Distributed Optimization
Introduction

Richard Combes

Jie Lu

Alexandre Proutiere

Objectives

- Cooperative setting
- Provide a survey of recent advances in distributed optimization techniques

$$\min \sum_{i=1}^{m} f_i(x)$$

  - m independent agents **cooperating** towards a single objective
  - How can they reach the desired minimizer?
  - How much should they communicate?
  - How fast can they reach the objective?

Objectives

- Competitive setting
- Provide a survey of recent advances in convergence to Nash Equilibria in games

$$\forall i = 1, \ldots, m, \quad \min_{x_i} f_i(x_i, x_{-i})$$

  - m independent agents **competing** towards different objectives
  - Does the notion of Nash Equilibrium make sense?
  - Are there natural learning algorithms leading to NEs?
  - Can agents / players select socially efficient NEs?
  - How fast can they reach equilibrium?

Applications

- Large networked systems
  - Internet
  - AdHoc networks
  - Data centers
  - Sensor networks
  - Social networks
  - Economic networks
  - …
- New interaction paradigms
  - Resource allocation
  - Coordination
  - Estimation
  - Games over networks
  - …

Decentralized interactions

- We need new tools to understand the way agents interact in these large-scale networked complex systems

- Challenges
  - Lack of central authority
  - Network dynamics
  - Stochastic phenomena
  - Lack of (or partial) local communication among agents
  - …

Concrete examples

- Resource allocation in communication networks
  - Internet Congestion Control
  - Power control in wireless systems
  - Routing
  - Load balancing
- Games
  - Load balancing games
  - Routing games
  - Power control games
  - Marriage problems
  - …

Schedule

- Sept 10 / 10:15AM-12:15PM | Brinellv. 23 (B24) | Overview and basic concepts in optimisation
- Sept 12 / 10:15AM-12:15PM | Osquldasv. 6 (Q22) | Convexity, gradient descent and sub-gradient method
- Sept 17 / 10:15AM-12:15PM | Osquarsbacke 14 (E52) | Optimal first order methods
- Sept 19 / 10:15AM-12:15PM | Lindstedtv. 3 (E34) | Duality, dual decomposition, and ADMM
- Sept 24 / 10:15AM-12:15PM | Drottning Krist. 30 (L42) | Iterative methods, parallel computing, and gossiping algorithms
- Sept 26 / 10:15AM-12:15PM | Drottning Krist. 30 (L43) | Project session 1

Schedule

- Oct 1 / 10:15AM-12:15PM | Drottning Krist. 30KV (L22) | Stochastic optimization - Stochastic approximation
- Oct 8 / 10:15AM-12:15PM | Drottning Krist. 30KV (L22) | Sampling-based optimization
- Oct 15 / 10:15AM-12:15PM | Brinellv. 23 (B23) | Learning in games 1
- Oct 17 / 10:15AM-12:15PM | Brinellv. 23 (B24) | Learning in games 2
- Oct 22 / 10:15AM-12:15PM | Brinellv. 23 (B24) | Project session 2

Outline

- Part I: Convex optimization (Sept 12, Sept 17)
- Part II: Distributed optimization (Sept 19, Sept 24)
- Part III: Stochastic optimization (Oct 1, Oct 8)
- Part IV: Dynamics in games (Opt 15, Oct 17)

Part I: Convex optimization
Jie Lu

- Convexity
- Gradient descent, sub-gradient descent algorithms
- Optimal first-order methods
  - Convergence rate: lower bounds
  - Order-optimal algorithms
- Material:
  - Y. Nesterov, *Introductory lectures on Convex Optimization: A Basic Course*. Norwell, MA: Kluwer Academic Publishers, 2004.
  - D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1999.
  - D. P. Bertsekas, A. Nedich, and A. Ozdaglar, *Convex Analysis and Optimization*. Belmont, MA: Athena Scientific, 2003.
  - S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY: Cambridge University Press, 2004.

