of the bilevel problem solutions is essential to connect the convergence of algorithms with the stability of steady states from previous sections. Therefore, we recall the concept of *local solutions* in [26, Ch. 8] to represent locals minima of (22) and their first and second-order necessary and sufficient conditions:

Definition 1 (local solution)**.** *[26, Ch. 8] A point (x_1^*, x_2^*) is a (strict) local solution of* (22) *if:*

1) The point x_2^* is a local minimum of $F_2(x_1^*, \cdot)$ with fixed x_1^*.
2) There exists a neighborhood N of (x_1^*, x_2^*) such that $F_1(x_1^*, x_2^*) \leq F_1(x_1, x_2)$ ($<$ for strict) for all $(x_1, x_2) \in N$ such that x_2 is a local minimum of $F_2(x_1, \cdot)$ with fixed x_1.

Proposition 6 (First-order necessary conditions)**.** *[26, Ch. 8] A local solution (x_1^*, x_2^*) is a stationary point, i.e., it satisfies $\nabla_{x_2} F_2(x_1^*, x_2^*) = 0,\ \nabla_{x_1} F_1^r(x_1^*) = 0$.*

Proposition 7 (Second-order conditions)**.** *[26, Ch. 8]*

- *Necessary conditions: A local solution (x_1^*, x_2^*) satisfies*
$$\nabla^2_{x_2 x_2} F_2(x_1^*, x_2^*) \succeq 0,\ \nabla^2_{x_1 x_1} F_1^r(x_1^*) \succeq 0$$
- *Sufficient conditions: A stationary point (x_1^*, x_2^*) satisfying*
$$\nabla^2_{x_2 x_2} F_2(x_1^*, x_2^*) \succ 0,\ \nabla^2_{x_1 x_1} F_1^r(x_1^*) \succ 0, \tag{25}$$
is a strict local solution.

B. Bilevel gradient flow

The steepest descent direction method [36] is a standard approach to iteratively solve (22). It follows the negative gradient of $F_1^r(x_1)$ with step size α^k in each iteration k:

$$\begin{aligned} x_1^{k+1} &= x_1^k - \alpha^k \nabla_{x_1} F_1^r(x_1^k) = x_1^k - \alpha^k \big(D_{x_1} F_1(x_1^k, x_2^k)\big) \\ x_2^k &= \arg\min_{x_2} F_2(x_1^k, x_2) \end{aligned} \tag{26}$$

If the lower-level update $x_2^k = \arg\min_{x_2} F_2(x_1^k, x_2)$ is not available in closed form, it can be solved iteratively using for example gradient descent with step size β^l and updates:

$$x_2^{l+1} = x_2^l - \beta^l \nabla_{x_2} F_2(x_1^k, x_2^l)$$

The corresponding continuous-time version of this bilevel gradient descent (26) can be represented on two time scales with $\epsilon \to 0$ and the singular perturbation interconnection (3):

$$\begin{aligned} \dot{x}_1 &= -D_{x_1} F_1(x_1, x_2) \\ \epsilon \dot{x}_2 &= -\nabla_{x_2} F_2(x_1, x_2) \end{aligned} \tag{27}$$

As mentioned before in Section II, these nested iterations (26) on two time scales (27) may slow down the algorithm convergence. On the other hand, the sensitivity-conditioning system (9) yields:

$$\begin{aligned} \dot{x}_1 &= -D_{x_1} F_1(x_1, x_2) \\ \dot{x}_2 &= -\nabla_{x_2} F_2(x_1, x_2) + S_{x_1}^{x_2}(x_1, x_2)\, \dot{x}_1 \end{aligned} \tag{28}$$

Corollary 4 (Local convergence of (28))**.** *A point (x_1^*, x_2^*) is a strict local solution of* (22) *satisfying the sufficient conditions in* (25) *if and only if it is a locally exponentially stable steady state of the sensitivity-conditioning bilevel gradient flow* (28)*.*

Proof. Note that Assumption 4 adapts Assumptions 1 and 2 for the bilevel problem (22), and the second-order total derivative is symmetric and satisfies

$$\begin{aligned} &D^2_{x_1 x_1} F_1(x_1, x_2^*(x_1)) := \nabla^2_{x_1 x_1} F_1^r(x_1) = \\ &\nabla_{x_1} D_{x_1} F_1(x_1, x_2^*(x_1)) + \nabla_{x_1} x_2{}^*(x_1)^T \nabla_{x_2} D_{x_1} F_1(x_1, x_2^*(x_1)) \end{aligned}$$

Hence, (x_1^*, x_2^*) is locally exponentially stable if and only if (25) holds [11, Cor. 4.3]. ■

Remark 7. *This convergence result can be stated to larger regions if similar conditions as in Proposition 3 hold.*

C. Time discretization and numerical simulation

The Euler-forward method [37] with time constant τ can be used to integrate the differential equations (28) and (27) for a fixed ϵ. Then we get a discrete-time descent algorithm:

$$\begin{aligned} & x_1^{k+1} = x_1^k - \tau D_{x_1} F_1(x_1^k, x_2^k) \\ (27):\ & x_2^{k+1} = x_2^k - \tfrac{\tau}{\epsilon} \nabla_{x_2} F_2(x_1^k, x_2^k) \\ (28):\ & x_2^{k+1} = x_2^k - \tau \nabla_{x_2} F_2(x_1^k, x_2^k) \\ & \qquad + S_{x_1}^{x_2}(x_1^k, x_2^k)(-\tau D_{x_1} F_1(x_1^k, x_2^k)), \end{aligned} \tag{29}$$

where the time constant τ plays the role of the step size in optimization [16]. If (28) is locally or globally exponentially stable, see Proposition 2 and 3, then under some conditions its Euler-forward discretization will retain this exponential stability for suitable time constants below a certain threshold $\tau < \bar{\tau}$ [38], [39]. More concretely, this can be proven by extending [40, Lemma 5] to the Krasovskii Lyapunov functions in Proposition 3. We will formalize further results for discrete-time sensitivity-conditioning in Subsection VI-C.

