Infinitesimal Interconnection Variation in Nonlinear Networked Systems

Insoon Yang Samuel A. Burden S. Shankar Sastry Claire J. Tomlin

***Abstract*— We propose a novel infinitesimal variation for a nonlinear networked system's behavior when its interconnection topology changes discontinuously. We introduce a *variational derivative* of system output with respect to the connectivity, and derive an analytic formula for the derivative using an adjoint formulation. We provide bounds relating the discontinuous change in system behavior to the proposed continuous infinitesimal variation. The variational derivative can be used as the sensitivity of the system output to the interconnection topology. The *separability* of the variational derivative allows us to develop a tractable algorithm for an *interconnection pursuit* problem applicable to optimization and inference in biochemical reaction networks.**

I. INTRODUCTION

Interconnected systems naturally model the dynamic interactions of mobile sensors [1], [2], social agents [3] and biochemical species [4]. Network topology plays a critical role in determining the behavior of these systems through coupling with the nonlinear continuous dynamics. Consequently, it is important to determine the effect of variations in network connectivity.

A standard method to variate networks involves adding or deleting a subset of connections between nodes. This *discrete* variational approach has been applied to a number of problems including connectivity control [5] and network identification [6]. In some applications it is more appropriate to consider a small variation in the strength of an interconnection rather than its discrete variation. Such *continuous* variational approaches are adopted in robust consensus of interconnected systems [7] and distributed routing problems in dynamical networks [8].

Sensitivity of system output with respect to connectivity can be a useful measure of the effect of variations in the interconnection on the output. To define sensitivity with respect to a graph, we consider a variation of the interconnection and compare outputs from the two interconnection structures of interest. Parametrizing the distance between any two interconnections by a scalar parameter that varies from 0 to 1, we define the sensitivity using a *variational derivative*, namely the ratio between the output variation and the scalar parameter as the parameter tends to zero. Motivated by the important problems of yield optimization [9] and network inference [10] for biochemical reaction networks, we apply the variational derivative-based sensitivity to propose a computationally–tractable approximation of an *interconnection pursuit* problem.

In Section II-B, we introduce a family of interconnected systems that interpolate between two network topologies as a scalar parameter varies from 0 to 1. The well-posedness of this system's dynamics enables us to define the variational derivative with respect to the interconnection change by computing the limit as the parameter tends to zero, as presented in Section II-C. Building on previous work in parameter optimization [11] and optimal control [12], in Section II-C, we explicitly derive an analytic expression for the variational derivative by introducing an adjoint system. Furthermore, we bound the difference between the variational derivative and the output difference relating two interconnections, and discuss usefulness of the variational derivative as a measure of the sensitivity of the system output. In Section III, we apply the variational derivative to develop a novel algorithm for design or inference of interconnection topologies in a class of systems relevant to applications, providing a tractable approximation to optimization [9] and inference [10] in biochemical reaction networks.

II. INTERCONNECTION VARIATION

A. *Interconnected Dynamical Systems*

Suppose that we have an interconnection that is invariant over time with the representative graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ denotes the (finite) set of vertices and $\mathcal{E}$ the set of edges. Consider the *interconnected system*

$$\dot{x}(t) = f_{\mathcal{E}}(x(t)),\ x(0) = x_0 \in \mathcal{X}, \tag{1}$$

where $\mathcal{X} \subseteq \mathbb{R}^n$ is the continuous state space and $f_{\mathcal{E}} : \mathcal{X} \to \mathbb{R}^n$ is the nonlinear vector field. We restrict our analysis to a subset $\Gamma \subseteq 2^{\mathcal{V}\times\mathcal{V}}$ of all possible graphs over the vertex set $\mathcal{V}$ containing only the *feasible interconnections*.

Definition 1. *The set* $\Gamma \subseteq 2^{\mathcal{V}\times\mathcal{V}}$ *of* feasible interconnections *for system* (1) *contains all* $\mathcal{E} \in 2^{\mathcal{V}\times\mathcal{V}}$ *such that the vector field* $f_{\mathcal{E}}(\cdot)$ *is twice differentiable, has a continuous second derivative, and is globally Lipshitz continuous in* $\mathcal{X}$.

Throughout this paper we will assume a continuously differentiable output function $h : \mathcal{X} \to \mathbb{R}$ is associated with the system (1) and we will let $y(t) = h(x(t))$ denote the output trajectory associated with any solution $x : \mathbb{R} \to \mathcal{X}$ of (1). Our aim is to investigate how y changes in response to a variation in $\mathcal{E}$. The set $\mathcal{V}$ of vertices will remain fixed even if a vertex does not connect to any other vertices.

This work was supported by NCI PS-OC program under grant number 29949-31150-44-IQPRJ1-IQCLT.

I. Yang, S. A. Burden, S. S. Sastry, and C. J. Tomlin are with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720, USA {iyang, sburden, sastry, tomlin}@eecs.berkeley.edu

C. J. Tomlin is also with the Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

Preprint submitted to 52nd IEEE Conference on Decision and Control .
Received March 11, 2013.

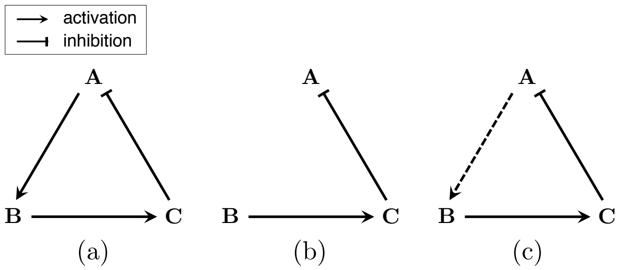

Fig. 1: Graphs with (a) $\mathcal{E}$; and (b) $\tilde{\mathcal{E}}$. The interconnections $\mathcal{E}$ and $\widetilde{\mathcal{E}}$ determine the vector field $f_{\mathcal{E}}$ and $f_{\widetilde{\mathcal{E}}}$, respectively. The interactions among nodes with the ϵ-variational vector field $f_{\epsilon(\mathcal{E},\widetilde{\mathcal{E}})}$ can be considered as (c).

We begin by introducing a family of dynamical systems that vary continuously between two interconnection topologies as a parameter ϵ is varied from 0 to 1. Subsequently, we develop an adjoint-based technique to compute the infinitesimal variation of trajectories to interconnection variations.

B. Parameterizing Interconnection Variation

For any pair $\mathcal{E}, \widetilde{\mathcal{E}} \in \Gamma$, we construct a vector field called the *variation from $\mathcal{E}$ to $\widetilde{\mathcal{E}}$*,

$$f_{(\mathcal{E},\widetilde{\mathcal{E}})} = f_{\widetilde{\mathcal{E}}} - f_{\mathcal{E}}, \tag{2}$$

and use this to define the *ϵ–variation from $\mathcal{E}$ to $\widetilde{\mathcal{E}}$*,

$$f_{\epsilon(\mathcal{E},\widetilde{\mathcal{E}})} = f_{\mathcal{E}} + \epsilon f_{(\mathcal{E},\widetilde{\mathcal{E}})}, \quad \epsilon \in [0,1]. \tag{3}$$

An illustration of this variation is shown in Fig. 1.