Part II: Distributed optimization
Jie Lu

- Duality, dual decomposition, and ADMM
- Iterative methods
- Parallel computing, and gossiping algorithms
- Material:
  - D. P. Bertsekas, A. Nedich, and A. Ozdaglar, *Convex Analysis and Optimization*. Belmont, MA: Athena Scientific, 2003.
  - S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY: Cambridge University Press, 2004.
  - S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*, Foundations and Trends in Machine Learning, 2011.

Part II: Distributed optimization
Jie Lu

- Material:
  - D. P. Bertsekas, J. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods,* Prentice-Hall, 1989.
  - IEEE JSAC special issue on Distributed Optimization, vol 8, 2006. *Mathematical decomposition techniques for distributed cross-layer optimization of data networks*, B. Johansson, P. Soldati and M. Johansson.

Part III: Stochastic optimization
Richard Combes

- Stochastic approximation
- Sampling-based optimization
  - Simulated annealing
  - Non-reversible dynamics
- Material:
  - V. Borkar, *Stochastic approximation: A dynamical systems viewpoint,* Cambridge University Press, 2008.
  - H. Kushner and G. G. Yin, *Stochastic approximation and recursive algorithm*. Springer, 2003.
  - S. Kirkpatrick; C. D. Gelatt; M. P. Vecchi, *Optimization by Simulated annealing*, Science, 1983.
  - J.R. Marden, P. Young, L.Y. Pao, *Achieving Pareto-optimality through distributed learning*, CDC, 2012.

Part IV: Learning in games
R. Combes / A. Proutiere

- Games and equilibria
- Nash dynamics
- Fictitious play
- No-regret dynamics
- Trial and error learning

- Material:
  - P. Young, *Strategic learning and its limit, Oxford* University Press, 2004.
  - D. Fudenberg and D. Levine, *The theory of learning in games*. MIT press, 2004.
  - P. Young, *Learning by trials and errors*, Games and economic behavior, 2009.

Grading policy

- Grading: P/F
- Credits: 8hp
- 2 project sessions
- Take-home exam

How to reach us?

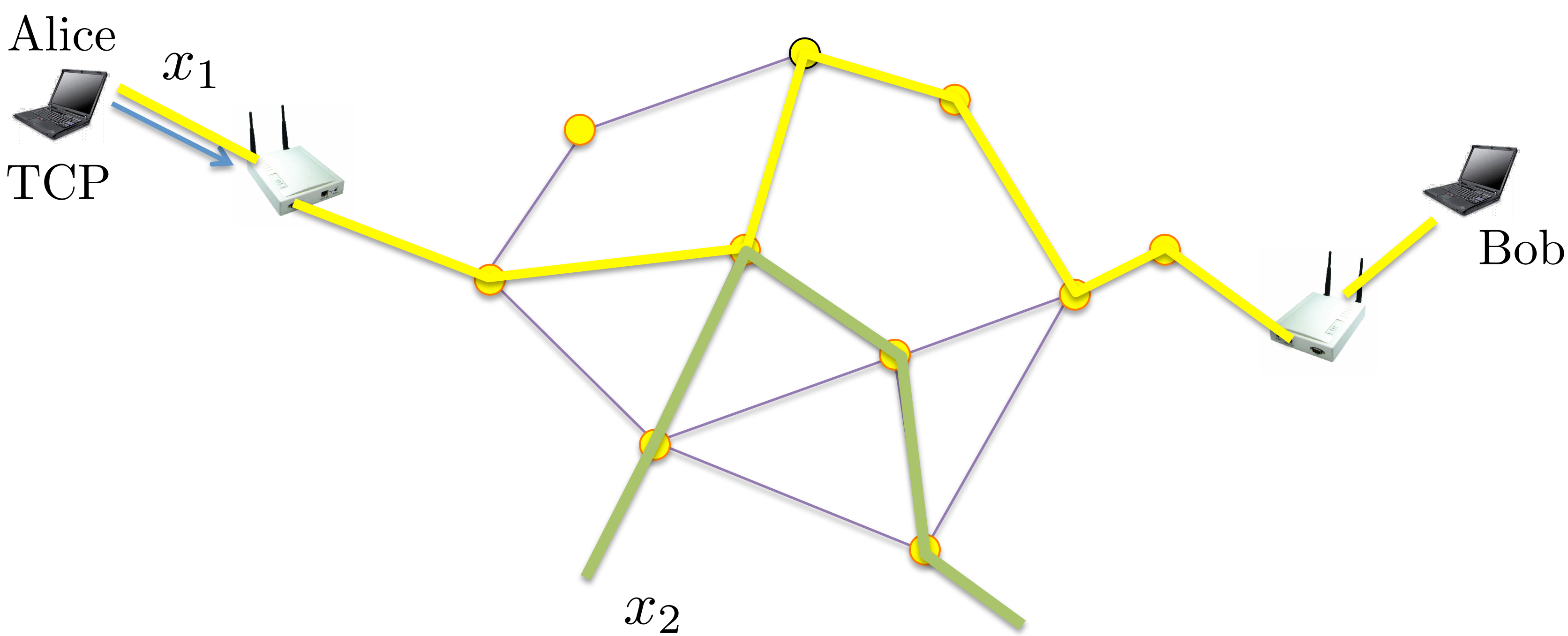
alepro@kth.se, jielu@kth.se, rcombes@kth.se