Remark 8. *Bilevel optimization problems like* (22) *can also be represented as Stackelberg games [41]. The singular perturbation dynamics* (27)*, with its discrete-time version in* (29)*, can be interpreted as a simultaneous gradient descent on both variables with different step sizes: $\tau, \frac{\tau}{\epsilon}$. This corresponds to deterministic Stackelberg learning dynamics [42], [43].*
The particular case when $F_2(x_1, x_2) = -F_1(x_1, x_2)$ in (22)*, is called a zero-sum or minimax game [42], [44]. In this context, this simultaneous gradient descent algorithm is known as the γ-gradient descent ascent (γ-GDA) [44], with $\gamma = \frac{1}{\epsilon}$. Then, the discrete-time sensitivity-conditioning application for bilevel optimziation in* (29)*, corresponds to the Stackelberg generalization of the algorithm in [23] for minimax games.*

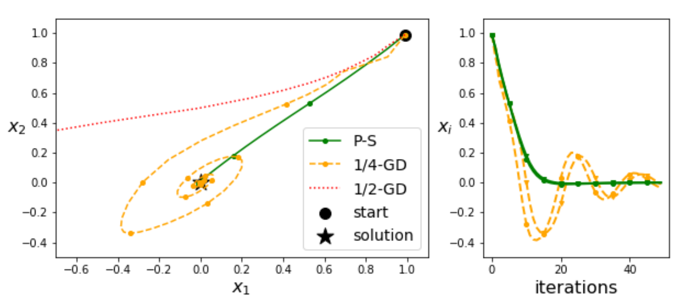


Fig. 6: Comparison of discrete-time methods (29) for the functions in (30): a simultaneous gradient descent (27) for two values $\epsilon \in \{\frac{1}{4}, \frac{1}{2}\}$ (ϵ-GD), against the sensitivity-conditioning (P-S) approach (28). Both are implemented in discrete time with markers every 5 iterations.

Consider an example similar to the ones in [23], [42]:

$$F_1(x_1,x_2) = -\frac{x_1^2}{2} + x_2^2,\ F_2(x_1,x_2) = \Big(\frac{x_2^2}{4} - \frac{x_1 x_2}{2}\Big)e^{-\frac{x_2^2}{2}} \tag{30}$$

The point $(0,0)$ is a *strict local solution* of the bilevel problem (22), since $\nabla_{x_2}F_2(0,0)=0, D_{x_1}F_1(0,0)=0, \nabla^2_{x_2x_2}F_2(0,0)=\frac{1}{2}>0, D^2_{x_1x_1}F_1(0,0)=1>0$.

In Fig. 6, we compare the approaches in (29) using $\tau=\frac{1}{4}$, singular perturbation values $\epsilon\in\{\frac{1}{4},\frac{1}{2}\}$, and the functions in (30). First, we observe that the sensitivity-conditioning (P-S) approach (28) is able to converge quickly to the solution $(0,0)$. For the case $\epsilon=\frac{1}{2}$, the simultaneous gradient descent (ϵ-GD) based on (27) fails to converge. It converges for $\epsilon=\frac{1}{4}$, but still oscillates around the solution causing a slower convergence.

VI. Sensitivity-conditioning for Multiple Systems

The theoretical results in Section III, and the applications to cascade control and bilevel optimization, Sections IV and V respectively, deal with two-time-scale systems. In this section we show how the sensitivity-conditioning interconnection (9) can be extended to multiple time scales arising in, e.g., multiple nested systems [12] or multilevel programming [45]. Consider N differential-algebraic-equation subsystems Σ_i, with states $x_i\in\mathbb{R}^{n_i}$ and vector fields $f_i(\cdot)$, operating on different time scales τ_i as in (2), ordered from slow Σ_1 to fast Σ_N:

$$\Sigma_i : \frac{dx_i}{d\tau_i} = f_i(x_1,\dots,x_N) \text{ s.t. } \begin{array}{l} f_j(x_1,\dots,x_N)=0\ \forall j>i \\ \frac{dx_j}{d\tau_i}=0\ \forall j<i \end{array} \tag{31}$$

As in (3), consider the corresponding singular-perturbed system with terms $0<\epsilon_N\ll\cdots\ll\epsilon_2\ll\epsilon_1=1$ [12]:

$$\begin{bmatrix}\epsilon_1 I & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \epsilon_N I\end{bmatrix}\begin{bmatrix}\dot{x}_1\\ \vdots\\ \dot{x}_N\end{bmatrix} = \begin{bmatrix} f_1(x_1,\dots,x_N)\\ \vdots \\ f_N(x_1,\dots,x_N)\end{bmatrix},$$

where the time-scale separation in (31) is recovered for $\tau_i=\frac{t}{\epsilon_i}$ in the *singular limit* $\frac{\epsilon_{i+1}}{\epsilon_i}\to 0,\ \forall i=1,\dots,N-1$.

A. Steady states, sensitivities and total derivatives

Assume that every subsystem Σ_i in (31) has isolated steady states, and that the implicit function theorem [28] can be to guarantee the existence of steady state maps $x_i^s(x_1,\dots,x_{i-1})$ recursively from fast to slow subsystems Σ_i: First, for some x_N such that $0=f_N(x_1,\dots,x_N)$ for the fastest Σ_N, the implicit function theorem guarantees the local existence of $x_N^s(x_1,\dots,x_{N-1})$. Under time-scale separation, the next system Σ_{N-1} has reduced-order dynamics $\dot{x}_{N-1}=f^r_{N-1}(x_1,\dots,x_{N-1}):=f_{N-1}(x_1,\dots,x_{N-1},x_N^s(x_1,\dots,x_{N-1}))$, which allow now to define the steady state map $x^s_{N-1}(x_1,\dots,x_{N-2})$, and recursively $x_i^s(x_1,\dots,x_{i-1})$, which depend only on the states of the slower systems $x_1,\dots,x_{i-1}$. To ease the notation, from now on we will use x_i^s instead of $x_i^s(x_1,\dots,x_{i-1})$ to denote the steady state map and its dependencies. Similarly, we use $(x_1,\dots,x_i,x^s_{i+1},\dots,x^s_N)$ to denote that $x^s_{i+1},\dots,x^s_N$ are all at steady-state for given values $x_1,\dots,x_i$. As in (7), the steady state maps allow to define the reduced-order dynamics:

$$\frac{dx_i}{d\tau_i} = f_i^r(x_1,\dots,x_i) := f_i(x_1,\dots,x_i,x^s_{i+1},\dots,x^s_N) \tag{32}$$

Under sufficient regularity, the implicit function theorem gives the sensitivity of each steady state x_i^s with respect to any x_j for $j<i$:

$$\begin{aligned}&\nabla_{x_j}x_i^s(x_1,\dots,x_{i-1})\\ &=-(\nabla_{x_i}f_i^r(x_1,\dots,x_{i-1},x_i^s))^{-1}\nabla_{x_j}f_i^r(x_1,\dots,x_{i-1},x_i^s)\end{aligned} \tag{33}$$