Example 1. *Fig. 1 contains an example of interconnections $\mathcal{E}$ and $\widetilde{\mathcal{E}}$ with an interpretation of $f_{\epsilon(\mathcal{E},\widetilde{\mathcal{E}})}$. The original graph in (a) consists of $\mathcal{V} = \{\mathbf{A},\mathbf{B},\mathbf{C}\}$ and $\mathcal{E} = \{(\mathbf{A},\mathbf{B}),(\mathbf{B},\mathbf{C}),(\mathbf{C},\mathbf{A})\}$. The interconnection in (b) is given by $\widetilde{\mathcal{E}} = \{(\mathbf{B},\mathbf{C}),(\mathbf{C},\mathbf{A})\}$. The ϵ-variational vector field $f_{\epsilon(\mathcal{E},\widetilde{\mathcal{E}})}$ takes into account the small degradation of the interaction via $\{(\mathbf{A},\mathbf{B})\}$ as depicted in (c). We will study this interconnection structure more detail in Section IV.*

Let $\mathcal{E}, \widetilde{\mathcal{E}} \in \Gamma$ and consider the dynamics of the ϵ–variation from $\mathcal{E}$ to $\widetilde{\mathcal{E}}$,

$$\dot{x}^\epsilon(t) = f_{\epsilon(\mathcal{E},\widetilde{\mathcal{E}})}\left(x^\epsilon(t)\right), \quad x^\epsilon(0) = x_0 \in \mathcal{X}, \tag{4}$$

which we call the *ϵ-variational system* associated with $(\mathcal{E}, \widetilde{\mathcal{E}})$. With $\Delta x^\epsilon(\cdot) = x^\epsilon(\cdot) - x(\cdot)$ denoting the difference between trajectories satisfying (1) and (4) with the same initial condition $x^\epsilon(0) = x(0) = x_0 \in \mathcal{X}$,

$$\begin{aligned} \dot{\Delta x^\epsilon} &= f_{\epsilon(\mathcal{E},\widetilde{\mathcal{E}})}\left(x^\epsilon\right) - f_{\mathcal{E}}(x) \\ &= f_{\mathcal{E}}(\Delta x^\epsilon + x) - f_{\mathcal{E}}(x) + \epsilon f_{(\mathcal{E},\widetilde{\mathcal{E}})}\left(\Delta x^\epsilon + x\right). \end{aligned} \tag{5}$$

Taking the Taylor approximation,

$$\dot{\Delta x^\epsilon} = \frac{\partial f_{\mathcal{E}}(x)}{\partial x}\Delta x^\epsilon + \epsilon f_{(\mathcal{E},\widetilde{\mathcal{E}})}\left(\Delta x^\epsilon + x\right) + H(\Delta x^\epsilon, x) \tag{6}$$

where $H := (H_1, \cdots, H_n)$ and H_i denotes the higher-order terms in the Taylor expansion of $(f_{\mathcal{E}})_i(\Delta x^\epsilon(t) + x(t))$, i.e. by applying the mean value theorem [13],

$$H_i(\Delta x^\epsilon(t), x(t)) = \int_0^1 (1-s)\Delta x^\epsilon(t)^\top D_x^2 (f_{\mathcal{E}})_i(x(t) + s\Delta x^\epsilon(t))\Delta x^\epsilon(t)\, ds.$$

Since the vector fields in (4) and (5) are Lipschitz continuous for all $\epsilon \in [0,1]$, they generate well-defined and bounded trajectories in a bounded time interval; this follows from Proposition 5.6.5 in [14].

Lemma 1. *For all $\epsilon \in [0,1]$, there exist unique trajectories $x^\epsilon(\cdot)$ and $\Delta x^\epsilon(\cdot)$ satisfying the dynamics in (4) and (6), respectively. In addition, $\|x^\epsilon(t)\|$ and $\|\Delta x^\epsilon(t)\|$ are bounded by some constant independent of ϵ for any $t \in (0,T]$.*

The next question to address is *how the ϵ-variational system behaves as ϵ tends to zero as compared to the original system.* The following Lemma shows that the difference $\Delta x^\epsilon = x^\epsilon - x$ is Lipschitz continuous in ϵ; a version of this result appears as Lemma 5.6.7 in [14].

Lemma 2. *There exists $L \in \mathbb{R}$ such that for all $t \in [0,T]$*

$$\|\Delta x^\epsilon(t)\| = \|x^\epsilon(t) - x(t)\| \le L\epsilon.$$

This Lemma allows us to directly deduce, for all $t \in [0,T]$,

$$\lim_{\epsilon \to 0^+} \|\Delta x^\epsilon(t)\| = 0, \quad \lim_{\epsilon \to 0^+} \frac{1}{\epsilon}\|\Delta x^\epsilon(t)\|^2 = 0.$$

The previous Lemmas with the dominated convergence theorem [15] also yields the following.

Corollary 1. *For any $t \in (0,T]$,*

$$\lim_{\epsilon \to 0^+} \frac{1}{\epsilon}\int_0^t \|\Delta x^\epsilon(s)\|^2\, ds = 0.$$

Lemmas 1 and 2 together with Corollary 1 imply that for $\epsilon \in [0,1]$ and $\mathcal{E}, \widetilde{\mathcal{E}} \in \Gamma$ the ϵ–variation from $\mathcal{E}$ to $\widetilde{\mathcal{E}}$ is well-posed in the sense of Hadamard [16].

We now compute the infinitesimal change in the output behavior due to the ϵ-variation of the interconnection as ϵ tends to zero.

C. Infinitesimal Interconnection Variation

We seek a first-order approximation to the change in system behavior caused by a variation of interconnection topology. Such an approximation is an intuitively appealing generalization of the *sensitivity* in continuous systems. Defining first-order variations with respect to the connectivity structure requires a distance metric defined over interconnection topologies. We define the distance between any two feasible connections $\mathcal{E}$ and $\widetilde{\mathcal{E}}$ as 1 and the distance between $\mathcal{E}$ and its ϵ-variation $\epsilon(\mathcal{E}, \widetilde{\mathcal{E}})$ as $\epsilon \in [0,1]$. We then use the the corresponding ϵ-variational system (4) to obtain the ϵ-variation of the system output $y^\epsilon := h(x^\epsilon)$. As ϵ tends to zero, we have a well-defined derivative of the output with respect to the interconnection, which we call the *variational derivative* of the output from $\mathcal{E}$ to $\widetilde{\mathcal{E}}$.

Preprint submitted to 52nd IEEE Conference on Decision and Control .
Received March 11, 2013.

Definition 2. *For $\mathcal{E}, \widetilde{\mathcal{E}} \in \Gamma$ we define the* variational derivative from $\mathcal{E}$ to $\widetilde{\mathcal{E}}$ *of the output as*

$$D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(t) := \lim_{\epsilon\to 0^+} \frac{1}{\epsilon}\left(h(x^\epsilon(t)) - h(x(t))\right) \tag{7}$$

where x^ϵ solves (4) *and x solves* (1) *with $x^\epsilon(0) = x(0)$.*

Qualitatively, the variational derivative $D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(t)$ measures the sensitivity of the output with respect to the network change from $\mathcal{E}$ to $\widetilde{\mathcal{E}}$. We aim to prove that the variational derivative (7) is well-defined and bounded for $t \in [0,T]$. We will show this by explicitly deriving an analytic formula of $D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(t)$ using the *adjoint system* of (1) at time t:

$$\begin{aligned} &-\dot{\mu}(s) = \frac{\partial f_{\mathcal{E}}(x(s))}{\partial x}^\top \left(\mu(s) - \frac{\partial h(x(t))}{\partial x}^\top\right),\ s \in [0,t] \\ &\mu(t) = 0. \end{aligned} \tag{8}$$

Here, $\mu(s)$ is called the *adjoint state*. Unique solutions to (8) exist for all $t \in [0,T]$ and $\mathcal{E} \in \Gamma$ since the definition of feasible interconnections guarantees continuous differentiability of $f_{\mathcal{E}}$ and boundedness of $x(s)$ for $s \in [0,t]$.

