Learning soft interventions in complex equilibrium systems

Abstract

Complex systems often contain feedback loops that can be described as cyclic causal models. Intervening in such systems may lead to counterintuitive effects, which cannot be inferred directly from the graph structure. After establishing a framework for differentiable interventions based on Lie groups, we take advantage of modern automatic differentiation techniques and their application to implicit functions in order to optimize interventions in cyclic causal models. We illustrate the use of this framework by investigating scenarios of transition to sustainable economies.

1 INTRODUCTION

Designing optimal interventions in complex systems, composed of many interacting parts, is a key objective in multiple fields. In the context of socio-economic systems, the design of public policies to improve economic and social welfare is a major source of scientific and political debate. Moreover, the positive aspects of socio-economic activities need to be traded-off with their environmental impacts, as their long term consequences may considerably affect societies [Dearing et al., 2014, Sherwood and Huber, 2010]. Interestingly, a priori intuitive interventions in such systems may lead to paradoxical outcomes. The rebound effect in energy economy, first reported by Jevons [1866], is paradigmatic: while the energy efficiency of devices may considerably increase due to technological improvements, this may trigger an overall increase of energy consumption due to increases in demand [Brockway et al., 2021]. This suggests in particular that efficiency alone may not be the best way to foster a transition towards sustainability, and calls for a quantitative study of optimal interventions in such complex systems [Arrobbio and Padovan, 2018]. As argued for the case of rebound effects [Wallenborn, 2018], such unexpected behaviors may reflect balanced causal relationships designed by evolution [Andersen, 2013], and feedback loops [Blom and Mooij, 2021] that maintain a system at an optimal "equilibrium" operating point independent from external perturbations, challenging classical causal inference assumptions of faithfulness and acyclicity.

While interventions have been extensively investigated theoretically in the field of causality [Pearl, 2000, Imbens and Rubin, 2015], the case of systems incorporating feedback loops remains particularly challenging, and therefore led to only limited applications to real-life complex systems. A possible first step to study such systems is to approximate them by a model that operates at an equilibrium point, and can thus be described by a cyclic structural causal model [Bongers et al., 2016]. Such models satisfy a self consistent set of equations that, under unique solvability assumptions, fully identifies the operating point, and allows to study interventions. For practical and ethical reasons, interventions that do not change the causal structure, called soft interventions, arguably provide a more realistic account of changes that can be performed in real life systems. While a restricted set of qualitative results exist for such interventions [Blom et al., 2020], their quantitative assessment and design in complex systems is made difficult by the analytical intractability of the self-consistency relations.

In this paper, we propose a framework for a general class of differentiable parametric soft interventions based on Lie groups and leverage recent technical and algorithmic developments allowing learning implicit functional relationships [Bai et al., 2019] to optimize such interventions. After defining Lie interventions and assaying their theoretical properties, we provide a computational framework to optimize them. We illustrate its application to economic models derived from real data, offering a novel approach to computational sustainability. Proofs are provided in Appendix B.

Related work. Various types of economic equilibrium models (EEM) have been used to investigate macroeconomic effect of specific interventions [Wiebe et al., 2018,

Submitted to the 38th Conference on Uncertainty in Artificial Intelligence (UAI 2022). **To be used for reviewing only**.

Wood et al., 2018]. In contrast to such work, we develop a general optimization framework that allows the optimal design of interventions to achieve specific goals. A restricted set of EEMs have been investigated more extensively from an optimization perspective (see, e.g., Esteban-Bravo 2004); however, these are restricted to rather specific assumptions and constraints that allow to address optimization with linear programming approaches. Instead, we rely on automated differentiation and backpropagation algorithms that allow studying mechanisms and interventions with a broad range of non-linearities. In the field of causality, several studies investigate the relationship between the equilibrium of dynamical systems and structural causal models (SCM) [Mooij et al., 2013, Peters et al., 2020] and how the causal structure can be learnt from data. In contrast, we focus on designing soft interventions in an known SCM at equilibrium. While the specificity of soft interventions have started to be investigated theoretically in structural causal models [Correa and Bareinboim, 2020, Blom et al., 2020], the present work is to the best of our knowledge the first to investigate theoretically and algorithmically the design of soft interventions in cyclic causal models. The algorithmic approach relies on modeling economic equilibrium with deep equilibrium models [Bai et al., 2019]. This approach belongs to the category of implicit deep learning [El Ghaoui et al., 2021], which has been used in a variety of applications such as model predictive control [Amos et al., 2018] and multi-agent trajectory modelling [Geiger and Straehle, 2020].

2 MOTIVATION AND BACKGROUND

2.1 ENVIRONMENTAL ECONOMIC MODELS

In the face of the increasing severity of climate change and further environmental impacts of human activities, our societies face challenges to transition to more sustainable economies. An overarching difficulty is the complexity of the systems that need to be intervened on, which comprise tightly intertwined components, ranging from economic agents to a broad variety of ecosystems [Haberl et al., 2019].

A classical way to represent the economy and its impacts are input-output (IO) multi-sector economic equilibrium models [Stadler et al., 2018], in which economic activities are divided in d interdependent *sectors* and described by a positive d-dimensional *output* vector $\boldsymbol{x}$ (see Appendix A). We take as a guiding example the demand-driven model introduced by Leontief [1951], which is the basis of the *Input-Output analysis* approach to environmental impact assessment. In such models, the sectors' outputs at economic equilibrium $\boldsymbol{x}^*$ are depend on the vector $\boldsymbol{y}$ gathering final demand for each product (consumed by users instead of being used to make another product). Satisfying the demand of all sectors implies the self-consistent equation

$$\boldsymbol{x}^* = A\boldsymbol{x}^* + \boldsymbol{y}\,, \tag{1}$$

where A is the so-called *technical coefficients matrix*, with A_{ij} the amount of each product i used as input to produce product j. An example of technical coefficient matrix estimated from economic data is provided in Fig. 1a. While such equilibria can be thought of as the asymptotic value of $\boldsymbol{x}$ in a dynamic model (see Appendix A) we focus our analysis on the equilibrium equations without consideration for the dynamics that gives rise to it. In turn, the socio-economic impacts (e.g., employment) and environmental stressors (e.g., GHG emissions, water use, ...) of each sector's activity is gathered in a vector of *impacts* $\boldsymbol{s}$ such that

$$\boldsymbol{s} = R\boldsymbol{x} \tag{2}$$

where R is a *footprint intensity* matrix such that R_{ij} is the amount of impact of type i generated by unit of output j. To mitigate major long term negative consequences of environmental stressors (see, e.g., Dearing et al. 2014, Sherwood and Huber 2010), a reorganization of the global economy is required, which may consist in intervening on economic sectors, their impacts and their interactions reflected in the matrices A and R. However, this faces three challenges.

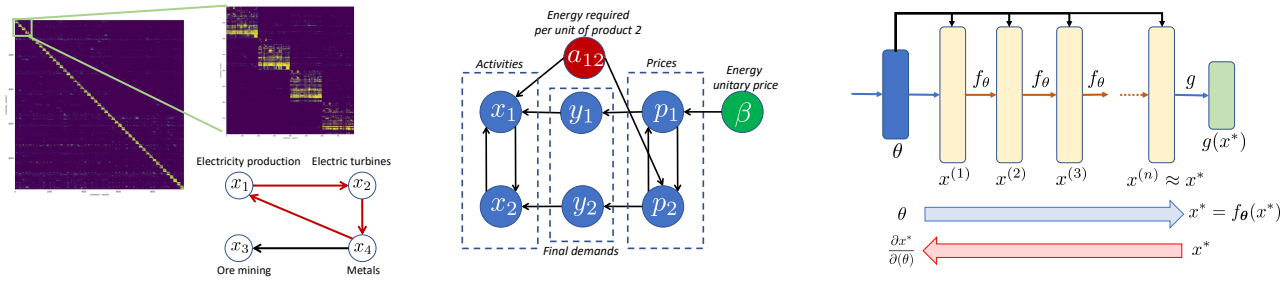
***Challenge 1*: social acceptability.** Reducing the activity in a sector may lead to both positive environmental effects (by reducing the amount of stressors) and negative socio-economic impacts (such as reducing economic growth and employment). Decision makers thus trade off environmental goals with the social acceptability of the chosen policies.