Now the concepts of extended sensitivities (6) and extended total derivatives (24) can be used to define compact analytical expressions for these sensitivities $\nabla_{x_j}x_i^s$ and an extension for a general point $(x_1,\dots,x_N)$. We define the extended total derivatives and sensitivities for every i and j recursively from N to 1:

$$\begin{aligned} D_{x_N}f_i &:= \nabla_{x_N}f_i\\ D_{x_j}f_i &:= \nabla_{x_j}f_i + \textstyle\sum_{k=\max(i,j)+1}^N D_{x_k}f_i S^{x_k}_{x_j}\\ S^{x_i}_{x_j} &:= -(D_{x_i}f_i)^{-1}D_{x_j}f_i,\ \forall j<i,\end{aligned} \tag{34}$$

where for clarity we omit the evaluation at $(x_1,\dots,x_N)$. These extended total derivatives of f_i with respect to x_j take into account the dependency of each intermediate x_k^s, $k\ge\max(i,j)+1$, with respect to x_j. When restricted, they coincide with the total derivatives (24) of $f_i^r(x_1,\dots,x_i)$, and the steady-state sensitivities (33):

$$\begin{aligned} D_{x_j}f_i(x_1,\dots,x_N)_{|(x_1,\dots,x_i,x^s_{i+1},\dots,x^s_N)} &= \nabla_{x_j}f_i^r(x_1,\dots,x_i)\\ S^{x_i}_{x_j}(x_1,\dots,x_N)_{|(x_1,\dots,x_i,x^s_{i+1},\dots,x^s_N)} &= \nabla_{x_j}x_i^s(x_1,\dots,x_{i-1})\end{aligned} \tag{35}$$

To guarantee that the implicit function theorem is applicable for every $f_i(x_1,\dots,x_i,x^s_{i+1},\dots,x^s_N)=0$, and thus ensure that steady states x_i^s, reduced-order dynamics (32), sensitivities (III-A), and extended sensitivities and total derivatives (34) are well-defined around steady states, we formalize the assumptions made in this section in the following one:

Assumption 5 (Extension of Assumption 1)**.** *For all $i>1$, the vector fields $f_i(\cdot), f_i^r(\cdot)$ are continuously differentiable, and the reduced-order systems* (32) *have isolated steady-states x_i^s, where the partial derivatives $\nabla_{x_i}f_i^r(x_1,\dots,x_{i-1},x_i^s)$ are invertible.*

This assumption also implies that the total derivatives $D_{x_i}f_i(\cdot)$ in (34) are invertible for $(x_i,\cdots,x_N)$ in a neighborhood of $(x_i^s,\cdots,x_N^s)$. Note that in contrast to Assumption 1, Assumption 5 relaxes the need for a single steady-state in each subsystem, since there may exist multiple steady state x_i^s for every subsystem Σ_i given the values values $x_1,\dots,x_{i-1}$, i.e., the set of steady states

$$\mathcal{X}_i^s(x_1,\dots,x_{i-1}) := \{x_i | f_i(x_1,\dots,x_i,x^s_{i+1},\dots,x^s_N)=0\}$$

is not necessarily a singleton. However, given the locally invertible $D_{x_i}f_i(\cdot)$, steady states will be isolated points. Moreover, given the implicit function theorem, the extended sensitivity $S^{x_i}_{x_j}(x_1,\dots,x_{i-1},x_i^s,\dots,x_N^s)$ (34) evaluated at each $x_i^s\in\mathcal{X}_i^s$ gives the actual sensitivity of each x_i^s with respect to $x_1, \dots, x_{i-1}$. Therefore, Theorem 1 presented later will allow to test the local stability of every combination of steady states $(x_1^s, \dots, x_N^s)$, where $x_i^s \in \mathcal{X}_i^s$ for every i.

B. Sensitivity-conditioning for multiple time scales

With the previously defined extended total derivatives and sensitivities (34), the sensitivity-conditioning system (9) from Section III can be extended and applied to the multiple-time-scales system (31):

$$\begin{bmatrix} I & 0 & 0 & 0 \\ -S_{x_1}^{x_2} & I & 0 & 0 \\ \vdots & \ddots & I & 0 \\ -S_{x_1}^{x_N} & \cdots & -S_{x_{N-1}}^{x_N} & I \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \vdots \\ \dot{x}_N \end{bmatrix} = \begin{bmatrix} f_1(x_1, \dots, x_N) \\ \vdots \\ f_N(x_1, \dots, x_N) \end{bmatrix}, \tag{36}$$

where for clarity we omit the evaluation at point $(x_1, \dots, x_N)$ in the conditioning matrix M. In the equivalent expression

$$\Sigma_i : \; \dot{x}_i = f_i(x_1, \dots, x_N) + \textstyle\sum_{j=1}^{i-1} S_{x_j}^{x_i}(x_1, \dots, x_N)\dot{x}_j, \tag{37}$$

the extra terms $\sum_{j=1}^{i-1} S_{x_j}^{x_i}(x_1, \dots, x_N)\dot{x}_j$ play again the role of predicting and anticipating the changes of steady states x_i^s due to the slower dynamics $\dot{x}_j$, $j < i$. Note that our approach (36) does not requires any specific cascaded structure.

As in Assumption 2, to guarantee local existence and uniqueness [11, Thm. 3.1] of a solution for (36), we assume:

Assumption 6. *The vector fields in* (36),(37) *are locally Lipschitz continuous.*

Then, the statements in Proposition 1 to 3 for the two systems sensitivity-conditioning (9), can be extended to the multiple time-scale case sensitivity-conditioning (36):

Theorem 1 (Extension of Propositions 1,2 and Corollary 2)**.**

1) ***Positive invariance****: For some $p \geq 1$, given the dynamics* (37) *of $\dot{x}_i$ for $i \geq p$, initialized at time t_0: $x_i(t_0) = x_i^s(x_1(t_0), \dots, x_{i-1}(t_0))$. Then, $x_i(t) = x_i^s(x_1(t), \dots, x_{i-1}(t))$ $\forall i \geq p$ is the unique solution on the open domain of existence.*
2) ***Local stability****: At a steady state $(x_1^s, \dots, x_N^s)$ the Jacobian J of* (36) *satisfies:*

$$J \sim \begin{bmatrix} \nabla_{x_1} f_1^r(x_1^s) & \star & \star & \star \\ 0 & \nabla_{x_2} f_2^r(x_1^s, x_2^s) & \star & \star \\ 0 & 0 & \ddots & \star \\ 0 & 0 & 0 & \nabla_{x_N} f_N(x_1^s, \dots, x_N^s) \end{bmatrix}, \tag{38}$$

where $\star$ are irrelevant terms.