http://www.ee.kth.se/~alepro/DistriOptCourse/

A first fundamental example: Internet congestion control

Based on:
Rate control for communication networks:
shadow prices, proportional fairness and stability
Kelly-Maulloo-Tan, J. Oper. Res. Soc., 1998.

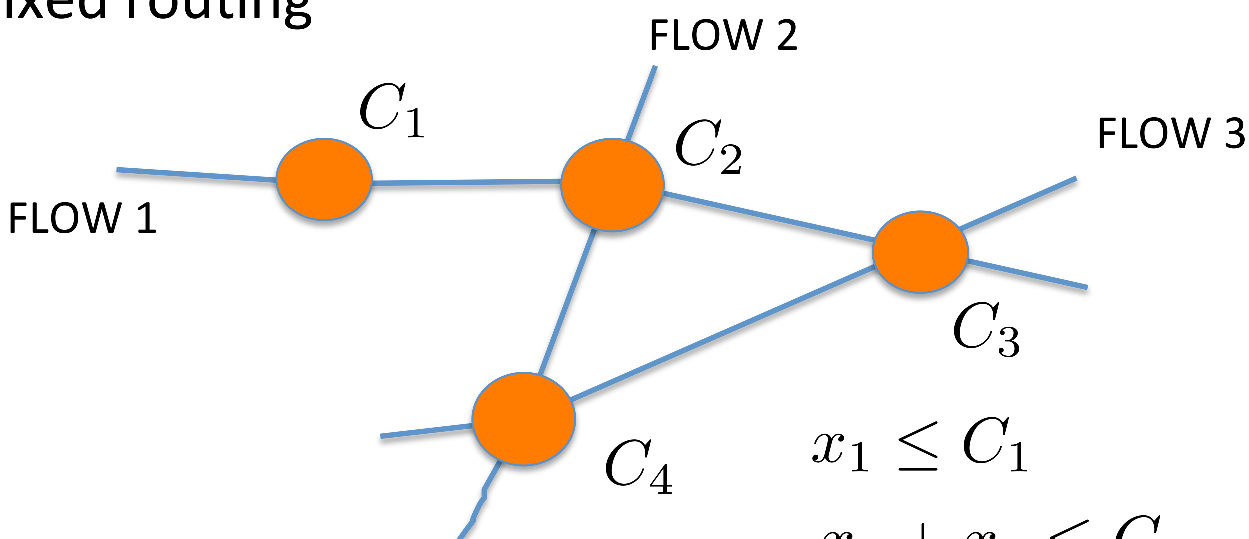
Internet congestion control



Objective of TCP: adapt the rates of sources to fairly and efficiently share network resources