***Challenge 2*: recurrence between sectors.** The sectors' activities are tightly intertwined by their reciprocal demands, as illustrated at the bottom of Fig. 1a: electricity production through renewable energy requires wind turbines, which require metals, while the metal industry requires itself electricity to extract metals from ores and transform them. Such cycles make it challenging to anticipate the system-wide consequences of interventions a particular sector.

***Challenge 3*: rebound effects.** The complexity of the economic system also manifests itself through balancing mechanisms that reflect the utility maximization behavior of economic agents, such as rebound effects. Consider $\boldsymbol{x}^*$ in eq. 1, which can be written as a function of final demand

$$\boldsymbol{x}^* = (I - A)^{-1}\boldsymbol{y}\,. \tag{3}$$

In practice, final demand is influenced by prices of each good and often modeled by a static demand curve d_i for good i such that $y_i = d_i(p_i)$. A final demand rebound through prices can be simulated in the Leontief model as follows. Energy efficiency of the production of a particular good j corresponds to a decrease of A_{ej}, where e indicates the energy sector, but this modification also affects the unit price through energy costs. For simplicity, we define the price vector $\boldsymbol{p}$ of goods such that it is proportional to the energy required in all sectors involved in the production



(a) (b) (c)

Figure 1: (a) Top left: technical coefficient matrix between 200 sectors and 49 world regions for 2011 (source: Exiobase 3, Stadler et al. 2018). Top right: magnification of the top left corner of this matrix. Diagonal blocks reflect the stronger sector dependency within a country. Bottom right: putative example of cyclic dependency between different sectors. (b) Illustration of the causal graph for *rebound trough prices* in a two sector economy. (c) Principle of deep equilibrium models.

of one unit of this good. It can thus be modelled by a self-consistent relation involving the technical coefficient matrix:

$$\boldsymbol{p}^* = A^\top \boldsymbol{p}^* + \beta \boldsymbol{\delta}_e \,,$$

where δ_e is a canonical basis vector which takes value 1 for the energy sector, and value 0 for all other sectors. For illustrative purposes, the overall causal model is shown in Fig. 1b in the case of a two sector economy, with sector 1 being the energy sector. The price-based rebound mechanisms then operates as follows: a decrease of A_{ej} will decrease energy demand on sector e, but will also decrease the unit price of goods for sector j (and downstream sectors consuming its goods). Because the demand curves d_j are monotonically decreasing, the price drop increases the final demand for these products, which in turn increases economic activity according to eq. 3, and their environmental footprint. The rebound may thus be avoided by simultaneously intervening on the unit price of energy β through a tax policy, so that price level is maintained high and prevents increases of final demand (see Fig. 1b). Importantly, while eq. (3) provides a linear relationship between activity and final demand, once we assume $\boldsymbol{p}$ is price dependent, the system of equations becomes non-linear and finding an analytic expression of the economic equilibrium is nontrivial. Our approach to designing interventions in cyclic causal models will be applied to models illustrating the above three challenges.

2.2 CYCLIC CAUSAL MODELS

Interventions and their effects on systems have been investigated using Structural Causal Models (SCM) [Pearl, 2000]. In this framework, relationships between observed variables X_k are described by a set of structural assignments

$$X_k := f_k(\mathbf{Pa}_k, \epsilon_k) \,,$$

where $\mathbf{Pa}_k$ indicates the parents of variable X_k in an associated directed causal graph, such as the one illustrated in Fig. 1a. Interventions turn an SCM into a different one, by applying a modification to at least one of its elements. Broadly construed, interventions range from "hard" interventions that modify the structure of the graph to "soft" interventions that do not [Eberhardt and Scheines, 2007].

While in acyclic graphs, interventions have generic effects on their descendants in the causal graph, and no effects on the parents, Blom et al. [2020] have shown that causal effects are less easy to read in graphs containing cycles. Anticipating the effect of interventions in cyclic graphs overall requires to estimate the changes in the equilibrium point, which is typically non-trivial. While a variety of approaches may be used (e.g., based on root finding approaches), designing optimal interventions for self-consistent equations that cannot be handled analytically is challenging, especially in high dimensional systems. Recent work in deep neural network has come up with techniques allowing gradient descent based optimization of such equilibrium models [Bai et al., 2019].

2.3 DEEP EQUILIBRIUM MODELS

Bai et al. [2019] introduced deep learning architecture elements with input-output functional relationships $\boldsymbol{x}^* = \mathbf{g}(\boldsymbol{\theta})$ between variables $\boldsymbol{x}^*$ and parameters $\boldsymbol{\theta}$ that are only defined through a self-consistent equation

$$\boldsymbol{x}^* = \mathbf{f}_{\boldsymbol{\theta}}(\boldsymbol{x}^*).$$

Assuming that for each value pair $\boldsymbol{\theta}$ there is a unique solution $\boldsymbol{x}^*$, the gradient with respect to the input or the parameter (indicated by the "(.)" notation) can be obtained through another self-consistent equation

$$\frac{\partial \boldsymbol{x}^*}{\partial(.)} = \frac{\partial \mathbf{f}}{\partial(.)}(\boldsymbol{x}^*, \boldsymbol{y}) + \frac{\partial \mathbf{f}}{\partial \boldsymbol{x}} \frac{\partial \boldsymbol{x}^*}{\partial(.)} \,.$$

Overall, $\mathbf{g}$ can be integrated as a layer in more complex differentiable models, which, as depicted in Fig. 1c, can be

understood as a cascade of multiple layers with identical functions and shared parameters, with specific accelerated fixed point iteration approaches to compute the forward and backward passes [Bai et al., 2019]. In this paper, we use Anderson's acceleration [Walker and Ni, 2011], which essentially generalizes the forward iteration approach (i.e. iterating $\boldsymbol{x}_{k+1} = \mathbf{f}(\boldsymbol{x}_k)$ until convergence) by leveraging the m previous estimates in order to find a better estimate. Overall, these layers offer a differentiable framework for investigating the behavior of cyclic graphs that we use to design interventions.

2.4 LIE GROUPS

Using deep equilibrium models, we can learn differentiable soft interventions compatible with classical optimization frameworks. We will use the concept of Lie groups, which are smooth manifolds of transformations (see Appendix A for more background), in order to implement smooth soft interventions. In short, a group G is a set of objects equipped with a group "multiplication" operation mapping $(g_1, g_2) \in G^2$ to $g_1 g_2 \in G$ and an inverse operation g^{-1} with the following properties:

- (associativity) $(g_1 g_2) g_3 = g_1 (g_2 g_3)$,
- (identity element) there exist a unique identity element e such that for all g, $eg = ge = g$,
- (inverse) for all $g \in G$, there exists a unique element g^{-1} such that $gg^{-1} = g^{-1}g = e$.

Groups perform transformations on objects in a set $\mathcal{X}$ through the definition of a group action operation φ mapping $(g, x) \in G \times \mathcal{X}$ to $\varphi_g(x) = g \cdot x \in \mathcal{X}$, such that

- (identity) for all $x \in \mathcal{X}$, $e \cdot x = x$,
- (compatibility) for all $(g, h) \in G \times G$, for all $x \in \mathcal{X}$, $g \cdot (h \cdot x) = (gh) \cdot x$.

A real Lie group is a group that is also a finite-dimensional real smooth manifold (see Appendix A), in which the group operations of multiplication and inversion are smooth maps. The differentiability of Lie groups will be leveraged to design smooth interventions.

3 INTERVENING IN SMOOTH MODELS

3.1 SMOOTH CAUSAL GRAPHICAL MODELS

We define a smooth structural causal model (SSCM) as a set of variables $\{x_j\}$ related to each other through structural equations and vertices in a directed graph as follows.