3) ***Global exponential stability****: Assume that the vector fields $f_i(\cdot)$ are globally Lipschitz continuous, that the total derivatives $D_{x_i} f_i(\cdot)$ in* (34) *are globally invertible, and that there exists positive definite matrices $P_i \succ 0$ and η_i such that for all $(x_1, \dots, x_N)$ the following contraction condition holds:* [1]

$$P_i D_{x_i} f_i + D_{x_i} f_i^T P_i^T \preceq -\eta_i P_i. \tag{39}$$

Then there exists a unique steady state $(x_1^s, \dots, x_N^s)$, which is a globally exponentially stable steady state of the sensitivity-conditioning approach (36)*.*

Proof. See Appendix II ∎

Corollary 5 (of Theorem 1.2, extending Corollary 1)**.** *The Jacobians* (35) *and* (38)*, for the systems under time-scale separation* (32) *and the sensitivity-conditioning* (36)*, respectively, have the same eigenvalues, and thus the same local stability.*

The multiple time-scale sensitivity-conditioning (36) can also be generalized as (16), to improve the performance. Theorem 1 can then be extended as in Proposition 4 and 5.

C. Discrete-time sensitivity-conditioning

Here we show how the multiple time-scale sensitivity-conditioning (36) can be extended to discrete-time systems, while preserving the local stability result in Theorem 1. Consider the discrete-time systems:

$$\Sigma_i : x_i^{k+1} = x_i^k + f_i(x_1^k, \dots, x_N^k),$$

where x_i^k denote the value of x_i at time t^k. Under time-scale separation as in (32), we can represent each discrete-time subsystem Σ_i in its own time scale k_i with reduced-order dynamics:

$$\begin{aligned} x_i^{k_i+1} &= x_i^{k_i} + f_i^r(x_1^{k_i}, \dots, x_{i-1}^{k_i}, x_i^{k_i}) \\ &:= x_i^{k_i} + f_i(x_1^{k_i}, \dots, x_{i-1}^{k_i}, x_i^{k_i}, x_{i+1}^s, \dots, x_N^s) \end{aligned} \tag{40}$$

where $x_j^{k_i+1} = x_j^{k_i}$ for $j < i$, and x_j^s is the steady state of x_j given $x_1^{k_i}, \dots, x_{i-1}^{k_i}, x_i^{k_i}, x_{i+1}^s, \dots, x_{j-1}^s$, for $j > i$. Similar to (35) the Jacobians of (40) are

$$J_i = I + \nabla_{x_i} f_i^r(x_1^s, \dots, x_i^s) = I + D_{x_i} f_i(x_1^s, \dots, x_N^s)$$

and thus systems (40) are locally asymptotically stable if the eigenvalues λ_i of $I + \nabla_{x_i} f_i^r$ satisfy $|\lambda_i| < 1$.

Using the same extended sensitivities $S_{x_j}^{x_i}$ as in (34), the discrete-time version of the sensitivity-conditioning (36) can be expressed as:

$$\begin{bmatrix} I & 0 & 0 & 0 \\ -S_{x_1}^{x_2} & I & 0 & 0 \\ \vdots & \ddots & I & 0 \\ -S_{x_1}^{x_N} & \cdots & -S_{x_{N-1}}^{x_N} & I \end{bmatrix} \begin{bmatrix} x_1^{k+1} - x_1^k \\ \vdots \\ x_N^{k+1} - x_N^k \end{bmatrix} = \begin{bmatrix} f_1 \\ \vdots \\ f_N \end{bmatrix}, \tag{41}$$

omitting for clarity the evaluation at the point $(x_1^k, \dots, x_N^k)$.

Proposition 8. *At the steady state $(x_1^s, \dots, x_N^s)$ the Jacobians of* (40) *and* (41) *have the same eigenvalues. Thus,* (40) *is locally exponentially stable if and only if* (41) *is so. Moreover, if for some i the Jacobian $I + \nabla_{x_i} f_i^r$ has any eigenvalue with norm larger than one, then both* (40) *and* (41) *are unstable.*

Proof. The proof follows similar steps as the local stability one in Theorem 1: After performing the same similarity transformations, the Jacobian J of (41) satisfies

$$J \sim I + \begin{bmatrix} \nabla_{x_1} f_1^r(x_1^s) & \star & \star \\ 0 & \ddots & \star \\ 0 & 0 & \nabla_{x_N} f_N(x_1^s, \dots, x_N^s) \end{bmatrix}$$

∎

[1] This would correspond to the condition $P_i \nabla_{x_i} f_i^r + (\nabla_{x_i} f_i^r)^T P_i^T \preceq -\eta_i P_i$ in Corollary 2 for the two time-scale system. However, here it is required to hold for an extended number of points $(x_1, \dots, x_N)$, not just for only $(x_1, \dots, x_i, x_{i+1}^s, \dots, x_N^s)$, hence it is more strict.

VII. Conclusion and Outlook

In this work, we have presented the sensitivity-conditioning: an alternative design tool for interconnected systems, that uses a predictive feed-forward term to preserve the stability of the system analysed at different time scales. This approach does not introduce a lower threshold on the actual time-scale separation between subsystems, in contrast to the usual singular perturbation approach. Moreover, we have shown examples of control design problems and optimization algorithms where our approach can be directly applied and improves the performance compared to a time-scale separation approach.

We believe that the applicability of our approach is not limited to these examples, but has the potential to be used in many other applications. For example, for any nested algorithms (e.g. in optimization or adaptive control) this sensitivity-conditioning could be used to design faster algorithms avoiding the need of time-scale separation between nested iterations. This is particularly promising for cases where iterations are computationally expensive, even when they are simple to evaluate, for example in distributed algorithms with communication bottlenecks.

Several directions for future research remain open: Since the sensitivities employed in the proposed conditioning are heavily model-based, we have established input-to-state stability robustness analysis against model errors. Nonetheless, a more sophisticated robust performance guarantees would be desirable. Moreover, through this work we have considered only continuously differentiable vector fields driving the dynamics. Thus, it remains to be seen how this method could be extended to nondifferentiable cases, arising often in optimization.

