

Dissertation ETH Zurich No. 23845

**Inference and control for populations of systems:
from aggregative games to systems biology**

A thesis submitted to attain the degree of
Doctor of Sciences of ETH Zurich
(Dr. sc. ETH Zurich)

presented by

Francesca Parise

Dott.sa Magistrale, Università degli studi di Padova, Italy
born 10.10.1988 in Verona
citizen of Italy

accepted on the recommendation of

Prof. Dr. John Lygeros, examiner
Prof. Dr. John N. Tsitsiklis, co-examiner
Prof. Dr. Maria Elena Valcher, co-examiner

2016

Abstract

The objective of this thesis is the analysis and eventually the control of two different types of populations of systems: populations of rational agents and populations of biological systems. The main challenge for both types of populations lies in their large scale, which makes traditional single-system modeling and control methods computationally intractable and architecturally unfeasible. To overcome these difficulties, we exploit global control signals that affect the whole population and are designed based on aggregate information only.

More in detail, in the first part of the thesis we consider the problem of coordinating multi-user systems. The main challenge for this type of populations is that most of today's multi-user networks are large scale, with lack of access to local information and consist of users with diverse requirements. Consequently, traditional distributed optimization methods, that focus on a unique control objective and typically assume that the users are cooperative, cannot be applied. On the other hand, a new distributed control paradigm, where the infrastructure is kept simple and the agents make their decisions independently ("selfishly") is emerging. This new framework motivates the interaction of control with game theory: control actions must in fact include economic notions related to markets and incentives. In the first part of the thesis, we propose a framework to implement this new paradigm in a scalable way by exploiting the aggregative nature of many multi-user systems of practical interest, as demand-response markets or social and traffic networks, to name just a few. Specifically, we derive control schemes, either decentralized or distributed, that can be used to coordinate very large populations of agents to a Nash equilibrium, that is, to a stable configuration where no agent has incentive to unilaterally change its strategy.

In the second part of the thesis we focus on populations of biological systems. Specifically, we consider populations of cells whose internal state is determined by a series of stochastic reactions whose rate can be influenced by an external control signal, applied to the whole population. The most prominent examples are gene expression reaction networks, which are responsible for the fundamental operations of any living organism, and whose rate can be influenced by means of light or concentration signals. In contrast to the previous part of the thesis, biochemical reaction networks are natural systems (instead of technological ones) and therefore are usually unknown and affected by high levels of noise. Our main concern is the analysis and modeling of such systems. Recent technological advancements have enabled researchers to observe populations with thou-

sands of cells and have accurate recordings of the internal state of large samples over time. These samples can be used to derive accurate statistics not only of the average state of the cells in the population but also of higher order moments (as for example the variance). As first result we propose a systematic procedure to infer a model of the unknown population by exploiting the available external control signals and the population measurements just described. Secondly, we use the identified model to study the range of behaviors that such controlled stochastic biochemical reaction networks can exhibit under different choices of the available control input. These are in fact the first two fundamental steps needed to understand these systems and consequently be able to use them, either as part of synthetic circuits or integrated in live organisms, for example for drug synthesis or targeted medicine. Our theoretical results are validated with in-vivo experiments on a light-inducible gene expression system.

Sommario

Questa tesi si occupa di tematiche riguardanti l'analisi e il controllo di due differenti tipi di popolazioni di sistemi: popolazioni di agenti razionali e popolazioni di sistemi biologici. In entrambi i casi le popolazioni considerate sono composte da un numero molto elevato di sistemi, pertanto tradizionali tecniche di controllo, basate su modelli specifici per ogni singolo sistema, non sono né computazionalmente né tecnologicamente implementabili. Per superare queste difficoltà, nella tesi viene suggerito l'utilizzo di un unico segnale di controllo che influenza l'intera popolazione ed è basato solo su informazioni aggregate.

Più nel dettaglio, la prima parte della tesi considera il coordinamento di sistemi multi-agente. Tali popolazioni sono solitamente caratterizzate da un elevato numero di agenti con obiettivi differenti e da una limitata disponibilità di informazioni locali, che sono per la maggior parte private. Pertanto tradizionali metodi di ottimizzazione distribuita, nei quali viene considerato un solo obiettivo e gli agenti sono tipicamente cooperativi, non possono essere utilizzati per il controllo di tali popolazioni. In contrapposizione sta emergendo un nuovo paradigma di controllo distribuito, nato dall'interazione tra la teoria dei giochi e del controllo, secondo il quale l'azione di controllo deve essere esercitata tramite incentivi e prezzi a cui i singoli agenti reagiscono indipendentemente (egoisticamente). La prima parte della tesi suggerisce come implementare questo nuovo paradigma, per grandi popolazioni, sfruttando la natura aggregativa della maggior parte dei sistemi multi-agente di interesse. Alcuni esempi di applicazione sono gestione della domanda in mercati energetici, controllo di reti di interazione sociale o direzione del traffico stradale. Nello specifico, il contributo di questa tesi è lo sviluppo di schemi di controllo, decentralizzato o distribuito, che possono essere utilizzati per coordinare gli agenti verso un equilibrio di Nash, ovvero verso una configurazione stabile di strategie, nella quale nessun agente è incentivato a modificare il suo comportamento dato il comportamento del resto della popolazione.

La seconda parte della tesi considera popolazioni di cellule il cui stato è determinato da una serie di reazioni chimiche stocastiche. Nello specifico, particolare attenzione è riservata a reti di reazioni la cui frequenza dipende da un segnale di controllo esterno, comune a tutta la popolazione. Come principale esempio il lettore può considerare le reti di espressione genica: queste reti sono alla base del funzionamento di ogni cellula e possono essere influenzate tramite stimoli luminosi o chimici. Contrariamente alla prima parte della tesi, i sistemi di reazioni biochimiche sono sistemi naturali (invece che tecnologici) le cui dinamiche sono perlopiù sconosciute e solitamente affette da rumore. Il principale

obbiettivo per tali popolazioni è pertanto l'analisi e l'identificazione di modelli matematici, che possono essere usati per descrivere tali sistemi. Grazie a recenti avanzamenti tecnologici è possibile osservare popolazioni composte da migliaia di cellule e ottenere di conseguenza statistiche accurate non solo del comportamento medio della popolazione, ma anche di momenti di ordine superiore (come ad esempio la varianza). Il primo obiettivo di questa parte della tesi è proporre una procedura per l'identificazione e la calibrazione sistematica di modelli matematici di reti biochimiche, sfruttando i segnali di controllo a disposizione e le misure di popolazione sopra descritte. Il secondo obiettivo è quello di utilizzare i modelli identificati per analizzare i possibili comportamenti della popolazione in risposta a diversi stimoli esterni. Questi sono infatti i primi due passi necessari per la comprensione delle reti biochimiche controllate e quindi per il loro utilizzo, sia come parte di circuiti sintetici sia integrati in organismi viventi, per esempio per la sintesi di nuovi farmaci o nello sviluppo di cure personalizzate. I risultati teorici di questa tesi sono affiancati da risultati sperimentali relativi ad una rete di espressione genica sintetica implementata in cellule di lievito.