Definition 1 (SSCM)**.** A d-dimensional smooth structural causal model is a triplet $(\mathcal{X}, \mathcal{T}, \mathbb{S}, \mathcal{G})$ consisting of

- two collections of smooth manifolds $\mathcal{X} = \{\mathcal{X}_i\}_{i=1..d}$ and $\mathcal{T} = \{\mathcal{T}_j\}_{j=1..d}$,
- a directed graph $\mathcal{G} = (V, E)$ with set V of d vertices and set E of directed edges between them, each vertex being associated to one variable $x_i \in \mathcal{X}_j$,
- a set $\mathbb{S}$ of structural assignments $\{x_j := f_j(\mathbf{Pa}_j, \theta_j), \theta_j \in \mathcal{T}_j\}_{j=1,\dots,d}$, where f_k are smooth maps, and $\mathbf{Pa}_j$ are the variables indexed by the set of parents of vertex j in $\mathcal{G}$.

Compared to classical definitions of SCMs (see, e.g., Peters et al. 2017), we have replaced exogenous random variables by deterministic parameters living on a manifold. This general definition does not prevent assigning random variables to some (components of) these parameters. In the cases considered here, $\mathcal{T}_i$ are subsets of Euclidean spaces. We are particularly interested in cyclic SCMs, where there exists at least one directed path linking one vertex to itself. As a consequence, the possible values achieved by each variable have to be chosen among the solutions of the d self-consistent structural equation constraints. We assume the base causal model is locally uniquely solvable.

Definition 2. A SSCM is locally uniquely solvable around a reference point $(\boldsymbol{x}^{\text{ref}}, \boldsymbol{\theta}^{\text{ref}})$ whenever there exists a neighborhood $U_{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}^{\text{ref}}$ and a neighborhood $U_{\boldsymbol{x}}$ of $\boldsymbol{x}^{\text{ref}}$ such that for all $\boldsymbol{\theta} \in U_{\boldsymbol{\theta}}$ there exists a unique (self-consistent) solution to the set of structural assignments $\boldsymbol{x}^*(\boldsymbol{\theta}) \in U_{\boldsymbol{x}}$.

Note that this is adapted to our SSCM definition and differs from the unique solvability definition of Bongers et al. [2016], which was formulated for causal models with random exogenous variables. This property is guaranteed by a condition on the Jacobian of the structural equations.

Proposition 1. We say the SSCM is locally diffeomorphic at $(\boldsymbol{x}^{\text{ref}}, \boldsymbol{\theta}^{\text{ref}})$ when $(\boldsymbol{x}^{\text{ref}}, \boldsymbol{\theta}^{\text{ref}})$ is a solution and the Jacobian of the mapping $\boldsymbol{x} \to \boldsymbol{x} - \mathbf{f}(\boldsymbol{x}, \boldsymbol{\theta}^{\text{ref}})$ is invertible. Such a SSCM is uniquely solvable around this reference point and the local mapping $\boldsymbol{\theta} \mapsto \boldsymbol{x}^*(\boldsymbol{\theta})$ is smooth.

In the context of IO analysis presented in Section 2.1, the variables can be the sector's outputs and unit prices. For eq. (1), the resulting SSCM thus contains the affine structural assignments associated to each component of $\boldsymbol{x}$

$$\mathbb{S} = \{x_k := \sum_j A_{kj} x_j + y_k\} ,$$

which are clearly smooth, and the $\{A_{kj}, y_k\}$'s may be assumed fixed or free parameters within an interval.

3.2 LIE INTERVENTIONS

We will consider interventions parameterized by an element u that turns the unintervened equilibrium solution $\boldsymbol{x}^*(\boldsymbol{\theta})$ into the *intervened equilibrium solution* $\boldsymbol{x}^{(u)}(\boldsymbol{\theta})$ over a range of values of $\boldsymbol{\theta}$. In particular, we define Lie interventions implemented through the action of a Lie group.

Definition 3 (Lie intervention)**.** A Lie intervention on an SSCM $\mathcal{M} = (\mathcal{X}, \mathcal{T}, \mathbb{S}, \mathcal{G})$ with a set of smooth structural assignments $\mathbb{S}$ is a pair (L, φ) where L is a Lie group and a smooth group action $\varphi : L \times \mathbb{S} \to \mathbb{S}$. The action defines a family of intervened SSCMs $\mathcal{M}^{(g)} = (\mathcal{X}, \mathcal{T}, \varphi(g, \mathbb{S}), \mathcal{G})$, for g in a neighborhood of the identity within L.

Note in particular that applying the identity element of the group leads to the original (unintervened) causal model. Such interventions preserve unique solvability.

Proposition 2 (Solvability)**.** For a Lie intervention on a locally diffeomorphic SSCM, there is a neighborhood U_L of the identity e in L such that the intervention is soft, the familly of intervened SSCMs is locally uniquely solvable and the local mapping to the intervened solution $(g, \boldsymbol{\theta}) \mapsto \boldsymbol{x}^{(g)}(\boldsymbol{\theta})$ is smooth.

Multiplicative Lie interventions. A simple way of intervening on an arbitrary system is to multiply one selected assignment by a strictly positive scalar coefficient. We can consider $\mathbb{R}_+^*$ equipped with multiplication as a Lie group, that acts on a node by rescaling its structural assignment. Several such *scalar* Lie interventions can then be combined into a *distributed* intervention on a set of nodes instead of a single one. A group element is then a strictly positive vector $\boldsymbol{\alpha} > 0$ acting on assignments indexed by I such that

$$\boldsymbol{\alpha} \cdot \mathbb{S}_{|I} = \{x_k \coloneqq \alpha_k f_k(x_k, \theta_k), k \in I\} .$$

In the context of Input-Output models presented in Section 2.1, applying this intervention can be seen as reducing or increasing the demand for products of specific sectors. Reducing the demand for a sector with large GHG emissions is for example a relevant objective for the transition to a sustainable economy and may be implemented by public policy in various ways (taxes, norms, ...). Such interventions are investigated in industrial ecology [Wood et al., 2018].

3.3 INVARIANT SOFT INTERVENTIONS

The rebound effect is paradigmatic of interventions that may trigger undesired effects that we wish to prevent. To this end, simultaneous interventions on other parts of a system have been considered in applications. For example, a rebound through prices can be prevented by a simultaneous auxiliary intervention of prices through taxes, such that the prices remain invariant to the overall intervention. Using the SSCM framework, we theoretically investigate the conditions under which some variables of the causal model can be maintained invariant to the Lie intervention on others.

Motivating example. Consider the following SSCM with parameters $\boldsymbol{\theta} = (\tau, \alpha, \beta, \gamma)$ with distributed multiplicative Lie intervention $\boldsymbol{u}$:

$$\begin{cases} x &= \tau , \\ y &= u_y(\alpha x + \beta z) , \\ z &= u_z \gamma y . \end{cases} \tag{4}$$

By choosing $u_z = \frac{1}{u_y}$, the intervened equilibrium solution component $z^{(u)}$ becomes insensitive to multiplicative interventions (u_y, u_z), such that $z^{(\boldsymbol{u})}(\boldsymbol{\theta}) = z^*(\boldsymbol{\theta})$ for arbitrary values of parameters $\boldsymbol{\theta}$ in a neighborhood of the reference parameter (see Appendix C). This result suggests that the influence of soft interventions (u_y in this example) can be restricted to a subset of nodes, by choosing an second intervention (u_z in this example) on an auxiliary variable. However, it is unclear whether this result still holds when the functional assignment of z becomes non-linear.

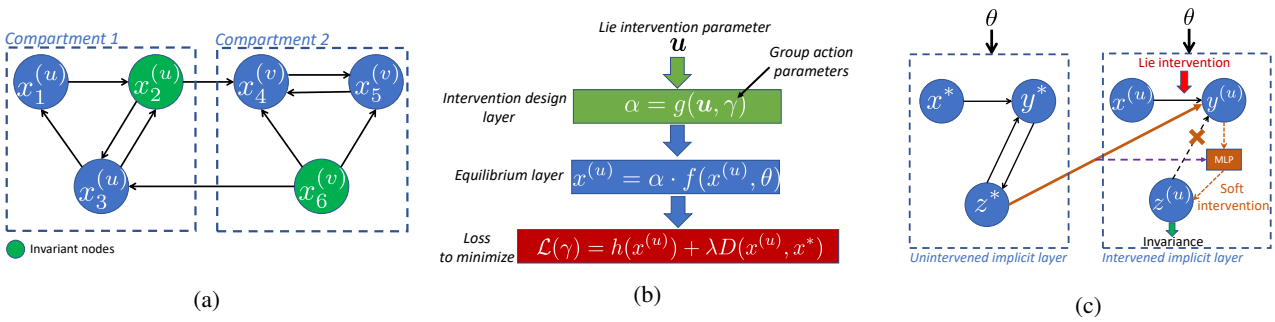
To frame this question in a general setting, we introduce a class of soft interventions under invariance constraint.