Theorem 1. *The variational derivative* (7) *can be obtained as*

$$D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(t) = \int_0^t \left[\frac{\partial h(x(t))}{\partial x} - \mu(s)^\top\right] f_{(\mathcal{E},\widetilde{\mathcal{E}})}(x(s))\, ds, \tag{9}$$

where x and μ solve (1) *and* (8)*, respectively, and $f_{(\mathcal{E},\widetilde{\mathcal{E}})}$ is defined in* (2)*. Hence, the variational derivative is well-defined and bounded for any $t \in [0,T]$.*

Proof. Let $\boldsymbol{O} := O(\int_0^t \|\Delta x^\epsilon(s)\|^2 ds) + O(\epsilon^2)$ denote high order terms. We then have $\lim_{\epsilon\to 0^+} \boldsymbol{O}/\epsilon = 0$ due to Lemma 2 and Corollary 1. Note that $\epsilon f_{(\mathcal{E},\widetilde{\mathcal{E}})}(x(t) + \Delta x^\epsilon(t)) = \epsilon f_{(\mathcal{E},\widetilde{\mathcal{E}})}(x(t)) + \boldsymbol{O}$, we can rewrite system (6) as

$$\dot{\Delta x^\epsilon}(t) = \frac{\partial f_{\mathcal{E}}(x(t))}{\partial x}\Delta x^\epsilon(t) + \epsilon f_{(\mathcal{E},\widetilde{\mathcal{E}})}(x(t)) + \boldsymbol{O}, \tag{10}$$

which implies

$$\Delta x^\epsilon(t) = \int_0^t \frac{\partial f_{\mathcal{E}}(x(s))}{\partial x}\Delta x^\epsilon(s) + \epsilon f_{(\mathcal{E},\widetilde{\mathcal{E}})}(x(s))\, ds + \boldsymbol{O}.$$

We now consider the difference

$$h(x^\epsilon(t)) - h(x(t)) = \frac{\partial h(x(t))}{\partial x}\Delta x^\epsilon(t) + \boldsymbol{O}, \tag{11}$$

in which we use the Taylor expansion of $h(\cdot)$. We now take the inner product of both sides of (10) with the adjoint state μ and subtract it from the difference (11) to obtain

$$\begin{aligned} &h(x^\epsilon(t)) - h(x(t)) \\ &= \frac{\partial h(x(t))}{\partial x}\int_0^t \frac{\partial f_{\mathcal{E}}(x(s))}{\partial x}\Delta x^\epsilon(s) + \epsilon f_{(\mathcal{E},\widetilde{\mathcal{E}})}(x(s))\, ds + \boldsymbol{O} \\ &\quad - \int_0^t \mu^\top\left[-\dot{\Delta x^\epsilon} + \frac{\partial f_{\mathcal{E}}(x)}{\partial x}\Delta x^\epsilon + \epsilon f_{(\mathcal{E},\widetilde{\mathcal{E}})}(x) + \boldsymbol{O}\right] ds \\ &= \epsilon\int_0^t \lambda(s,t) f_{(\mathcal{E},\widetilde{\mathcal{E}})}(x(s))\, ds + \boldsymbol{O} \\ &\quad + \int_0^t \left(-\dot{\mu}(s)^\top + \lambda(s,t)\frac{\partial f_{\mathcal{E}}(x(s))}{\partial x}\right)\Delta x^\epsilon(s)\, ds, \end{aligned} \tag{12}$$

in which we let $\lambda(s,t) = \left(\frac{\partial h(x(t))}{\partial x} - \mu(s)^\top\right)$ for notational simplicity, In the second equality, we used integration by parts $\int_0^t \mu(s)^\top \dot{\Delta x^\epsilon}(s)\, ds = \mu(t)^\top\Delta x^\epsilon(t) - \mu(0)^\top\Delta x^\epsilon(0) - \int_0^t \dot{\mu}(s)^\top \Delta x^\epsilon(s)\, ds$ with $\mu(t) = 0$ and $\Delta x^\epsilon(0) = 0$. The second integral in the righthand side of the last equality equals zero due to the definition of μ in (8). The limit

$$\begin{aligned} &\lim_{\epsilon\to 0^+} \frac{h(x^\epsilon(t)) - h(x(t))}{\epsilon} \\ &\qquad = \lim_{\epsilon\to 0^+}\left[\int_0^t \lambda(s,t) f_{(\mathcal{E},\widetilde{\mathcal{E}})}(x(s))\, ds + \frac{\boldsymbol{O}}{\epsilon}\right] \end{aligned}$$

yields the formula in (9) because $\lim_{\epsilon\to 0^+} \boldsymbol{O}/\epsilon = 0$. We also notice that $D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(t)$ is unique and bounded due to the uniqueness and boundedness of x and μ. $\square$

From the proof, for $y^\epsilon := h(x^\epsilon)$ we have the estimate

$$y^\epsilon(t) = y(t) + \epsilon D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(t) + O(\epsilon^2)$$

for sufficiently small ϵ, which is analogous to the Taylor expansion of $y^\epsilon(t)$ with respect to ϵ. For the important case $\epsilon = 1$, we have the following.

Proposition 1. *Let $\Delta_{(\mathcal{E},\widetilde{\mathcal{E}})}y := y^1 - y$ denote the output difference between* (5) *and* (1) *when $\epsilon = 1$. Then there exists $K \in \mathbb{R}$ such that for all $s \in (0,t)$ we have*

$$|\Delta_{(\mathcal{E},\widetilde{\mathcal{E}})}y(t) - D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(t)| \le K\int_0^t \|\Delta x^1(s)\|\, ds.$$

Proof. Choose $\epsilon = 1$ in (6). Then, instead of (10), we have

$$\dot{\Delta x^\epsilon}(s) = \frac{\partial f_{\mathcal{E}}(x(s))}{\partial x}\Delta x^\epsilon(s) + \epsilon f_{(\mathcal{E},\widetilde{\mathcal{E}})}(x(s)) + \tilde{\boldsymbol{O}}(s)$$

for small $\|\Delta x^1(s)\|$, where $\tilde{\boldsymbol{O}}(s) = O(\|\Delta x^1(s)\|)$. Therefore, (11) should be rewritten with $\tilde{\boldsymbol{O}}$ as follows:

$$\begin{aligned} &h(x^1(t)) - h(x(t)) \\ &= \int_0^t \left(\frac{\partial h(x(t))}{\partial x} - \mu^\top(s)\right) f_{(\mathcal{E},\widetilde{\mathcal{E}})}(x(s))\, ds + \int_0^t \tilde{\boldsymbol{O}}(s) ds \\ &= D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(t) + \int_0^t \tilde{\boldsymbol{O}}(s) ds. \end{aligned}$$

Recall that $\Delta_{(\mathcal{E},\widetilde{\mathcal{E}})}y(t) = h(x^1(t)) - h(x(t))$ by definition, we have the desired estimate. $\square$

This estimate leads us to interpret the variational derivative as measuring the output sensitivity with respect to the interconnection change $(\mathcal{E}\setminus\widetilde{\mathcal{E}})\cup(\widetilde{\mathcal{E}}\setminus\mathcal{E})$ in the following sense. Let $\epsilon \in [0,1]$ parameterize the intensity of the interaction among nodes via $(\mathcal{E}\setminus\widetilde{\mathcal{E}})\cup(\widetilde{\mathcal{E}}\setminus\mathcal{E})$. Then the variational derivative yields the sensitivity of the output with respect to the perturbation that slightly degrades the efficiency of $\mathcal{E}\setminus\widetilde{\mathcal{E}}$ and infinitesimally generates the interaction through $\widetilde{\mathcal{E}}\setminus\mathcal{E}$. For this reason, we believe the variational derivative will be useful for analyzing output sensitivity in nonlinear biochemical and power system models.