Appendix I
Proof of Lemma 1

For clarity we omit the evaluation at a given x. Consider the singular value decomposition $\nabla_x f = USV^T$. Let $\sigma_{\min} = \min_i S_{i,i}$ denote the minimum singular value of $\nabla_x f$, and $v_{\min}, u_{\min}$ the columns of V, U corresponding to $\sigma_{\min}$. Since $P \succ 0$ and $\|v_{\min}\|_2 = \|u_{\min}\|_2 = 1$, we have

$$\begin{aligned} 0 < \eta\lambda_{\min}(P) &\leq \eta v_{\min}^T P v_{\min} \\ &\leq -v_{\min}^T (P\nabla_x f + \nabla_x f^T P^T) v_{\min} \\ &= -2\sigma_{\min} v_{\min}^T P u_{\min} \\ &\leq 2\sigma_{\min}\|v_{\min}\|_2\|u_{\min}\|_2\lambda_{\max}(P) = 2\sigma_{\min}\lambda_{\max}(P), \end{aligned}$$

where the third inequality is due to $P\nabla_x f(x) + \nabla_x f(x)^T P^T \preceq -\eta P \ \forall x$. Thus, $\nabla_x f^{-1} = VS^{-1}U^T$ is well-defined and

$$\|\nabla_x f^{-1}\|_2 = \max_i \tfrac{1}{S_{i,i}} = \tfrac{1}{\sigma_{\min}} \leq \tfrac{2\lambda_{\max}(P)}{\eta\lambda_{\min}(P)}.$$

Next, with $x(\delta) = x^s + \delta(x - x^s)$ for $\delta \in [0,1]$, then

$$\begin{aligned} &(x - x^s)^T P\big(f(x) - \overbrace{f(x^s)}^{=0}\big) + \big(f(x) - \overbrace{f(x^s)}^{=0}\big)^T P(x - x^s) \\ &= \textstyle\int_0^1 (x - x^s)^T \big(\underbrace{P\nabla_x f(x(\delta)) + \nabla_x f(x(\delta)^T)P}_{\preceq -\eta P,\ \text{since } x(\delta)\in\mathcal{B}_r(x^s)}\big)(x - x^s)d\delta \\ &\leq -\eta\|x - x^s\|_P^2 \leq 0. \end{aligned}$$

Hence, $2\|P^{\frac{1}{2}} f(x)\|_2 = 2\|f(x)\|_P \|x - x^s\|_P \geq \eta\|x - x^s\|_P^2$, and $\|f(x)\|_P \geq \frac{\eta}{2}\|x - x^s\|_P$.

Appendix II
Proof of Theorem 1

1) For any $i \geq p$, consider the dynamics of $\dot{x}_i$ in (37), initialized at time t_0 as $x_i(t_0) = x_i^s(x_1(t_0), \ldots, x_{i-1}(t_0))$. Local existence and uniqueness of a solution is guaranteed by Assumption 6. Then, $x_i(t) = x_i^s(x_1(t), \ldots, x_{i-1}(t))$ is the unique solution on the open domain of existence, since the derivative

$$\begin{aligned} &\dot{x}_i(t) \\ &= f_i(x_1(t), \ldots, x_N(t)) + \textstyle\sum_{j=1}^{i-1} S_{x_j}^{x_i}(x_1(t), \ldots, x_N(t))\dot{x}_j(t) \\ &= \cancel{f_i(x_1(t), \ldots, x_{i-1}(t), x_i^s(t), \ldots, x_N^s(t))} \\ &\quad + \textstyle\sum_{j=1}^{i-1} \underbrace{S_{x_j}^{x_i}(x_1(t), \ldots, x_{i-1}(t), x_i^s(t), \ldots, x_N^s(t))}_{\overset{(35)}{=}\nabla_{x_j} x_i^s(x_1(t),\ldots,x_{i-1}(t))} \dot{x}_j(t) \\ &= \tfrac{dx_i^s(x_1(t),\ldots,x_{i-1}(t))}{dt} \end{aligned}$$

and the initial conditions coincide for $t \geq t_0$ on the domain of existence.

2) By recursively using the expression in (37), the sensitivity-conditioning system (36) can be rewritten as

$$\begin{aligned} \begin{bmatrix} \dot{x}_1 \\ \vdots \\ \dot{x}_N \end{bmatrix} &= \begin{bmatrix} I & 0 \\ \begin{bmatrix} S_{x_1}^{x_N} \cdots S_{x_{N-1}}^{x_N} \end{bmatrix} & I \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \vdots \\ \dot{x}_{N-1} \\ f_N \end{bmatrix} \\ &= \begin{bmatrix} I & 0 \\ \begin{bmatrix} S_{x_1}^{x_N} \cdots S_{x_{N-1}}^{x_N} \end{bmatrix} & I \end{bmatrix} \cdots \begin{bmatrix} I & 0 & 0 \\ S_{x_1}^{x_2} & I & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} f_1 \\ \vdots \\ f_N \end{bmatrix}, \end{aligned}$$

where the matrices $\begin{bmatrix} S_{x_1}^{x_i} \cdots S_{x_{i-1}}^{x_i} \end{bmatrix}$ are at the row block i. Since at a steady state $(x_1^s, \ldots, x_N^s)$ we have $f_i(x_1^s, \ldots, x_N^s) = 0$, then the Jacobian of (36) at $(x_1^s, \ldots, x_N^s)$ simplifies to

$$J = \begin{bmatrix} I & 0 \\ \begin{bmatrix} S_{x_1}^{x_N} \cdots S_{x_{N-1}}^{x_N} \end{bmatrix} & I \end{bmatrix} \cdots \begin{bmatrix} I & 0 & 0 \\ S_{x_1}^{x_2} & I & 0 \\ 0 & 0 & I \end{bmatrix} \nabla_x f,$$

where the i,j-block of $\nabla_x f$ is $(\nabla_x f)_{i,j} = \nabla_{x_j} f_i$.