Definition 4 (Invariant soft interventions)**.** Given an SSCM with Lie intervention from group L on node i. The intervention leaves node j *invariant* by *leveraging* node k if for all group elements u in a neighborhood $\mathcal{N}$ of the identity, there exists a soft intervention on node k, $f_k^{(u)}(\mathbf{Pa}_k, \boldsymbol{\theta})$, replacing functional assignment f_k such that the intervened node value $x_j^{(u)}$ satisfies $x_j^{(u)}(\boldsymbol{\theta}) = x_j^*(\boldsymbol{\theta})$ in a neighborhood of the reference parameter. Node i is called the intervened node, node j is called the invariant node, and node k is called the auxiliary node.

Remarks: The soft intervention property is key, as it entails that the use of an auxiliary variable to enforce the invariance constraint must only exploit the information available to this node as defined by its parents in the unintervened graph (and no parameter values). This constraint makes deployment more realistic in a complex system, as intervening does not require supervision by an external entity monitoring the whole system. Unless otherwise stated, the auxiliary node will be chosen identical to the invariant node.

Let us denote $\boldsymbol{x}_{-j}$ and $\mathbf{f}_{-j}$ the vector and mapping with the j-th component removed. We also define two quantities important for the existence of such interventions. The partial derivative $\frac{\partial x_j^*}{\partial x_k}_{|\boldsymbol{\theta}=\boldsymbol{\theta}^{\mathrm{ref}}}$ is obtained by performing a hard intervention $x_k = \lambda$ leading to equilibrium value $x_j^{(\lambda)}(\boldsymbol{\theta}^{\mathrm{ref}})$, and computing the derivative $\frac{dx_j^{(\lambda)}}{d\lambda}_{|\lambda=x_k^*(\boldsymbol{\theta}^{\mathrm{ref}})}$. The Jacobian $J^{\boldsymbol{\theta}}_{x^*_{\mathbf{Pa}_k}}(\boldsymbol{\theta}^{\mathrm{ref}})$ is the Jacobian of the mapping from the parameters $\boldsymbol{\theta}$ to the vector consisting of the parent nodes of k at equilibrium. Based on these two quantities, we have the following sufficient condition.

Proposition 3. Consider an SSCM locally diffeomorphic at $(\boldsymbol{x}^{\mathrm{ref}}, \boldsymbol{\theta}^{\mathrm{ref}})$ with intervened/invariant/auxiliary triplet of nodes $(i, j \neq i, k \neq i)$. If the Jacobian of the mapping $\boldsymbol{x}_{-j} \to \boldsymbol{x}_{-j} - \mathbf{f}_{-j}(\boldsymbol{x}_{-j}, \boldsymbol{\theta}^{\mathrm{ref}})$ is invertible, $J^{\boldsymbol{\theta}}_{x^*_{\mathbf{Pa}_k}}(\boldsymbol{\theta}^{\mathrm{ref}})$ has full column rank, and $\frac{\partial x_j^*}{\partial x_k}_{|\boldsymbol{\theta}=\boldsymbol{\theta}^{\mathrm{ref}}} \neq 0$, then the intervention on i leaves node j invariant by leveraging node k.



(a) (b) (c)

Figure 2: (a) Illustration of a compartmentalized intervention: enforcing invariance of the green nodes allows each compartment to be independently influenced by two (invariant) interventions u and v. (b) Architecture for Lie intervention optimization. The equilibrium layer is controlled by intervention parameters and a loss is applied to its output. (c) Schematic representation of the procedure to learn invariant soft intervention (y: intervened node, z: invariant and auxiliary node). A multilayer perceptron (MLP) learns the soft intervention enforcing invariance of $z^{(u)}$ over a range of parameter values.

This result suggests that the motivating example of eq. (4) can be extended, in a neighborhood of the identity, beyond the linear case, when the number of free parameters considered remains low relative to the number of parents of the auxiliary node. However, as can be seen in the proof, the soft intervention on the auxiliary variable is given by an implicit function theorem, suggesting non-parametric models are necessary to learn it (based e.g. on automated differentiation methods). This will be described in Sec. 4.

3.4 COMPARTMENTALIZED INTERVENTIONS

Invariant interventions allow to restrict the propagation of effects to a subset of nodes. If a complex system can be partitioned into sparsely connected subsets of nodes, we can consider designing such interventions in order to modify the equilibrium values of each compartment independently from each other.

Definition 5. Given a partition of the SSCM nodes into compartments $\{C_k\}$. Interventions are compartmentalized when they affect only values of a single compartment.

The following result guaranties that if the nodes influencing other compartments are made invariant, interventions on each compartment can be designed and performed independently from each other as their effects remain confined to their own compartment.

Proposition 4. Given a partition $\{C_k\}$ of the SSCM nodes. If for each compartment k there exists one Lie intervention performed on structural equations such that intervened, auxiliary and invariant nodes belong to the compartment, and all nodes of this compartment that have an outgoing arrow pointing to a different compartment are invariant, then those interventions are compartmentalized.

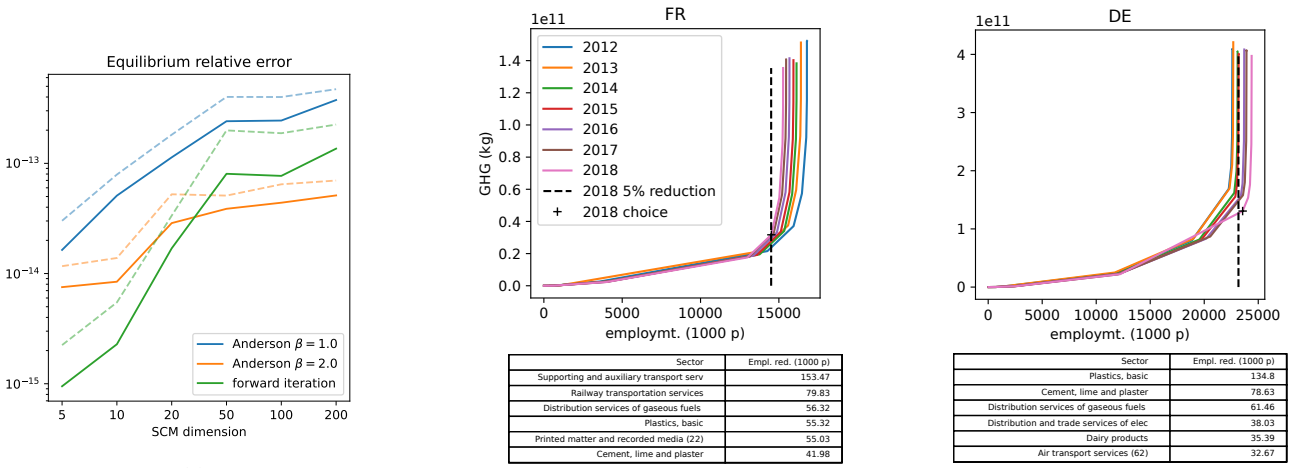
A fundamental aspect of this result is that, from the definition of invariant interventions, compartmentalization is valid over a range of parameters of the causal model (a neighborhood of the reference point) and a range of Lie interventions parameters (a neighborhood of the identity). This can be seen as a way to enforce interpretability of interventions by restricting their influence to a specific subsystem, at least for a range of parameter values. An illustration of a setting compatible with Prop. 4 is provided in Fig. 2a, where the equilibria of two sparsely connected compartments are interdependent (notably, the causal ordering algorithm described in Blom et al. [2020] returns a single cluster merging both compartments). Enforcing invariance of the green nodes, each associated to one intervention (u and v) within their compartment allows applicability of Prop. 4.