D. Computing Infinitesimal Interconnection Variation

Computation of the variational derivative $D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(t)$ using (9) requires the solutions of the system (1) and its adjoint (8). Due to nonlinearity, it is not generally possible to find an analytic expression for solutions of (1) (and hence (8)). Algorithm 1 contains a straightforward algorithm to numerically approximate the variational derivative.

We discretize $[0,T]$ into $N \in \mathbb{N}$ equally–spaced time intervals, then apply the Forward Euler algorithm to obtain the numerical approximation x_k to (1) for each $k \in \{0,\dots,N\}$. To evaluate the variational derivative at time $t = iT/N$, we apply Forward Euler to the adjoint dynamics (8) backward in time to obtain μ_k for each $k \in \{0,\dots,i\}$, then numerically integrate (i.e. sum) the formula in (9).

Algorithm 1: Variational derivative

Initialization:
Given $x_0 \in \mathbb{R}^n$ and $\mathcal{E},\widetilde{\mathcal{E}} \in \Gamma$;

Approximate solution of (1)**:**
for $k = 0 : N-1$ **do**
  $x_{k+1} \leftarrow x_k + \frac{1}{N} f_{\mathcal{E}}(x_k)$;
end

Variational derivative:
for $i = 0 : N$ **do**
  Approximate solution of (8)**:**
  $\mu_i \leftarrow 0$;
  for $k = i-1 : 0$ **do**
    $\mu_k \leftarrow \mu_{k+1} - \frac{1}{N}\frac{\partial f_{\mathcal{E}}(x_k)}{\partial x}^\top \left(\mu_{k+1} - \frac{\partial h(x_i)}{\partial x}^\top\right)$;
  end
  Approximation of (9)**:**
  $D_{(\mathcal{E},\widetilde{\mathcal{E}})}y_i \leftarrow \frac{T}{N}\sum_{k=0}^{i}\left(\frac{\partial h(x_k)}{\partial x} - \mu_k^\top\right) f_{(\mathcal{E},\widetilde{\mathcal{E}})}(x_k)$;
end

E. Infinitesimal Interconnection Variation at Equilibria

An important special case concerns the sensitivity of the system (1) at an equilibrium $\xi \in \mathcal{X}$ where $f_{\mathcal{E}}(\xi) = 0$. In applications, exponentially stable equilibria may correspond to a setpoint in a biochemical reaction network or nominal load conditions in an electrical power system, and there exists a rich set of tools to ensure existence and stability of equilibria for such interconnected systems [17], [18].

Consider an exponentially stable equilibrium trajectory x for interconnection $\mathcal{E}$ of (1) so that $x(t) = \xi$ for all $t \in [0,T]$ and $\operatorname{Re}\lambda < 0$ for all $\lambda \in \operatorname{spec}\frac{\partial f_{\mathcal{E}}(\xi)}{\partial x}$. Then we have $\frac{\partial h(x(t))}{\partial x} =: b$, $f_{\mathcal{E}}(x(t)) = 0$, and $\frac{\partial f_{\mathcal{E}}(x(t))}{\partial x} =: A$ independent of $t \in [0,T]$, whence the adjoint satisfies the linear time–invariant differential equation

$$-\dot{\mu}(s) = A^\top\left(\mu(s) - b^\top\right),\ s \in [0,t], \quad \mu(t) = 0 \tag{13}$$

with solution $\mu(s) = -e^{A^\top(t-s)}b^\top + b^\top$, $s \in [0,t]$. This leads to the simplified expression for $D_{(\mathcal{E},\widetilde{\mathcal{E}})}y$, for $t \in [0,T]$,

$$\begin{aligned} D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(t) &= \int_0^t \left(b - \mu(s)^\top\right)\left(f_{\widetilde{\mathcal{E}}}(\xi) - f_{\mathcal{E}}(\xi)\right) ds \\ &= b\int_0^t e^{A(t-s)} ds\, f_{\widetilde{\mathcal{E}}}(\xi) \\ &= -bA^{-1}\left(I - e^{At}\right) f_{\widetilde{\mathcal{E}}}(\xi). \end{aligned} \tag{14}$$

For sufficiently large T, we find that

$$D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(T) = -\frac{\partial h(\xi)}{\partial x}\left(\frac{\partial f_{\mathcal{E}}(\xi)}{\partial x}\right)^{-1} f_{\widetilde{\mathcal{E}}}(\xi) + O(e^{aT}). \tag{15}$$

where $a = \max\{\operatorname{Re}\lambda : \lambda \in \operatorname{spec}A\} < 0$ by assumption. This expression shows that the variational derivative at equilibria is most sensitive to changes in states that converge slowly as measured by the output function. To see this, consider the case where $A = \frac{\partial f_{\mathcal{E}}(\xi)}{\partial x}$ is diagonal and the output function measures the i^{th} state. In this case, $A = \operatorname{diag}(a_1,\dots,a_n)$ where $\operatorname{Re} a_i < 0$ and $\frac{\partial h(\xi)}{\partial x} = e_i^\top$ where e_i is the i^{th} standard Euclidean basis vector. For large T we see that $D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(T) \simeq -\frac{\overline{a_i}}{|a_i|^2} e_i^\top f_{\widetilde{\mathcal{E}}}(\xi)$.

III. Interconnection Pursuit

Motivated by applications, we consider an optimization problem over interconnection topologies. Specifically, given an initial state $x(0) \in \mathcal{X}$, interconnection topology $\mathcal{E} \in \Gamma$, and time $\tau \in [0,T]$, we seek a new interconnection $\widetilde{\mathcal{E}} \in \Gamma$ that maximizes $\Delta_{\mathcal{E},\widetilde{\mathcal{E}}}y(\tau) := h(x_{\widetilde{\mathcal{E}}}(\tau)) - h(x_{\mathcal{E}}(\tau))$, the difference in the outputs at time τ.

Optimization 1 (Interconnection Pursuit)**.**

$$\max_{\widetilde{\mathcal{E}}\in\Gamma} \Delta_{(\mathcal{E},\widetilde{\mathcal{E}})}y(\tau) \tag{16}$$

This single optimization subsumes several important problems in the context of biochemical reaction networks: design of *in vitro* circuits that maximize yield of a desirable species x_i can be obtained by choosing $h(x) = x_i$; inference of the *in vivo* network most compatible with observation data $\eta(\tau)$ can be achieved using $h(x) = -\|x - \eta(\tau)\|$. Unfortunately, Optimization 1 is generally nonconvex and NP-hard, whence we seek a tractable approximation.