Since $\begin{bmatrix} I & 0 & 0 \\ -\begin{bmatrix} S_{x_1}^{x_i} \cdots S_{x_{i-1}}^{x_i} \end{bmatrix} & I & 0 \\ 0 & 0 & I \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 & 0 \\ \begin{bmatrix} S_{x_1}^{x_i} \cdots S_{x_{i-1}}^{x_i} \end{bmatrix} & I & 0 \\ 0 & 0 & I \end{bmatrix} \ \forall i$, multiplying J on both sides by these matrices we can iteratively construct matrices similar to J:

$$\begin{aligned} J \sim \tilde{J}_N &= \begin{bmatrix} I & 0 \\ -\begin{bmatrix} S_{x_1}^{x_N} \cdots S_{x_{N-1}}^{x_N} \end{bmatrix} & I \end{bmatrix} J \begin{bmatrix} I & 0 \\ \begin{bmatrix} S_{x_1}^{x_N} \cdots S_{x_{N-1}}^{x_N} \end{bmatrix} & I \end{bmatrix} \\ &= \begin{bmatrix} I & 0 & 0 \\ \begin{bmatrix} S_{x_1}^{x_{N-1}} \cdots S_{x_{N-2}}^{x_{N-1}} \end{bmatrix} & I & 0 \\ 0 & 0 & I \end{bmatrix} \cdots \begin{bmatrix} I & 0 & 0 \\ S_{x_1}^{x_2} & I & 0 \\ 0 & 0 & I \end{bmatrix} \\ &\quad \begin{bmatrix} \begin{bmatrix} (\nabla_{x_j} f_i + \nabla_{x_N} f_i S_{x_j}^{x_N})_{i,j} \end{bmatrix} & \star \\ 0 & D_{x_N} f_N \end{bmatrix}, \end{aligned}$$

where $\begin{bmatrix} (\nabla_{x_j} f_i)_{i,j} \end{bmatrix}$ indicates a matrix with block elements

$(\nabla_{x_j} f_i)_{i,j}$. Assume that for $l+1$ we have

$$\tilde{J}_{l+1} = \begin{bmatrix} & I & & 0\ 0 \\ \left[S^{x_l}_{x_1} \cdots S^{x_l}_{x_{N-2}}\right] & & & I\ 0 \\ & 0 & & 0\ I \end{bmatrix} \cdots \begin{bmatrix} I & 0 & 0 \\ S^{x_2}_{x_1} & I & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} \left[(\nabla_{x_j} f_i + \sum_{k=l+1}^N D_{x_k} f_i S^{x_k}_{x_j})_{i,j}\right] & \star & \star & \star \\ 0 & D_{x_{l+1}} f_{l+1} & \star & \star \\ 0 & 0 & \ddots & \star \\ 0 & 0 & 0 & D_{x_N} f_N \end{bmatrix}$$

Note that the terms in the last row and column blocks of $\left[(\nabla_{x_j} f_i + \sum_{k=j+1}^N D_{x_k} f_i S^{x_k}_{x_j})_{i,j}\right]$ are total derivatives: $D_{x_j} f_l = \nabla_{x_j} f_l + \sum_{k=l+1}^N D_{x_k} f_l S^{x_k}_{x_j}$ for $j \le l$, and $D_{x_l} f_i = \nabla_{x_l} f_i + \sum_{k=l+1}^N D_{x_k} f_i S^{x_k}_{x_l}$ for $i \le l$. Therefore,

$$\begin{aligned} \tilde{J}_l &= \begin{bmatrix} & I & 0\ 0 \\ -\left[S^{x_l}_{x_1} \cdots S^{x_l}_{x_{N-2}}\right] & & I\ 0 \\ & 0 & 0\ I \end{bmatrix} \tilde{J}_{l+1} \begin{bmatrix} & I & 0\ 0 \\ \left[S^{x_l}_{x_1} \cdots S^{x_l}_{x_{N-2}}\right] & & I\ 0 \\ & 0 & 0\ I \end{bmatrix} \\ &= \begin{bmatrix} & I & 0\ 0 \\ \left[S^{x_{l-1}}_{x_1} \cdots S^{x_{l-1}}_{x_{N-2}}\right] & & I\ 0 \\ & 0 & 0\ I \end{bmatrix} \cdots \begin{bmatrix} I & 0 & 0 \\ S^{x_2}_{x_1} & I & 0 \\ 0 & 0 & I \end{bmatrix} \\ &\quad \begin{bmatrix} \left[(\nabla_{x_j} f_i + \sum_{k=l}^N D_{x_k} f_i S^{x_k}_{x_j})_{i,j}\right] & \star & \star & \star \\ \left[(D_{x_j} f_l + D_{x_l} f_l S^{x_l}_{x_j})_j\right] & D_{x_l} f_l & \star & \star \\ 0 & 0 & \ddots & \star \\ 0 & 0 & 0 & D_{x_N} f_N \end{bmatrix}, \end{aligned}$$

where $D_{x_j} f_l + D_{x_l} f_l S^{x_l}_{x_j} = 0\ \forall j < l$ by definition of $S^{x_l}_{x_j}$ in (34). Finally, $D_{x_1} f_1 = \nabla_{x_1} f_1 + \sum_{k=2}^N D_{x_k} f_1 S^{x_k}_{x_1}$ implies:

$$J \sim \tilde{J}_N \sim \cdots \sim \tilde{J}_2 = \begin{bmatrix} D_{x_1} f_1 & \star & \star \\ 0 & \ddots & \star \\ 0 & 0 & D_{x_N} f_N \end{bmatrix}$$

and at steady-state $D_{x_i} f_i = \nabla_{x_i} f_i^r\ \forall i$.

3) There is a unique steady state as a consequence of the contraction condition (39) and Lemma 1. The extended total derivatives (34) can be rewritten as

$$\left[D_{x_1} f_i \cdots D_{x_N} f_i\right] S_i = \left[\nabla_{x_1} f_i \cdots \nabla_{x_N} f_i\right], \tag{42}$$

where S_i is the matrix with sensitivities in (36) truncated at i:

$$S_i = \left[\begin{array}{ccc|ccc} & I & & & 0 & \\ \hline -S^{x_{i+1}}_{x_1} & \cdots & -S^{x_{i+1}}_{x_i} & I & 0 & 0 \\ \vdots & & & & \ddots\ I & 0 \\ -S^{x_N}_{x_1} & \cdots & -S^{x_N}_{x_i} & \cdots & -S^{x_N}_{x_{N-1}} & I \end{array}\right]$$