4 INTERVENTION DESIGN

To address *Challenge 2* of Sec. 2.1, we design interventions with implicit layers (see Appendix D for additional details).

Differentiable architecture. Base optimization relies on a differentiable architecture comprising one central module representing the cyclic SSCM. Essentially, the cyclic model is represented by an equilibrium layer following Bai et al. [2019], schematized in Fig. 1c: the differentiable module is designed such that forward and backward passes through the equilibrium layer use Anderson acceleration to solve a fixed point equation. This equilibrium layer is cascaded if necessary with parametric layers to achieve specific goals. The architectures are implemented using the PyTorch library.

Lie intervention optimization We design an architecture around the equilibrium module to optimize multiplicative intervention according to a loss, as represented in Fig. 2b. Parameters $\boldsymbol{u}$ of the Lie group element are optimized in order to minimize an objective $\mathcal{L}(\boldsymbol{x}^{(u)})$ achieved by the equilibrium solution of the SSCM. This objective may include an additional regularization term, $D(\boldsymbol{x}^{(u)}, \boldsymbol{x}^*)$ with



Sector	Empl. red. (1000 p)
Supporting and auxiliary transport serv	153.47
Railway transportation services	79.83
Distribution services of gaseous fuels	56.32
Plastics, basic	55.32
Printed matter and recorded media (22)	55.03
Cement, lime and plaster	41.98

Sector	Empl. red. (1000 p)
Plastics, basic	134.8
Cement, lime and plaster	78.63
Distribution services of gaseous fuels	61.46
Distribution and trade services of elec	38.03
Dairy products	35.39
Air transport services (62)	32.67

(a) (b) (c)

Figure 3: (a) Equilibrium relative error for different methods and SCM dimensions (solid: mean, dashed: mean+std). (b-c) Outcome of Lie intervention optimization on country models of GHG emission reduction in France (b) and Germany (c) for varying values of λ in eq. (5), based on economic models estimated from different years. Dashed line indicates 5% reduction in employment and the cross the corresponding λ choice. Tables show sectors with largest employment reduction for 2018.

regularization parameter λ, to enforce that some properties of the intervened system remain invariant or close to the original, non-intervened, equilibrium solution $\boldsymbol{x}^*$.

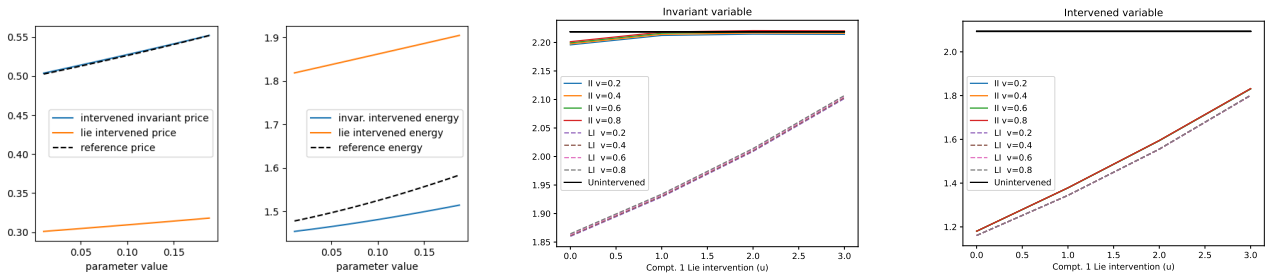
Learning invariant interventions. In order to enforce invariance of interventions based on Sec. 3.3, we follow the procedure exemplified in Fig. 2c. We design two implicit layers with shared parameters $\boldsymbol{\theta}$, the first layer being unintervened giving the corresponding equilibrium values of the nodes, and the second one being invariantly intervened, for a fixed value of Lie intervention u on the intervened node. In the intervened layer, we replace putative incoming arrows from the invariant node by arrows from the same node in the unintervened graph (as this replacement encodes the invariance assumption) and we replace the functional assignment of the auxiliary node by a Multi-Layer Perceptron (MLP), relying on universal approximation properties to learn a soft intervention that satisfies invariance. We use a least square loss between the intervened and unintervened equilibrium values of the invariant node in order to train the MLP.

5 EXPERIMENTS

The following toy and semi-synthetic experiments illustrate how our framework contributes addressing sustainability challenges exposed in Sec. 2.1.

Evaluation of equilibrium estimation. We first evaluate the performance of equilibrium layers in computing an accurate estimate of the SSCM solution $\boldsymbol{x}^*$. For that we use the SSCM associated to the economic equilibrium of equation (1) where we select a subset of sectors in order to vary the dimension of A. The full matrices A, as well as the final demands $\boldsymbol{y}$ are estimated from the Exiobase 3 dataset [Stadler et al., 2018] for years 2012-2018, using the *Pymrio* library [Stadler, 2021] for five countries (France, Germany, Italy, USA, Great-Britain). We compare Anderson acceleration (see [Walker and Ni, 2011]) for two different choices of the mixing parameter β, together with the baseline forward iteration approach that simply consists in iterating $\boldsymbol{x}_{k+1} = \mathbf{f}(\boldsymbol{x}_k)$. For each choice of dimension and fixed-point iteration algorithm, we compute the relative error $\frac{\|\boldsymbol{x}^* - \mathbf{f}_{\boldsymbol{\theta}}(\boldsymbol{x}^*)\|}{\|\boldsymbol{x}^*\|}$. The results, averaged across countries and years, show that although forward iteration is the most accurate in lower dimensions, Anderson acceleration with a relaxation parameter $\beta = 2.0$ performs better for SSCM dimensions larger than 50. Interestingly, Anderson acceleration with $\beta = 1.0$ gives the worst performance, suggesting an appropriate choice of β is key.

Optimization of multiplicative Lie interventions. In order to investigate *Challenge 1*, we optimize the IO demand driven model of eq. (1). The matrices A and R, as well as the final demands $\boldsymbol{y}$ and sector output at equilibrium $\boldsymbol{x}^*$ are estimated from yearly activity available in the Exiobase 3 dataset [Stadler et al., 2018], using the *Pymrio* library [Stadler, 2021]. While the data describes economic interactions across multiple countries, we design an economic equilibrium model of each country by neglecting those interactions, an extracting the blocs of matrices A and R relevant to the country under consideration. We design a distributed multiplicative Lie intervention on the activity of all 200 sectors of the database. The coefficient vector $\boldsymbol{\alpha}$ is then optimized in order to reduce the overall greenhouse gaz (GHG) emissions cumulated across sectors (estimated by one component of the stressor vector $\boldsymbol{s}$), while enforcing that



(a) (b) (c) (d)

Figure 4: (a-b) Outcome of Lie and invariant interventions on energy efficiency in the rebound model described in Fig. 1b: (a) unit price of the target sector, (b) total energy demand. (c-d) Outcome of the design of compartmentalized interventions for model of Fig. 2a. The unintervened node values (in black) are compared to invariant interventions (II, solid lines) and their corresponding Lie intervention (LI, dashed lines) (without enforcing invariance), for multiple values of the Lie interventions' parameters (u, v). (c) Value of the invariant node in compartment 1. (d) Value of the intervened node in compartment 1.

the overall employment distribution over the sectors stays closest to the non-intervened economy, in order to mitigate challenges associated to reorganizing of economic activities (e.g. mass unemployment and the need for large scale professional reorientation programs). Using the ℓ_1 norm for regularization, this leads to the following loss:

$$\mathcal{L}(\boldsymbol{u}) = \boldsymbol{c}^\top \boldsymbol{x}^{(u)} + \lambda \|\boldsymbol{e}^{(u)} - \boldsymbol{e}^*\|_1 \tag{5}$$

where $\boldsymbol{c}$ is the GHG emission intensity of each sector, and $\boldsymbol{e}^{(u)}$ and $\boldsymbol{e}^*$ the intervened and unintervened distributions of employment across sectors (estimated by entry wise multiplication of $\boldsymbol{x}^*$ with one row of matrix R). The graphs shown in Figs. 3b-3c (top), illustrate the trade off between employment preservation and GHG emission reduction achieved by varying λ for two different countries. Interestingly, the left tail of these curves reflect differences across countries, with Germany having less room than France for reducing emissions before starting reducing employment significantly. The sectors yielding the largest employment reduction also differ across countries, likely influenced both by the overall structure of each economy.