A. Additive Interconnections

Motivated by applications, we consider the special case where interconnections enter additively in (1). This means that for each $\mathcal{E} \in \Gamma$ the vector field $f_{\mathcal{E}}$ is obtained by adding one term for each edge $e \in \mathcal{E}$, and we assume all possible

Preprint submitted to 52nd IEEE Conference on Decision and Control .
Received March 11, 2013.

interconnections are feasible for the set $\mathcal{V}$ of nodes, i.e. $\Gamma = 2^{\mathcal{V}\times\mathcal{V}}$. In the following definition, we regard the vector field as a function $f : \mathcal{X} \times \Gamma \to \mathbb{R}^n$ where $f(\cdot, \mathcal{E}) = f_{\mathcal{E}}(\cdot)$.

Definition 3. *The vector field $f : \mathcal{X} \times \Gamma \to \mathbb{R}^n$ has* additive interconnections *if for all $\mathcal{E} \in \Gamma$*

$$f_{\mathcal{E}} = f_{\emptyset} + \sum_{e\in\mathcal{E}} \left(f_{\{e\}} - f_{\emptyset}\right), \tag{17}$$

where $f_{\emptyset}$ denotes the vector field with no interconnections.

The vector field with additive interconnections implies that the variational dervative is *separable over edges*, i.e.,

$$\begin{aligned} D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(t) = &\sum_{e\in\widetilde{\mathcal{E}}\setminus\mathcal{E}} D_{(\mathcal{E},\mathcal{E}\cup\{e\})}y(t) \\ &+ \sum_{e\in\mathcal{E}\setminus\widetilde{\mathcal{E}}} D_{(\mathcal{E},\mathcal{E}\setminus\{e\})}y(t). \end{aligned} \tag{18}$$

As we will see in Section III-B, this decomposition will allow us to solve an approximate version of the interconnection pursuit problem with an algorithm that only requires the evaluation of $D_{(\mathcal{E},\mathcal{E}\cup\{e\})}y(\tau)$ for $e \notin \mathcal{E}$ and $D_{(\mathcal{E},\mathcal{E}\setminus\{e\})}y(\tau)$ for $e \in \mathcal{E}$.

Proposition 2. *If $f : \mathcal{X} \times \Gamma \to \mathbb{R}^n$ has additive interconnections, then for all $\mathcal{E}, \widetilde{\mathcal{E}} \in \Gamma$ the variational derivative $D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(t)$ satisfies* (18).

Proof. Since f has additive interconnections, the variation $f_{(\mathcal{E},\widetilde{\mathcal{E}})}$ is given by

$$\begin{aligned} f_{(\mathcal{E},\widetilde{\mathcal{E}})} &= f_{\widetilde{\mathcal{E}}} - f_{\mathcal{E}} = \sum_{\widetilde{e}\in\widetilde{\mathcal{E}}}(f_{\{\widetilde{e}\}} - f_{\emptyset}) - \sum_{e\in\mathcal{E}}(f_{\{e\}} - f_{\emptyset}) \\ &= \sum_{\widetilde{e}\in\widetilde{\mathcal{E}}\setminus\mathcal{E}} (f_{\{\widetilde{e}\}} - f_{\emptyset}) + \sum_{e\in\mathcal{E}\setminus\widetilde{\mathcal{E}}} (-f_{\{e\}} + f_{\emptyset}) \\ &= \sum_{\widetilde{e}\in\widetilde{\mathcal{E}}\setminus\mathcal{E}} (f_{\mathcal{E}\cup\{\widetilde{e}\}} - f_{\mathcal{E}}) + \sum_{e\in\mathcal{E}\setminus\widetilde{\mathcal{E}}} (f_{\mathcal{E}\setminus\{e\}} - f_{\mathcal{E}}). \end{aligned}$$

Now since the variation $f_{(\mathcal{E},\widetilde{\mathcal{E}})}$ enters linearly into (9) and letting $\lambda(s,t) = \left(\frac{\partial h(x(t))}{\partial x} - \mu(s)^\top\right)$ for notational simplicity,

$$\begin{aligned} D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(t) &= \int_0^t \lambda(s,t) f_{(\mathcal{E},\widetilde{\mathcal{E}})}\left(x(s)\right) ds \\ &= \sum_{\widetilde{e}\in\widetilde{\mathcal{E}}\setminus\mathcal{E}} \int_0^t \lambda(s,t) f_{(\mathcal{E},\mathcal{E}\cup\{\widetilde{e}\})}\left(x(s)\right) ds \\ &\quad + \sum_{e\in\mathcal{E}\setminus\widetilde{\mathcal{E}}} \int_0^t \lambda(s,t) f_{(\mathcal{E},\mathcal{E}\setminus\{e\})}\left(x(s)\right) ds, \end{aligned}$$

which is equivalent to (18). □

When the vector field has additive interconnections, we can bound the difference between $D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(T)$ and $\Delta_{(\mathcal{E},\widetilde{\mathcal{E}})}y(T)$ that are associated with equilibria to obtain a more explicit expression than the estimate in Proposition 1.

Proposition 3. *Suppose that the vector field f has additive interconnections. Let ξ and $\tilde{\xi}$ be exponentially stable equilibria of $\dot{x} = f_{\mathcal{E}}(x)$ and $\dot{x} = f_{\widetilde{\mathcal{E}}}(x)$, respectively. Then*

$$\begin{aligned} &|\Delta_{(\mathcal{E},\widetilde{\mathcal{E}})}y(T) - D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(T)| \\ &\quad\leq \left\| \frac{\partial h(\xi)}{\partial x}\left(\frac{\partial f_{\mathcal{E}}(\xi)}{\partial x}\right)^{-1} \frac{\partial f_{(\mathcal{E},\widetilde{\mathcal{E}})}(\xi)}{\partial x}\right\| \|\xi - \tilde{\xi}\| \\ &\quad\quad + O(e^{aT}) + O(\|\xi - \tilde{\xi}\|^2) \end{aligned}$$

where $a = \max\left\{\operatorname{Re}\lambda : \lambda \in \operatorname{spec}\frac{\partial f_{\mathcal{E}}(\xi)}{\partial x}\right\} < 0$.

The proof is contained in the Appendix. Proposition 3 implies that for sufficiently large T and small $\|\xi - \tilde{\xi}\|$, the variation $D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(T)$ approximates the actual difference in system behavior $\Delta_{(\mathcal{E},\widetilde{\mathcal{E}})}y(T)$ as the interconnection changes from $\mathcal{E}$ to $\widetilde{\mathcal{E}}$. Although the relationship between $\Delta_{(\mathcal{E},\widetilde{\mathcal{E}})}y(\tau)$ and $D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(\tau)$ for any $\tau \in (0,T]$ warrents further study beyond Proposition 1, we will use $D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(\tau)$ in the *approximate interconnection pursuit* problem proposed below. Additive interconnections and the separability of the variational derivative enable us to solve this approximate pursuit problem efficiently relative to the original (NP-hard) interconnection pursuit problem.

B. Approximate Interconnection Pursuit

The analysis in the previous section suggests a tractable approximation to the interconnection pursuit problem (16) for systems with additive interconnections. We assume that $\tau \in (0,T]$ is fixed, all possible interconnections are feasible, and the vector field has additive interconnections.