As in the proof of Proposition 3, let L_{f_i} denote the Lipschitz constant of f_i. Given Lemma 1, it can be certified that the extended sensitivities $S^{x_j}_{x_i}$ and total derivatives $D_{x_j} f_i$ are all bounded, so there exists $L_{D,f_i,x_j} > 0$ such that $\|D_{x_j} f_i\|_2 \le L_{D,f_i,x_j}$. Defining the Lyapunov functions $V_i(x_1,\dots,x_N) = \|f_i(x_1,\dots,x_N)\|^2_{P_i}$, then we have

$$\begin{aligned} \dot{V}_i &= f_i^T P_i \left[\nabla_{x_1} f_i \cdots \nabla_{x_N} f_i\right] \left[\dot{x}_1^T \cdots \dot{x}_N^T\right]^T \\ &\quad + (\text{transpose terms}) \\ &\overset{(42)}{=} f_i^T P_i \left[D_{x_1} f_i \cdots D_{x_N} f_i\right] S_i \left[\dot{x}_1^T \cdots \dot{x}_N^T\right]^T \\ &\quad + (\text{transpose terms}) \\ &\overset{(36)}{=} f_i^T P_i \left[D_{x_1} f_i \cdots D_{x_N} f_i\right] \left[\dot{x}_1^T \cdots \dot{x}_i^T\ f_{i+1}^T \cdots f_N^T\right]^T \\ &\quad + (\text{transpose terms}) \\ &\overset{(37)}{=} f_i^T P_i \big(\cancel{\textstyle\sum_{j=1}^{i-1} D_{x_j} f_i \dot{x}_j} + D_{x_i} f_i \underbrace{(f_i + \cancel{\textstyle\sum_{j=1}^{i-1} S^{x_i}_{x_j} \dot{x}_j})}_{\dot{x}_i} \\ &\quad + \textstyle\sum_{j=i}^N D_{x_j} f_i\ f_j\big) \\ &\quad + (\text{transpose terms}) \\ &\overset{(34)}{=} \textstyle\sum_{j=i}^N f_i^T P_i (D_{x_j} f_i) f_j + (f_i^T P_i (D_{x_j} f_i) f_j)^T \\ &\overset{(39)}{\le} -\eta_i \|f_i\|^2_{P_i} + 2 \textstyle\sum_{j=i+1}^N \underbrace{L_{D,f_i,x_j} \frac{\|P_i\|_2}{\lambda_{\min}(P_j)}}_{\nu_{i,j}} \|f_i\|_{P_i} \|f_j\|_{P_j} \end{aligned}$$

Consider the parameters $\theta_i > 0$, and the Lyapunov function $V(x_1,\dots,x_N) = \sum_i \theta_i V_i(x_1,\dots,x_N)$. Then we have

$$\begin{aligned} \dot{V} &\le -\left[\|f_N\|_{P_N} \cdots \|f_1\|_{P_1}\right] W \left[\|f_N\|_{P_N} \cdots \|f_1\|_{P_1}\right]^T \\ W &= \begin{bmatrix} \theta_N \eta_N & \cdots & -\theta_i \nu_{i,N} & \cdots & -\theta_1 \nu_{1,N} \\ \vdots & \ddots & \vdots & & \vdots \\ -\theta_i \nu_{i,N} & \cdots & \theta_i \eta_i & \cdots & -\theta_1 \nu_{1,i} \\ \vdots & & \vdots & \ddots & \vdots \\ -\theta_1 \nu_{1,N} & \cdots & -\theta_1 \nu_{1,i} & \cdots & \theta_1 \eta_1 \end{bmatrix} \end{aligned}$$

Let W_i denote the upper left minor i of W. The first minor satisfies $W_1 = \zeta_1 := \theta_N \eta_N > 0$. We continue by recursion: assume there exists $\zeta_{i-1} > 0$ so that $W_{i-1} \succeq \zeta_{i-1} I$, then $W_i \succ 0$ if

$$\begin{aligned} &\theta_i \eta_i - \theta_i^2 \begin{bmatrix} \nu_{i,N} \\ \vdots \\ \nu_{i,i+1} \end{bmatrix}^T W_{i-1}^{-1} \begin{bmatrix} \nu_{i,N} \\ \vdots \\ \nu_{i,i+1} \end{bmatrix} \\ &\ge \theta_i \eta_i - \frac{\theta_i^2}{\zeta_{i-1}} \textstyle\sum_{j=i+1}^N \nu_{i,j}^2 > 0, \end{aligned}$$

or equivalently $\theta_i < \frac{\eta_i \zeta_{i-1}}{\sum_{j=i+1}^N \nu_{i,j}^2}$, which provides a recursive set of conditions on θ_i so that all minors W_i are positive definite. In the end, $W_N = W \succ 0$, so there exists $\zeta_N > 0$ such that $W \succeq \zeta_N I$ and

$$\dot{V} \le -\zeta_N \textstyle\sum_{i=1}^N V_i \le -\frac{\zeta_N}{\max_i \theta_i} V.$$

References

[1] J. H. Chow, Ed., *Time-scales in interconnected systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1982, pp. 1–5.

[2] C. Kuehn, *Multiple time scale dynamics*. Springer, 2015, vol. 191.

[3] A. N. Tikhonov, "Systems of differential equations containing small parameters in the derivatives," *Matematicheskii sbornik*, vol. 73, no. 3, pp. 575–586, 1952.

[4] G. Peponides, P. Kokotovic, and J. Chow, "Singular perturbations and time scales in nonlinear models of power systems," *IEEE Transactions on Circuits and systems*, vol. 29, no. 11, pp. 758–767, 1982.

[5] I. Subotic, D. Gross, M. Colombino, and F. Dorfler, "A Lyapunov framework for nested dynamical systems on multiple time scales with application to converter-based power systems," *IEEE Transactions on Automatic Control*, pp. 1–1, 2020.

[6] S. Muratori and S. Rinaldi, "Low- and high-frequency oscillations in three-dimensional food chain systems," *SIAM Journal on Applied Mathematics*, vol. 52, no. 6, pp. 1688–1706, 1992.

[7] A. Tavasoli, M. Eghtesad, and H. Jafarian, "Two-time scale control and observer design for trajectory tracking of two cooperating robot manipulators moving a flexible beam," in *American Control Conference*, 2007, pp. 735–740.

[8] A. Mezouar, M. Fellah, and S. Hadjeri, "Adaptive sliding mode observer for induction motor using two-time-scale approach," *Electric Power Systems Research*, vol. 77, no. 5, pp. 604 – 618, 2007.

[9] P. Kokotovic, R. O'Malley, and P. Sannuti, "Singular perturbations and order reduction in control theory — an overview," *Automatica*, vol. 12, no. 2, pp. 123 – 132, 1976.