Control of rebound effects. To illustrate how *Challenge 3* of Sec. 2.1 can be addressed, we used our invariant intervention framework to prevent price rebound effects. We use a toy 3-sector model, with one energy sector and one target sector for which energy efficiency is increased, modeled by a multiplicative Lie intervention on the energy requirements coefficient of the Leontief matrix. We use the final demand of this target sector as an invariant node, and achieve this by softly intervening on it through a modification of the unit price of this sector. The soft intervention is learnt using an MLP with two hidden layers (see Appendix D for details). Fig. 4a-4b compares 3 models: unintervened (called "reference" in the figure), Lie intervened (without enforcing invariance), and invariantly intervened. For a range of one parameter left free in the Leontief matrix, the results show the invariant intervention maintains the price close to the unintervened model (Fig. 4a), while this price is much lower for the Lie intervention (due to the rebound effect). The benefit of invariance is demonstrated by the effect on the activity of overall energy demand of the economy (Fig. 4b): for the Lie intervention, the rebound through prices leads to the so-called *backfire* scenario: the actual energy savings are negative because usage increased beyond potential savings. In contrast, invariant intervention leads to a reduction of energy demand (relative to the unintervened system), as the rebound through prices is prevented.

Compartmentalized interventions design. We further implement compartmentalized interventions and show its benefits for addressing *Challenge 2* in Sec. 2.1 in muli-sector economic models. We design a two compartment Leontiev model according to Fig. 2a. We optimize two invariant interventions, u on compartment 1 and v on compartment 2, to follow the conditions of Prop. 4. The results provided in Fig. 4, show the invariant node of compartment 1 is unchanged by both values of u and v (Fig. 4c), while the intervened node of this compartment changes value only as a function of its corresponding intervention u (Fig. 4d), in a way similar to the (non-invariant) Lie intervention.

6 CONCLUSION

We introduced a differentiable soft intervention design framework for general equilibrium systems. We argue those are more likely to approximate deployable policy changes in real-world socio-economic systems and address key challenges of the transition to sustainable economies. Theoretical results and algorithmic tools are provided to design interventions with desirable invariance properties under the assumption that the considered system is in equilibrium and model parameters are known. Further work in this direction will need to address identifiability of the considered models from observational or experimental data.

References

Brandon Amos, Ivan Dario Jimenez Rodriguez, Jacob Sacks, Byron Boots, and J Zico Kolter. Differentiable MPC for end-to-end planning and control. *arXiv preprint arXiv:1810.13400*, 2018.

Holly Andersen. When to expect violations of causal faithfulness and why it matters. *Philosophy of Science*, 80(5): 672–683, 2013.

Osman Arrobbio and Dario Padovan. A vicious tenacity: The efficiency strategy confronted with the rebound effect. *Frontiers in Energy Research*, 6:114, 2018.

Shaojie Bai, J Zico Kolter, and Vladlen Koltun. Deep equilibrium models. *arXiv preprint arXiv:1909.01377*, 2019.

Tineke Blom and Joris M Mooij. Causality and independence in perfectly adapted dynamical systems. *arXiv preprint arXiv:2101.11885*, 2021.

Tineke Blom, Mirthe M van Diepen, and Joris M Mooij. Conditional independences and causal relations implied by sets of equations. *arXiv preprint arXiv:2007.07183*, 2020.

Stephan Bongers, Patrick Forré, Jonas Peters, Bernhard Schölkopf, Joris M Mooij, et al. Foundations of structural causal models with cycles and latent variables. *arXiv preprint arXiv:1611.06221*, 2016.

Paul E Brockway, Steve Sorrell, Gregor Semieniuk, Matthew Kuperus Heun, and Victor Court. Energy efficiency and economy-wide rebound effects: A review of the evidence and its implications. *Renewable and Sustainable Energy Reviews*, page 110781, 2021.

Juan Correa and Elias Bareinboim. General transportability of soft interventions: Completeness results. *Advances in Neural Information Processing Systems*, 33, 2020.

John A Dearing, Rong Wang, Ke Zhang, James G Dyke, Helmut Haberl, Md Sarwar Hossain, Peter G Langdon, Timothy M Lenton, Kate Raworth, Sally Brown, et al. Safe and just operating spaces for regional social-ecological systems. *Global Environmental Change*, 28:227–238, 2014.

Frederick Eberhardt and Richard Scheines. Interventions and causal inference. *Philosophy of science*, 74(5):981–995, 2007.

Laurent El Ghaoui, Fangda Gu, Bertrand Travacca, Armin Askari, and Alicia Tsai. Implicit deep learning. *SIAM Journal on Mathematics of Data Science*, 3(3):930–958, 2021.

Mercedes Esteban-Bravo. Computing equilibria in general equilibrium models via interior-point methods. *Computational Economics*, 23(2):147–171, 2004.

Philipp Geiger and Christoph-Nikolas Straehle. Learning game-theoretic models of multiagent trajectories using implicit layers. *arXiv preprint arXiv:2008.07303*, 2020.

Helmut Haberl, Dominik Wiedenhofer, Stefan Pauliuk, Fridolin Krausmann, Daniel B Müller, and Marina Fischer-Kowalski. Contributions of sociometabolic research to sustainability science. *Nature Sustainability*, 2 (3):173–184, 2019.

Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

William S Jevons. *The coal question*. Routledge, 1866.

John M Lee. Smooth manifolds. In *Introduction to Smooth Manifolds*, pages 1–31. Springer, 2013.

Wassily W Leontief. The structure of american economy, 1919-1939: an empirical application of equilibrium analysis. Technical report, 1951.

J. Mooij, D. Janzing, and B. Schölkopf. From ordinary differential equations to structural causal models: the deterministic case. In *Proceedings of the Twenty-Ninth Conference Annual Conference on Uncertainty in Artificial Intelligence*, pages 440–448, Corvallis, OR, 2013. AUAI Press.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

Judea Pearl. *Causality: models, reasoning and inference*, volume 29. Cambridge Univ Press, 2000.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of Causal Inference – Foundations and Learning Algorithms*. MIT Press, 2017.

Jonas Peters, Stefan Bauer, and Niklas Pfister. Causal models for dynamical systems. *arXiv preprint arXiv:2001.06208*, 2020.

Steven C Sherwood and Matthew Huber. An adaptability limit to climate change due to heat stress. *Proceedings of the National Academy of Sciences*, 107(21):9552–9555, 2010.

Konstantin Stadler. Pymrio–a python based multi-regional input-output analysis toolbox. *Journal of Open Research Software*, 9(1), 2021.

Konstantin Stadler, Richard Wood, Tatyana Bulavskaya, Carl-Johan Södersten, Moana Simas, Sarah Schmidt, Arkaitz Usubiaga, José Acosta-Fernández, Jeroen Kuenen, Martin Bruckner, et al. Exiobase 3: Developing a time series of detailed environmentally extended multi-regional input-output tables. *Journal of Industrial Ecology*, 22(3):502–515, 2018.

Homer F Walker and Peng Ni. Anderson acceleration for fixed-point iterations. *SIAM Journal on Numerical Analysis*, 49(4):1715–1735, 2011.

Grégoire Wallenborn. Rebounds are structural effects of infrastructures and markets. *Frontiers in Energy Research*, 6:99, 2018.

Kirsten Svenja Wiebe, Eivind Lekve Bjelle, Johannes Többen, and Richard Wood. Implementing exogenous scenarios in a global MRIO model for the estimation of future environmental footprints. *Journal of Economic Structures*, 7(1):1–18, 2018.

Richard Wood, Daniel Moran, Konstantin Stadler, Diana Ivanova, Kjartan Steen-Olsen, Alexandre Tisserant, and Edgar G. Hertwich. Prioritizing consumption-based carbon policy based on the evaluation of mitigation potential using input-output methods. *Journal of Industrial Ecology*, 22(3):540–552, 2018.