Optimization 2 (Approximate Interconnection Pursuit)**.**

$$\max_{\widetilde{\mathcal{E}}\in\Gamma} D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(\tau) \tag{19}$$

To study how efficiently we can solve this problem, we first let $\mathcal{E} := \{e_1^-, \cdots, e_L^-\}$ and $\mathcal{E}^* \setminus \mathcal{E} := \{e_1^+, \cdots, e_{M-L}^+\}$ and , where $\mathcal{E}^*$ is the maximal interconnection (i.e., arbitrary two nodes are connected) and $L := |\mathcal{E}|$ and $M := |\mathcal{E}^*|$. We also introduce the following vectors of variational derivatives

$$\begin{aligned} c^+ &:= \left(D_{(\mathcal{E},\mathcal{E}\cup\{e_1^+\})}y(\tau), \cdots, D_{(\mathcal{E},\mathcal{E}\cup\{e_{M-L}^+\})}y(\tau)\right), \\ c^- &:= \left(D_{(\mathcal{E},\mathcal{E}\setminus\{e_1^-\})}y(\tau), \cdots, D_{(\mathcal{E},\mathcal{E}\setminus\{e_L^-\})}y(\tau)\right), \end{aligned} \tag{20}$$

and let $c := [c^+, c^-] \in \mathbb{R}^M$. Due to the separability of the variational derivative, we have

$$D_{\mathcal{E},\widetilde{\mathcal{E}}}y(\tau) = c^\top \alpha$$

with $\alpha \in \mathbb{R}^M$ such that

$$\alpha_i = \begin{cases} 1 & \text{if } e_i^+ \in \widetilde{\mathcal{E}} \text{ or } e_i^- \notin \widetilde{\mathcal{E}}, \\ 0 & \text{otherwise.} \end{cases} \quad i = 1, \cdots, M.$$

Optimization 2 is then equivalent to

$$\begin{aligned} &\max_{\alpha\in\mathbb{R}^M} \ c^\top \alpha \\ &\text{subject to } \alpha_i \in \{0,1\}, \quad i = 1, \cdots, M. \end{aligned}$$

This problem has an explicit solution given by

$$\alpha_i = \begin{cases} 1 & \text{if } c_i > 0, \\ 0 & \text{otherwise.} \end{cases} \qquad i = 1, \cdots, M,$$

which motivates the following tractable algorithm:

Algorithm 2: Approximate interconnection pursuit

1 **Initialization:**
2 $\widetilde{\mathcal{E}} \leftarrow \mathcal{E}$;
3 Fix $\tau \in (0, T]$;
4 Compute the vectors of variational derivates c^+ and c^- given by (20);
5 **Interconnection pursuit:**
6 **for** $i = 1 : M - L$ **do**
7   **if** $c_i^+ > 0$ **then**
8         $\widetilde{\mathcal{E}} \leftarrow \widetilde{\mathcal{E}} \cup \{e_i^+\}$;
9   **end**
10 **end**
11 **for** $i = 1 : L$ **do**
12   **if** $c_i^- > 0$ **then**
13         $\widetilde{\mathcal{E}} \leftarrow \widetilde{\mathcal{E}} \setminus \{e_i^-\}$;
14   **end**
15 **end**

Note that lines 6–15 can be completely parallelized because the order of the additions and the deletions does not matter. In practice, we choose a *threshold* $\delta \geq 0$, and add $e_i^+ \notin \mathcal{E}$ if $D_{(\mathcal{E}, \mathcal{E} \cup \{e_i^+\})} y(\tau) > \delta$ and remove $e_i^- \in \mathcal{E}$ if $D_{(\mathcal{E}, \mathcal{E} \setminus \{e_i^-\})} y(\tau) > \delta$ to account for error introduced by the approximation (3). While the original interconnection pursuit problem requires us to search all 2^M possible interconnections to find a global optimal solution, the approximate interconnection pursuit problem can be very efficiently solved with Algorithm 2 that has $O(M)$ complexity.

IV. Example

The three-species biochemical circuit has played an important role in illuminating fundamental properties of complex signaling networks such as biochemical adaptation [19] and dynamic correlations in biochemical noise [20]. In this section, we consider the biochemical network in Figure 2 (a) and investigate which edges should be deleted to maximize the increase in the concentration of active species $p\mathbf{B}$. We first represent the biochemical circuit in Figure 2 (a) as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$ and $\mathcal{E} = \{e_1, e_2, e_3\}$, where $e_1 = (\mathbf{A}, \mathbf{B})$, $e_2 = (\mathbf{B}, \mathbf{C})$ and $e_3 = (\mathbf{C}, \mathbf{A})$. In the biochemical network, species $\mathbf{A}$ activates (i.e., phosphorylates) $\mathbf{B}$, $\mathbf{B}$ activates $\mathbf{C}$, and species $\mathbf{C}$ inhibits (i.e., dephosphorylates) $\mathbf{A}$. The following biochemical equations allow us to examine the interaction among the species:

- Activation of $\mathbf{B}$ by $\mathbf{A}$ (edge $e_1 = (\mathbf{A}, \mathbf{B})$)

$$p\mathbf{A} + \mathbf{B} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} p\mathbf{AB}, \quad p\mathbf{AB} \underset{k_4}{\overset{k_3}{\rightleftharpoons}} p\mathbf{A} + p\mathbf{B}$$

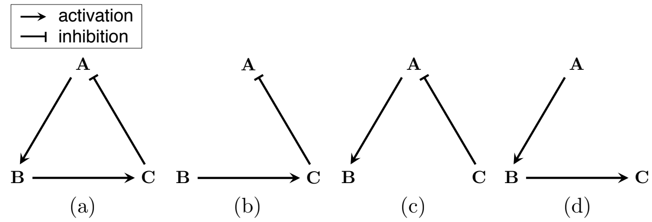

Fig. 2: Three component biochemical network $\mathcal{G}$ with connection (a) $\mathcal{E}$; (b) $\mathcal{E} \setminus \{e_1\}$; (c) $\mathcal{E} \setminus \{e_2\}$; and (d) $\mathcal{E} \setminus \{e_3\}$.

- Activation of $\mathbf{C}$ by $\mathbf{B}$ (edge $e_2 = (\mathbf{B}, \mathbf{C})$)

$$p\mathbf{B} + \mathbf{C} \underset{k_6}{\overset{k_5}{\rightleftharpoons}} p\mathbf{BC}, \quad p\mathbf{BC} \underset{k_8}{\overset{k_7}{\rightleftharpoons}} p\mathbf{B} + p\mathbf{C}$$

- Inhibition of $\mathbf{A}$ by $\mathbf{C}$ (edge $e_3 = (\mathbf{C}, \mathbf{A})$)

$$p\mathbf{C} + p\mathbf{A} \underset{k_{10}}{\overset{k_9}{\rightleftharpoons}} p\mathbf{CA}, \quad p\mathbf{CA} \underset{k_{12}}{\overset{k_{11}}{\rightleftharpoons}} p\mathbf{C} + \mathbf{A}.$$