A simple model

- Resources: a set of *L* links shared by a fixed population of *n* connections or data flows
- Fixed routing



$$x_1 \leq C_1$$

$$x_1 + x_2 \leq C_2$$

$$x_1 + x_3 \leq C_3$$

$$x_2 + x_3 \leq C_4$$

Network Utility Maximization

- The goal is to design distributed protocols converging to the solution of:

$$\text{maximize } \sum_{i=1}^{n} U_i(x_i)$$
$$\text{subject to } Rx \leq C$$

Previous example:

FLOWS

$$C = \begin{pmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{pmatrix} \qquad R = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

LINKS

Network Utility Maximization

- Utility functions:
  - Proportional fairness (Kelly): $U_i(\cdot) = \log(\cdot)$
  - α-fairness: $U_i(\cdot) = (\cdot)^{(1-\alpha)}/(1-\alpha)$
  - Max-min fairness (Rawls): $\alpha = \infty$
  - Max-thru: $\alpha = 0$

Decomposition

- Lagrangean:

$$L(x,\mu) = \sum_{i=1}^{n}(U_i(x_i) - x_i \sum_{l:R_{li}=1} \mu_l) + \sum_{l} \mu_l C_l$$

- Dual function:

$$q(\mu) = \sum_{i=1}^{n} \underbrace{\max_{x_i}(U_i(x_i) - x_i \sum_{l:R_{li}=1} \mu_l)}_{\text{Source sub-problems}} + \sum_{l} \mu_l C_l$$

Dual decomposition

- Link price update: for each link *l*

$$\mu_l(k+1) = \left[\mu_l(k) + \beta(\sum_{i:R_{li}=1} x_i(k) - C_l)\right]^+$$

- Source rate update:

$$x_i(k+1) = \arg\max_{x_i}(U_i(x_i) - x_i \sum_{l:R_{li}=1} \mu_l)$$

Convergence of dual GD algorithm

- The gradient of the dual function is lipschitz
  - Assume that $-U_i''(x_i) \geq 1/g > 0$
  - Let *L* and *S* be the length of the longest route and maximum number of sources using a given link, respectively

Lemma* We have:

$$\|\nabla q(\mu) - \nabla q(\mu')\|_2 \leq gLS\|\mu - \mu'\|_2$$

- … which ensures convergence of the algorithm

* Optimization flow control-I: Basic algorithm and convergence, **Low-Lapsley**, ACM/IEEE trans. on Networking, 1999.

Primal decomposition

- Source rate update:

$$x_i(k+1) = x_i(k) + \beta \left(U_i'(x_i(k)) - \sum_{l:R_{li}=1} \mu_l \right)$$

- Price update:

$$\mu_l(k+1) = p_l\left(\sum_{i:R_{li}=1} x_i(k+1) \right)$$

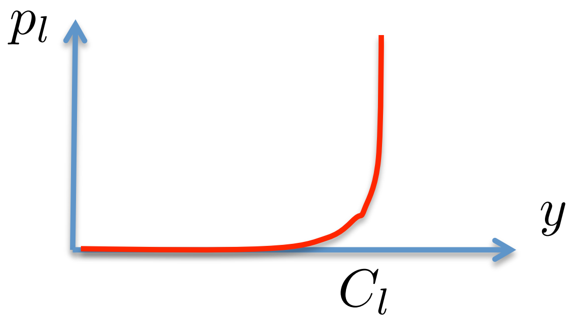
p_l : barrier function (to be defined later)

Convergence of the primal algorithm

Theorem* For appropriate choice of β, the primal algorithm converges to a solution of:

$$\max \sum_i U_i(x_i) - \sum_l \int_0^{\sum_{i:R_{li}=1} x_i} p_l(y)dy$$

- The barrier functions are increasing, and can be chosen so that we obtain a good approximation of the initial NUM problem



* Rate control for communication networks: shadow prices, proportional fairness and stability, **Kelly-Maulloo-Tan**, J. Oper. Res. Soc., 1998.