A ADDITIONAL BACKGROUND

A.1 SMOOTH MANIFOLDS

While many non-equivalent definitions exist for smooth manifold, we follow Lee [2013] in defining smoothness as infinite continuously differentiability of functions. A diffeomorphism is then a smooth bijection whose inverse is also smooth.

For an n-dimensional topological manifold M, an atlas is a collection of coordinate charts (U_k, φ_k) such that U_k's are open sets of M covering it, and such that the mappings $\varphi_k : U_k \mapsto \varphi_k(U_k) \subset \mathbb{R}^n$ are homeomorphisms (continuous bijection with continuous inverse). Briefly, the atlas is smooth whenever $\varphi_k \circ \varphi_n^{-1}$ is are diffeomorphisms whenever well defined, and a smooth manifold is a topological manifold associated with a maximal smooth atlas.

A smooth map $F : M \to N$ between two smooth manifolds M and N is a function such that for any chart (U, φ) and (V, ψ), $\psi \circ F \circ \varphi^{-1}$ is smooth whenever well defined.

A.2 CYCLIC CAUSAL MODELS

A classical type of hard interventions are *perfect interventions*, which replace the structural assignments of a given variable X_k by an assignment $X_k := \xi_k$, with ξ_k constant [Blom et al., 2020]. It thus eliminates the arrows in the causal graph pointing to this variables, and makes this variable deterministic.

In particular, tracing the effects of perfect interventions requires special assumptions. In contrast, soft interventions may be read from the so-called causal ordering graph, which can be built from the original SCM graph. Broadly construed, a unique causal ordering graph can be constructed with several algorithms [Blom et al., 2020]. This is a directed cluster graph that contains groups of variables connected by oriented edges (starting from single variable in a given cluster, and pointing to another cluster). By construction, the resulting graph between clusters entailed by these edges is directed and contains no cycles. As a consequence, the effect of generic soft intervention on clustered variables can be easily read from this graph.

A.3 LINK BETWEEN EQUILIBRIUM AND DYNAMIC MODELS

The equilibrium of eq. (1) can be thought of as the asymptotic value of $\boldsymbol{x}$ in a dynamic model (see Appendix A)

$$\frac{d\boldsymbol{x}}{dt} = A\boldsymbol{x} + \boldsymbol{y} - \boldsymbol{x}\,,$$

where the increase or decrease of the sectors' activity is controlled by the imbalance between their demand $A\boldsymbol{x} + y$ and their current output $\boldsymbol{x}$. More generally, any fixed point-equation can be thought of a the equilibrium value of some dynamical system, for example by considering a numerical algorithm that converges to it. However, the relationship between dynamical systems and self-consistent equation is not one to one. Notably, we can rescale the time evolution of a stable dynamical system to create many other that converge to the same self-consistent equation. Moreover, by inverting the arrow of time, we can obtain systems for which the self-consistent equation is an unstable equilibrium. As mentioned in main text, in this work we leave aside the dynamical aspects to focus on the equilibrium properties.

A.4 MRIO MODELS

Multi-regional input-output models are built based on macro-economic information, notably the one provided by the National Accounts of the countries involved in the model. The technical coefficient matrix of eq. (1) is computed from so-called *Supply and Use Tables* that form the basis of National Accounts. The unit used to measure output is frequently monetary (e.g., EUR) due to the data collection process and to allow an homogeneous treatment of the economic flows. However, under homogeneity and linearity assumptions, the output of each sector may be converted in appropriate physical units using unit prices and material flow data. Moreover, there also exist hybrid MRIO models which include information regarding physical flows in the economy (energy, raw materials, ...) and the are combined with monetary information to ensure the best level of self-consistency.

B PROOF OF MAIN TEXT RESULTS

B.1 PROOF OF PROPOSITION 1

Proof. Assuming the SSCM is locally diffeomorphic entails that the Jacobian of $\boldsymbol{x} \to \boldsymbol{x} - \mathbf{f}(\boldsymbol{x}, \boldsymbol{\theta}^{\text{ref}})$ is invertible at $\boldsymbol{x} = \boldsymbol{x}^{\text{ref}}$. Then the Jacobian of $(\boldsymbol{x}, \boldsymbol{\theta}) \to (\boldsymbol{x} - \mathbf{f}(\boldsymbol{x}, \boldsymbol{\theta}),\, \boldsymbol{\theta})$ is also invertible at $(\boldsymbol{x}^{\text{ref}}, \boldsymbol{\theta}^{\text{ref}})$ (due to its block triangular structure). Using the inverse function theorem for smooth maps between smooth manifolds [Lee, 2013, Theorem 4.5], this implies that there exists connected open neighborhoods (U, V) of $(\boldsymbol{x}^{\text{ref}}, \boldsymbol{\theta}^{\text{ref}})$ and $(\mathbf{0}, \boldsymbol{\theta}^{\text{ref}})$ such that

$$\begin{aligned} g\colon \quad & U \to V \\ & (\boldsymbol{x}, \boldsymbol{\theta}) \mapsto (\boldsymbol{x} - \mathbf{f}(\boldsymbol{x}, \boldsymbol{\theta}),\, \boldsymbol{\theta}) \end{aligned}$$

is a diffeomorphism. As a consequence, self-consistent solutions $(\boldsymbol{x}, \boldsymbol{\theta})$ in U are given by $S = g^{-1}((\{0\} \times \mathcal{T}) \cap V)$. It is a submanifold of same dimension as $\mathcal{T}$ for the following reasons:

- S is a manifold diffeomorphic to $(\{0\} \times \mathcal{T}) \cap V$ and thus has the same dimension [Lee, 2013, Theorem 2.17],
- $(\{0\} \times \mathcal{T}) \cap V$ is an open submanifold because V is open, and thus has the same dimension as $\{0\} \times \mathcal{T}$ [Lee, 2013, Proposition 5.1]
- $\{0\} \times \mathcal{T}$ has the same dimension as $\mathcal{T}$ because it is diffeomorphic to it [Lee, 2013, Propositions 5.3 and 2.17].

Let us now define the cartesian projection

$$\begin{aligned} \pi\colon \quad & U \to \mathcal{T} \\ & (\boldsymbol{x}, \boldsymbol{\theta}) \mapsto \boldsymbol{\theta}\,, \end{aligned}$$

we want to establish that there exist an open neighborhood $U_{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}^{\text{ref}}$ such that there is a unique self-consistent solution for each parameter choice in this set $\pi_{|S}$ is a smooth embedding because it is an injective smooth immersion, and is open[1], by Lee [2013, Proposition 4.22]). As a consequence $\pi(S)$ is an embedded submanifold of $\mathcal{T}$ diffeomorphic to S (by Lee [2013, Proposition 5.2]). Since we have shown that the dimension of S is the dimension of $\mathcal{T}$, then $\pi(S)$ is a submanifold of same codimension 0 (same dimension as its ambient manifold) and is thus an open submanifold of $\mathcal{T}$ (Proposition 5.1 in Lee [2013]). As a consequence, $\pi(S)$ is open, such that there is an open neighborhood of $\boldsymbol{\theta}^{\text{ref}}$ included in it. Then for any parameter chosen in this neighborhood, there is one solution to the self-consistency equation, by definition of the image. Assume there are two distinct solution for this parameter, then the mapping $(\boldsymbol{x}, \boldsymbol{\theta}) \to (\boldsymbol{x} - \mathbf{f}(\boldsymbol{x}, \boldsymbol{\theta}),\, \boldsymbol{\theta})$ would not be a diffeomorphism. □