[10] P. Kokotović, H. K. Khalil, and J. O'reilly, *Singular perturbation methods in control: analysis and design*. SIAM, 1999.

[11] H. K. Khalil and J. W. Grizzle, *Nonlinear systems*. Prentice hall Upper Saddle River, NJ, 2002, vol. 3.

[12] G. Grammel, "On nonlinear control systems with multiple time scales," *Journal of Dynamical and Control Systems*, vol. 10, no. 1, pp. 11–28, 2004.

[13] P. V. Kokotović, "Applications of singular perturbation techniques to control problems," *SIAM review*, vol. 26, no. 4, pp. 501–550, 1984.

[14] P. Kokotovic and P. Sannuti, "Singular perturbation method for reducing the model order in optimal control design," *IEEE Transactions on Automatic Control*, vol. 13, no. 4, pp. 377–384, 1968.

[15] Y. Lee, S. Park, and M. Lee, "PID controller tuning to obtain desired closed-loop responses for cascade control systems," *IFAC Proceedings Volumes*, vol. 31, no. 11, pp. 613 – 618, 1998, 5th IFAC Symposium on Dynamics and Control of Process Systems 1998 (DYCOPS 5).

[16] D. P. Bertsekas, "Nonlinear programming," *Journal of the Operational Research Society*, vol. 48, no. 3, pp. 334–334, 1997.

[17] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[18] A. Simonetto, A. Mokhtari, A. Koppel, G. Leus, and A. Ribeiro, "A class of prediction-correction methods for time-varying convex optimization," *IEEE Transactions on Signal Processing*, vol. 64, no. 17, pp. 4576–4591, 2016.

[19] M. Fazlyab, S. Paternain, V. M. Preciado, and A. Ribeiro, "Prediction-correction interior-point method for time-varying convex optimization," *IEEE Transactions on Automatic Control*, vol. 63, no. 7, pp. 1973–1986, 2018.

[20] V. M. Zavala and L. T. Biegler, "The advanced-step NMPC controller: Optimality, stability and robustness," *Automatica*, vol. 45, no. 1, pp. 86 – 93, 2009.

[21] M. Diehl, H. J. Ferreau, and N. Haverbeke, *Efficient Numerical Methods for Nonlinear MPC and Moving Horizon Estimation*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 391–417.

[22] J. N. Foerster, R. Y. Chen, M. Al-Shedivat, S. Whiteson, P. Abbeel, and I. Mordatch, "Learning with opponent-learning awareness," in *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2018.

[23] Y. Wang, G. Zhang, and J. Ba, "On solving minimax optimization locally: A follow-the-ridge approach," in *International Conference on Learning Representations (ICLR)*, 2020.

[24] A. V. Fiacco, "Sensitivity analysis for nonlinear programming using penalty methods," *Mathematical programming*, vol. 10, no. 1, pp. 287–311, 1976.

[25] K. Jittorntrum, "Solution point differentiability without strict complementarity in nonlinear programming," in *Sensitivity, Stability and Parametric Analysis*. Springer, 1984, pp. 127–138.

[26] J. F. Bard, *Practical bilevel optimization: algorithms and applications*. Springer Science & Business Media, 1998.

[27] K. J. Åström and B. Wittenmark, *Adaptive control*. Courier Corporation, 2013.

[28] S. G. Krantz and H. R. Parks, *The implicit function theorem: history, theory, and applications*. Springer Science & Business Media, 2012.

[29] W. Lohmiller and J.-J. E. Slotine, "On contraction analysis for non-linear systems," *Automatica*, vol. 34, no. 6, pp. 683–696, 1998.

[30] Z. Aminzare and E. D. Sontag, "Contraction methods for nonlinear systems: A brief introduction and some open problems," in *53rd IEEE Conference on Decision and Control*. IEEE, 2014, pp. 3835–3847.

[31] S. Sastry, *Nonlinear systems: analysis, stability, and control*. Springer Science & Business Media, 2013, vol. 10.

[32] D. Del Vecchio and J.-J. E. Slotine, "A contraction theory approach to singularly perturbed systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 3, pp. 752–757, 2012.

[33] A. Yazdani and R. Iravani, *Voltage-sourced converters in power systems: modeling, control, and applications*. John Wiley & Sons, 2010.

[34] A. Sinha, P. Malo, and K. Deb, "A review on bilevel optimization: From classical to evolutionary approaches and applications," *IEEE Transactions on Evolutionary Computation*, vol. 22, no. 2, pp. 276–295, 2018.

[35] S. Dempe, B. S. Mordukhovich, and A. B. Zemkoho, "Sensitivity analysis for two-level value functions with applications to bilevel programming," *SIAM Journal on Optimization*, vol. 22, no. 4, pp. 1309–1343, 2012.

[36] G. Savard and J. Gauvin, "The steepest descent direction for the nonlinear bilevel programming problem," *Oper. Res. Lett.*, vol. 15, no. 5, p. 265–272, Jun. 1994.

[37] K. E. Atkinson, *An introduction to numerical analysis*. John wiley & sons, 2008.

[38] H. J. Stetter, *Analysis of discretization methods for ordinary differential equations*. Springer, 1973, vol. 23.

[39] A. M. Stuart and A. Humphries, "Numerical analysis of dynamical systems," *Acta numerica*, vol. 3, no. 1, pp. 467–572, 1994.

[40] G. Qu and N. Li, "On the exponential stability of primal-dual gradient dynamics," *IEEE Control Systems Letters*, vol. 3, no. 1, pp. 43–48, Jan 2019.

[41] H. Von Stackelberg, *Market structure and equilibrium*. Springer Science & Business Media, 2010.

[42] T. Fiez, B. Chasnov, and L. J. Ratliff, "Convergence of learning dynamics in Stackelberg games," *arXiv preprint arXiv:1906.01217*, 2019.

[43] T. Fiez, B. Chasnov, and L. Ratliff, "Implicit learning dynamics in Stackelberg games: Equilibria characterization, convergence analysis, and empirical study," in *International Conference on Machine Learning (ICML)*, 2020.

[44] C. Jin, P. Netrapalli, and M. Jordan, "What is local optimality in nonconvex-nonconcave minimax optimization?" in *International Conference on Machine Learning*. PMLR, 2020, pp. 4880–4889.

[45] J. F. Bard and J. E. Falk, "An explicit solution to the multi-level programming problem," *Computers & Operations Research*, vol. 9, no. 1, pp. 77 – 100, 1982.