Here, $p\mathbf{A}$ and $\mathbf{A}$ denote the active (phosphorylated) version and the inactive version of $\mathbf{A}$, respectively, and $p\mathbf{AB}$ is the complex of $\mathbf{A}$ and $\mathbf{B}$. Others are defined in a similar way. Let $x_1 = [p\mathbf{A}]$, $x_2 = [\mathbf{A}]$, $x_3 = [p\mathbf{B}]$, $x_4 = [\mathbf{B}]$, $x_5 = [p\mathbf{C}]$, $x_6 = [\mathbf{C}]$, $x_7 = [p\mathbf{AB}]$, $x_8 = [p\mathbf{BC}]$ and $x_9 = [p\mathbf{CA}]$, where $[\mathbf{M}]$ denotes the concentration level of protein $\mathbf{M}$. The dynamics of the biochemical concentrations with the signaling network $\mathcal{G}$ in Figure 2 (a) can then be modeled by (1) with the vector field $f_{\mathcal{E}}$ such that

$$\begin{aligned}
(f_{\mathcal{E}})_1 &= -k_1 x_1 x_4 + k_2 x_7 + k_3 x_7 - k_4 x_1 x_3 - k_9 x_1 x_5 + k_{10} x_9 \\
(f_{\mathcal{E}})_2 &= k_{11} x_9 - k_{12} x_2 x_5 \\
(f_{\mathcal{E}})_3 &= k_3 x_7 - k_4 x_1 x_3 - k_5 x_3 x_6 + k_6 x_8 + k_7 x_8 - k_8 x_3 x_5 \\
(f_{\mathcal{E}})_4 &= -k_1 x_1 x_4 + k_2 x_7 \\
(f_{\mathcal{E}})_5 &= k_7 x_8 - k_8 x_3 x_5 - k_9 x_1 x_5 + k_{10} x_9 + k_{11} x_9 - k_{12} x_2 x_5 \\
(f_{\mathcal{E}})_6 &= -k_5 x_3 x_6 + k_6 x_8 \\
(f_{\mathcal{E}})_7 &= k_1 x_1 x_4 - k_2 x_7 - k_3 x_7 + k_4 x_1 x_3 \\
(f_{\mathcal{E}})_8 &= k_5 x_3 x_6 - k_6 x_8 - k_7 x_8 + k_8 x_3 x_5 \\
(f_{\mathcal{E}})_9 &= k_9 x_1 x_5 - k_{10} x_9 - k_{11} x_9 + k_{12} x_2 x_5,
\end{aligned}$$

where we used the mass-action kinetics. We can see that the vector field has additive interconnections, i.e.,

$$f_{\mathcal{E}} = f_{\{e_1\}} + f_{\{e_2\}} + f_{\{e_2\}}$$

with $f_{\emptyset} = 0$. Note that the vector field is *quasi-positive*, i.e.,

$$(f_{\mathcal{E}})_i(x_1, \cdots, x_{i-1}, 0, x_{i+1}, \cdots, x_9) \geq 0, \ \forall i = \{1, \cdots, 9\},$$

and the quasi-positivity guarantees the solution to be non-negative invariant. If we let $X := x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + 2x_7 + 2x_8 + 2x_9$, then $\dot{X} = 0$. Hence, $x_i \leq X(0)$ due to the non-negative invariance. Therefore, we can set the state space as the compact subset $\mathcal{X} = [0, X(0)]^9 \subset \mathbb{R}^9$. Furthermore, the vector field $f_{\widetilde{\mathcal{E}}}$ is Lipschitz continuous and twice differentiable in $\mathcal{X}$ for any interconnection $\widetilde{\mathcal{E}}$.

In this example, we choose the output to be the concentration of the active protein $p\mathbf{B}$, i.e.,

$$y(t) = h(x(t)) = x_3(t).$$

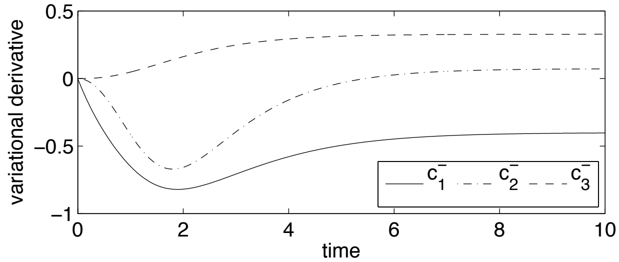

Fig. 3: Trajectory of the variational derivatives $(c_1^-, c_2^-, c_3^-) := (D_{(\mathcal{E},\mathcal{E}\setminus\{e_1\})}y, D_{(\mathcal{E},\mathcal{E}\setminus\{e_2\})}y,\ D_{(\mathcal{E},\mathcal{E}\setminus\{e_3\})}y)$ over time interval $t \in [0, 10]$.

We assume that $x_i(0) = 1$ for all $i = 1, \cdots, 9$, and $k_{2j-1} = 1$ and $k_{2j} = 10^{-3}$ for $j = 1, \cdots, 5$.

Our goal is to solve the interconnection pursuit problem to maximize the increase in $[p\mathbf{B}]$, i.e., $\Delta_{(\mathcal{E},\widetilde{\mathcal{E}})}y(\tau)$, where the maximal interconnection is given by $\mathcal{E}^* := \{e_1, e_2, e_3\}$. We must consider all possible $\widetilde{\mathcal{E}}$'s that are subsets of $\{e_1, e_2, e_3\}$. By comparing all 2^3 cases, we have the optimal interconnection as

$$\{e_1, e_2\} = \arg\max_{\widetilde{\mathcal{E}}\in\Gamma} \Delta_{(\mathcal{E},\widetilde{\mathcal{E}})}y(\tau), \quad \tau \in (0, 10].$$

However, for the approximate interconnection pursuit problem, i.e, Optimization 2, however, it is sufficient to compute $c^- = (D_{(\mathcal{E},\mathcal{E}\setminus\{e_1\})}y, D_{(\mathcal{E},\mathcal{E}\setminus\{e_2\})}y,\ D_{(\mathcal{E},\mathcal{E}\setminus\{e_3\})}y)$ because the vector field has a additive interconnections. Note that $\mathcal{E}^* = \mathcal{E}$, we have $c^+ = \emptyset$. In Fig. 2 (b)–(d), interconnections $\mathcal{E} \setminus \{e_1\}$, $\mathcal{E} \setminus \{e_2\}$ and $\mathcal{E} \setminus \{e_3\}$ are depicted. Using the adjoint-based formula for the variational derivative (9) and Algorithm 1, we compute the variational derivatives over $(0, 10]$ as shown in Fig. 3. If we use Algorithm 2 (with threshold $\delta = 0$), then the optimum of the approximate problem is as follows:

$$\begin{aligned}\{e_1, e_2\} &= \arg\max_{\widetilde{\mathcal{E}}\in\Gamma} D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(\tau), \quad \tau \in (0, T_1],\\ \{e_1\} &= \arg\max_{\widetilde{\mathcal{E}}\in\Gamma} D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(\tau), \quad \tau \in (T_1, 10],\end{aligned}$$

where $T_1 \approx 5.46$. In other words, the approximate problem, Optimization 2, finds the global optimum of the original problem, Optimization 1, in $(0, T_1]$. Even in $(T_1, 10]$, the solution obtained by the approximate problem is the third best interconnection out of the total eight connections. Note that, if we choose the threshold δ as 0.075 or higher, then we can obtain the global optimum of the original problem by solving the approximate problem for all $T \in (0, 10]$.

V. Conclusion And Future Work

Inspired by the directional derivative approach in functional analysis, we introduced a variational derivative for nonlinear networked systems with respect to the interconnection topology. We derived an analytical expression for the derivative by introducing an adjoint state, and provided bounds relating our derivative to the actual change in system behavior. The additivity assumption on the vector field yielded separability of the variational derivative that further enabled us to develop an efficient solution to an approximate interconnection pursuit problem with applications to optimization and inference of biochemical reaction networks. We also believe the variational derivative will be useful for analyzing output sensitivity in nonlinear biochemical and power system models with respect to the interconnection topology.

In future work we plan to further study convergence and suboptimality of the approximate interconnection pursuit problem and to generalize the variational method to the networked control systems setting where we anticipate finding applications to estimating sensitivity of power transmission systems to connectivity structure.