B.2 PROOF OF PROPOSITION 2

Proof. We extend the smooth parameterization of function f by $\boldsymbol{\theta}$ to get a smooth parameterization of the intervened functional assignments by $\bar{\boldsymbol{\theta}} = (g, \boldsymbol{\theta})$. Indeed, the mapping

$$(x, \bar{\boldsymbol{\theta}}) \mapsto g \cdot f(x, \theta)$$

is smooth as a composition of the following smooth maps

$$(\boldsymbol{x}, \boldsymbol{\theta}, g) \underset{f \text{ smooth}}{\mapsto} (f(\boldsymbol{x}, \boldsymbol{\theta}), g) \underset{\varphi \text{ smooth}}{\mapsto} \varphi(g, f(\boldsymbol{x}, \boldsymbol{\theta})) = g \cdot f(\boldsymbol{x}, \boldsymbol{\theta})$$

where the smoothness of each transformation stem from the definition of SSCM and Lie interventions, respectively. Proposition 1 applied around the extend parameter $(e, \boldsymbol{\theta}^{\text{ref}})$ implies that there exists a neighborhood $U_{(e, \boldsymbol{\theta}^{\text{ref}})}$ of this point such that the intervened solution is uniquely solvable and the mapping from the extended parameter to the solution is smooth. There exists moreover a product neighborhood $U_L \times U_{\boldsymbol{\theta}} \subset U_{(e, \boldsymbol{\theta}^{\text{ref}})}$ (this is a basic property of neighborhoods on product spaces). By continuity of the partial derivative of the intervened functional assignment (due to smoothness of the Lie group action), dependency on the parents of the intervened variables is preserved in a neighborhood of the identity, such that the intervention is soft in the considered neighborhood. □

[1] $\pi_{|S}$ is open because $\pi_{|S} \circ g^{-1}_{|(\{0\} \times \mathcal{T}) \cap V}$ is a smooth submersion and thus open by Proposition 4.28 in Lee [2013], and g_S is also open as the restriction of a diffeomorphism.

B.3 PROOF OF PROPOSTION 3

Proof. The Lie intervention parameterized by u guaranties solvability of the SSCM is preserved in a neighborhood of the identity (Proposition 2), and we denote $x^{(u)}(\boldsymbol{\theta})$ the unique solution in such neighborhood, with $x^{(e)}(\boldsymbol{\theta}) = x^*(\boldsymbol{\theta})$. The Jacobian $J^{\boldsymbol{\theta}}_{x^*_{\mathbf{Pa}_k}}(\boldsymbol{\theta}^{\text{ref}})$ is the Jacobian of the mapping from the parameters $\boldsymbol{\theta}$ to the vector consisting of the parent nodes of k at equilibrium. Because this Jacobian is full column rank, there exists a neighborhood of e such that for any fixed u in it, the mapping $\boldsymbol{\theta} \mapsto \boldsymbol{x}^{(u)}_{\mathbf{Pa}_k}(\boldsymbol{\theta})$ is injective in a neighborhood of the reference parameter. As a consequence the restriction to its image is a diffeomorphic map between manifolds. Let us denote $\psi^{(u)}$ its inverse.

Consider the SSCM obtained by performing a hard intervention $x_j := x^*_j(\boldsymbol{\theta})$. Because the original SSCM is locally diffeomorphic at $(\boldsymbol{x}^{\text{ref}}, \boldsymbol{\theta}^{\text{ref}})$, $\{x_j := x^*_j(\boldsymbol{\theta})\}$ is a smooth assignment, and because additionally the Jacobian of the mapping $\boldsymbol{x}_{-j} \to \boldsymbol{x}_{-j} - \mathbf{f}_{-j}(\boldsymbol{x}_{-j}, \boldsymbol{\theta}^{\text{ref}})$ is invertible, then this hard intervened system is also locally diffeomorphic at $(\boldsymbol{x}^{\text{ref}}, \boldsymbol{\theta}^{\text{ref}})$ (exploiting the block diagonality of the Jacobian of its assignment). As a consequence, Lie intervention with parameter u on node i of this (already hard-intervened) system leads to a smooth intervened equilibrium $x^{(u)}$.

Let us recall that the partial derivative $\frac{\partial x^*_j}{\partial x_k}_{|\boldsymbol{\theta}=\boldsymbol{\theta}^{\text{ref}}}$ corresponds to the derivative with respect to the hard interventions value. The assumption $\frac{\partial x^*_j}{\partial x_k}_{|\boldsymbol{\theta}=\boldsymbol{\theta}_{\text{ref}}} \neq 0$ thus entails, by the inverse function theorem, that there exists also a smooth mapping $\phi^{(u)}$ such that $x^{(u)}_k = \phi^{(u)}(x^*_j(\boldsymbol{\theta}))$ in a neighborhood of $(e, \boldsymbol{\theta}_{\text{ref}})$. As a consequence, the mapping defined as $f^{(u)}_k = \phi^{(u)} \circ x^*_j(.) \circ \psi^{(u)}$ is a soft intervention replacing f_k achieving the same equilibrium values as the above hard-intervened system under Lie interventions, and in particular the invariance constraint $x^{(u)}_j(\boldsymbol{\theta}) = x^*_j(\boldsymbol{\theta})$. □

B.4 PROOF OF PROPOSITION 4

Proof. We proceed iteratively by adding one intervention after the next. First intervention on compartment C_1 leaves invariant the equilibrium values of the remaining compartments C_{-1} as the only node from C_1 influencing them is invariant.

Given $C_1, ..., C_n$ satisfy invariance with respect to each others interventions, consider intervening on C_{n+1}. As C_{n+1} receives only inputs from intervened upon compartments $C_1, ..., C_n$ through invariant nodes, the invariant intervention on it can be designed identical as for the non-intervened system. Moreover, invariance of the outgoing node ensures the equilibrium values of (potentially intervened upon) other compartments C_{-n+1}. □

C ADDITIONAL THEORETICAL RESULTS

C.1 MOTIVATING EXAMPLE OF SEC. 3.3

Let us restate the unintervened assignments of this example.

$$\begin{aligned} x &= \tau\,, \\ y &= (\alpha x + \beta z)\,, \\ z &= \gamma y\,. \end{aligned}$$

The equilibrium solution then writes

$$\begin{aligned} x^* &= \tau\,, \\ y^* &= \frac{\alpha\tau}{1-\beta\gamma}\,, \\ z^* &= \frac{\gamma\alpha\tau}{1-\beta\gamma}\,. \end{aligned}$$

Applying multiplicative Lie interventions on both x an y leads to the assignments

$$\begin{aligned} x &= \tau\,, \\ y &= u_y(\alpha x + \beta z)\,, \\ z &= u_z \gamma y\,. \end{aligned}$$

which leads to the intervened equilibrium

$$\begin{aligned} x^{(\boldsymbol{u})} &= \tau\,, \\ y^{(\boldsymbol{u})} &= \frac{u_y \alpha \tau}{1 - u_z u_y \beta \gamma}\,, \\ z^{(\boldsymbol{u})} &= \frac{u_y u_z \gamma \alpha \tau}{1 - u_y u_z \beta \gamma}\,. \end{aligned}$$

We can thus notice that choosing $u_z = \frac{1}{u_y}$ makes the intervened equilibrium value invariant for any choice of parameters $(\tau, \alpha, \beta, \gamma)$.

D METHODS

Following [Bai et al., 2019], we implemented implicit layers using the pyTorch library [Paszke et al., 2019]. We use Anderson acceleration with $m = 5$ previous iterations and a relaxation parameter $\beta = 2.0$ (based on preliminary analysis) to compute iteratively the fixed points of the implicit layers, for both forward and backward passes, with a maximum number of iterations of 5000 and a tolerance of .0001.

Optimization of interventions is done using backpropagation with adaptive moment estimation (Adam), with a learning rate of .001 and 10000-20000 iterations. Soft interventions to enforce invariance are learned with two hidden layer perceptrons, with 20 and 10 hidden units respectively for the first and second layers, and all layers have ReLU activation functions. At each iteration, free parameters β are sample from an factorized Gaussian distribution whose means and variances are chosen to cover the neighborhood of the reference point.

Code for the toy optimization experiments is provided in Supplementary material:

- leontiefPriceinvarUAI.py reproduces the invariant intervention experiment of paragraph “Control of rebound effects” in main text Sec. 5,
- leontiefCompartv2.py reproduces the compartmentalized intervention experiment of paragraph “Compartmentalized intervention design” in main text Sec. 5.