Appendix
Proof of Proposition 3

We first note that

$$\begin{aligned}0 &= f_{\mathcal{E}}(\xi),\\ 0 &= f_{\widetilde{\mathcal{E}}}(\tilde{\xi}) = f_{\mathcal{E}}(\tilde{\xi}) + f_{(\mathcal{E},\widetilde{\mathcal{E}})}(\tilde{\xi}).\end{aligned}$$

Subtracting one with another, we have

$$0 = \frac{\partial f_{\mathcal{E}}(\xi)}{\partial x}(\tilde{\xi} - \xi) + f_{(\mathcal{E},\widetilde{\mathcal{E}})}(\tilde{\xi}) + \mathcal{O},$$

where $\mathcal{O} := O(\|\xi - \tilde{\xi}\|^2)$. This implies that

$$\tilde{\xi} - \xi = -\left(\frac{\partial f_{\mathcal{E}}(\xi)}{\partial x}\right)^{-1} f_{(\mathcal{E},\widetilde{\mathcal{E}})}(\tilde{\xi}) + \mathcal{O}.$$

The Taylor expansion allows us to deduce

$$\begin{aligned}\Delta_{(\mathcal{E},\widetilde{\mathcal{E}})}y(T) = h(\tilde{\xi}) - h(\xi) &= \frac{\partial h(\xi)}{\partial x}(\tilde{\xi} - \xi) + \mathcal{O}\\ &= -\frac{\partial h(\xi)}{\partial x}\left(\frac{\partial f_{\mathcal{E}}(\xi)}{\partial x}\right)^{-1} f_{(\mathcal{E},\widetilde{\mathcal{E}})}(\tilde{\xi}) + \mathcal{O}.\end{aligned} \tag{21}$$

On the other hand, we can rewrite (15) as

$$D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(T) = -\frac{\partial h(\xi)}{\partial x}\left(\frac{\partial f_{\mathcal{E}}(\xi)}{\partial x}\right)^{-1} f_{(\mathcal{E},\widetilde{\mathcal{E}})}(\xi) + O(e^{aT}) \tag{22}$$

because $f_{\mathcal{E}}(\xi) = 0$ and the interconnection additivity. Comparing (21) with (22), we obtain

$$\begin{aligned}&|\Delta_{(\mathcal{E},\widetilde{\mathcal{E}})}y(T) - D_{(\mathcal{E},\widetilde{\mathcal{E}})}y(T)|\\ &= -\frac{\partial h(\xi)}{\partial x}\left(\frac{\partial f_{\mathcal{E}}(\xi)}{\partial x}\right)^{-1} (f_{(\mathcal{E},\widetilde{\mathcal{E}})}(\tilde{\xi}) - f_{(\mathcal{E},\widetilde{\mathcal{E}})}(\xi)) + \mathcal{O}_2\\ &\le \left\|\frac{\partial h(\xi)}{\partial x}\left(\frac{\partial f_{\mathcal{E}}(\xi)}{\partial x}\right)^{-1} \frac{\partial f_{(\mathcal{E},\widetilde{\mathcal{E}})}(\xi)}{\partial x}\right\| \|\xi - \tilde{\xi}\| + \mathcal{O}_2,\end{aligned}$$

as desired, where $\mathcal{O}_2 := O(e^{aT}) + O(\|\xi - \tilde{\xi}\|^2)$. $\square$

References

[1] J. Cortés, S. Martínez, T. Karataş, and F. Bullo, "Coverage control for mobile sensing networks," *IEEE Transactions on Robotics and Automation*, vol. 20, no. 2, pp. 243–255, 2004.

[2] N. E. Leonard, D. A. Paley, F. Lekien, R. Sepulchre, D. M. Fratantoni, and R. E. Davis, "Collective motion, sensor, networks, and ocean sampling," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 48–74, 2007.

[3] V. D. Blondel, J. M. Hendrickx, and J. N. Tsitsiklis, "Continuous-time average-preserving opinion dynamics with opinion-dependent communications," *SIAM Journal on Control and Optimization*, vol. 48, no. 8, pp. 5214–5240, 2010.
[4] L. Scardovi, M. Arcak, and E. D. Sontag, "Synchronization of interconnected systems with applications to biochemical networks: an input-output approach," *IEEE Transactions on Automatic Control*, vol. 55, no. 6, pp. 1367–1379, 2010.
[5] M. M. Zavlanos and G. J. Pappas, "Distributed connectivity control of mobile networks," *IEEE Transactions on Robotics*, vol. 24, no. 6, pp. 1416–1428, 2008.
[6] M. Nabi-Abdolyousefi and M. Mesbahi, "Network identification via node knockout," *IEEE Transactions on Automatic Control*, vol. 57, no. 12, pp. 3214–3219, 2012.
[7] Y.-P. Tian and C.-L. Liu, "Robust concensus of multi-agent systems with diverse input delays and asymmetric interconnection perturbations," *Automatica*, vol. 45, pp. 1347–1353, 2009.
[8] G. Como, K. Savla, D. Acemoglu, M. A. Dahleh, and E. Frazzoli, "Robust distributed routing in dynamical networks - part I: locally responsive policies and weak resilience," *IEEE Transactions on Automatic Control*, vol. 58, no. 2, pp. 317–332, 2013.
[9] H. Alper, Y.-S. Jin, J. Moxley, and G. Stephanopoulos, "Identifying gene targets for the metabolic engineering of lycopene biosynthesis in escherichia coli," *Metabolic Engineering*, vol. 7, no. 3, pp. 155–164, 2005.
[10] M. Bansal, V. Belcastro, A. Ambesi-Impiombato, and D. di Bernardo, "How to infer gene networks from expression profiles," *Molecular Systems Biology*, vol. 3, no. 78, 2007.
[11] P. Kokotović and J. Heller, "Direct and adjoint sensitivity equations for parameter optimization," *IEEE Transactions on Automatic Control*, vol. 12, no. 5, pp. 609–610, 1967.
[12] A. Jameson, "Aerodynamic design via control theory," *Journal of Scientific Computing*, vol. 3, no. 3, pp. 233–260, 1988.
[13] T. M. Apostol, *Calculus*, 2nd ed. New York: Wiley, 1967.
[14] E. Polak, *Optimization: algorithms and consistent approximations*. Springer Verlag, 1997, vol. 124.
[15] H. L. Royden and P. M. Fitzpatrick, *Real Analysis*, 4th ed. Boston: Prentice Hall, 2010.
[16] J. Hadamard, "Sur les problèmes aux dèrivèes partielles et leur signification physique," *Princeton University Bulletin*, vol. 13, 1902.
[17] M. Arcak and E. D. Sontag, "A passivity-based stability criterion for a class of biochemical reaction networks," *Mathematical Biosciences and Engineering*, vol. 5, no. 1, pp. 1–19, 2008.
[18] P. Kundar, *Power System Stability and Control*, N. J. Balu and M. G. Lauby, Eds. New York: McGraw–Hill, 1994.
[19] W. Ma, A. Trusina, H. El-Samad, W. A. Lim, and C. Tang, "Defining network topologies that can achieve biochemical adaptation," *Cell*, vol. 138, pp. 760–773, 2009.
[20] M. J. Dunlop, R. S. Cox III, J. H. Levine, R. M. Murray, and M. B. Elowitz, "Regulatory activity revealed by dynamic correlations in gene expression noise," *Nature Genetics*, vol. 40, no. 12, pp. 1493–1498, 2008.